

Insight Academic Research Journal

Summer
Publication
2026

Table of Contents

Page 2: The Unintended Consequences of Affirmative Action on Minority Groups Outside the Policy's Focus *By Chantal Massachi*

Page 8: A Review of Gamification in Cognitive Assessment for Dementia *By Qiwen Hu*

Page 22: EcoFiber: Biodegradable Packaging Based on Oyster Mushroom Mycelium and Filamentous Algae Grown in a Photobioreactor *By Assylzhan Sailau, Semyon Chistyakov, Najmudin Adikhanov, Mansur Kakharman, Rudik Ivan, Aldabergen Amirali*

Page 35: Can a Dual-Reporter Senolytic Biosensor Visualize and Eliminate Senescent Cells Driving Tau and Amyloid-Beta Pathology in Alzheimer's Disease? *By Alinur Zaitkhan*

Page 41: The Correlation Between *Raphanus sativus L.* Seed Germination and Exposure to Zinc Oxide *By Kaiden Kim*

Page 57: "Tenzo" – Adaptive Tensegrity Exoskeleton *By Semyon Chistyakov, Najmudin Adikhanov, Assylzhan Sailau, Jaleel Gabitov, Yerkenur Zhanabekova, Arnur Togambayev*

Page 72: How Can Artificial Intelligence and Machine Learning Be Applied to Recruiting Athletes at the College Level, in Particular in Football? *By Jonah Khoshabe*

Page 81: Adolescent Depression and School-Based Mental Health Interventions *By Roxanna A. Abdollahi, Hannah L. Dunn*

Page 94: "Óskin" – A Smart Incubator Supporting a Favorable Environment for Plant Seed Germination *By Najmudin Adikhanov, Semyon Chistyakov, Assylzhan Sailau, Aidana Kobentai, Alinur Akezhanov, Alikhan Zhetpis*

Page 104: Uncovering Correlations Between the Length and Type of Parental Leave Taken by Elementary, Middle, and High School Parent-Teachers in Virginia and Teachers' Sense of Responsibility Towards Their Students *By Joy Svendsen, Andrew McCarthy*

Page 117: Evaluating the Reliability of AI-Generated Explanations for High School Mathematics Problems *By Abil Sagdayldaev*

Page 131: Development and Evaluation of a Lightweight Edge-Based Computer Vision System for Real-Time Driver Drowsiness and Micro-Sleep Detection *By Alisher Sabitov*

Page 150: Why Kyrgyzstan Is Not Yet a Regional Business Hub: A Marketing and Management Strategy for Promoting Kyrgyz Traditional Wool Products Globally *By Aruuke Meimanbaeva*

Page 173: Fear of Missing Out as a Mediator Between Social Media Screen Time and Anxiety in Adolescents *By Nazarova Khamida Mutalimovna*

Page 190: Artificial Intelligence in UFC: Fight Outcome Prediction and Computer Vision-Based Judging Support *By Iskhak Razakov*

Page 221: Cybersecurity Awareness and Online Safety Behaviour Among Teenagers: A Quantitative Survey of Secondary-School Students *By Nuriia Karagulova*

Page 250: Generative Artificial Intelligence and Employment in the Software Development Industry: A Mathematical and Labour-Market Analysis of Job Creation, Productivity, and Technological Unemployment from 2022 to 2026 *By Ruslan Tai*

Page 274: Higher-Altitude Wind Access and the Power Output of Airborne Wind Energy Systems: A Mathematical and Empirical Comparison with Conventional Wind Turbines in China *By Sultan Karachalov*

Page 286: Gender, Social Comparison, and Clothing Expenditure Among High School Students: A Behavioural Economics Perspective *By Mikhail Razygraev*

The Unintended Consequences of Affirmative Action on Minority Groups Outside the Policy's Focus By Chantal Massachi

Does affirmative action negatively impact minority groups, not in its policy? According to the Pew Research Center, a nonpartisan fact tank that informs the public, "49% of Americans said the consideration of race and ethnicity makes the overall admissions process less fair, while only 20% said it makes the process fairer." Affirmative action (AA) began in the 1960s, as part of the Civil Rights Movement, to ensure equal opportunities in jobs and education for all races, colors, genders, religions, sexes, and origins, leading colleges to provide more opportunities for disadvantaged groups (Policy Circle, 10). Aiming to help historically underrepresented students who are Black, Hispanic/Latino, American-Indian/Alaska-Native, and Native-Hawaiian/Pacific Islander groups. In 2003, Supreme Court cases helped shape AA's legal policies by addressing the balance between promoting diversity and discrimination against non-minorities and minorities not mentioned in the policy (EBSCO 2). Before the Supreme Court's decision in the 2014 Harvard case, they decided that promoting diversity in higher education is a government interest. However, it was hard for places of higher education to show that their AA programs were part of their mission and would pass judicial standards. While there are loopholes, like how private colleges and universities don't have the same constitutional barriers as public institutions. AA programs at private universities are still able to be legally challenged because of Title VI of the Civil Rights Act. Stating that private institutions that get federal funding are unable to discriminate against applicants. This makes sure that affirmative action programs at private universities meet the same standards as public institutions (NAICU 2). So, if AA only takes into consideration Black, Hispanic/Latino, American-Indian/Alaska-Native, and Native-Hawaiian/Pasicifc Islander students, what happens to those who aren't considered, such as Asians and Middle Easterns?

One of the ethnicities at a disadvantage due to affirmative action is the Asian applicants. According to the Harvard Law Review, a student-run journal of legal scholarship publishing, Asian American students have to do better than all other groups to have equal chances of admission. "One study showed that in order to be admitted to certain selective institutions, Asian American applicants needed to score — on the 1600 point scale of the "old SAT" — 140 points higher than whites, 270 points higher than Hispanics, and 450 points higher than African Americans if other factors are held equal." This illustrates the consequences of AA to Asian minorities when claiming how much higher the standard is for Asian applicants. Additionally, because Asian Americans aren't recognized in AA, it makes it even harder for them to get a 'leg up' in the admission process. According to the PEW Research Center, a nonpartisan fact tank, Asian Americans say that selective colleges that consider race and ethnicity will unfairly impact admissions and will affect the quality of admitted students. Based on a PEW Survey, 53% of English-speaking Asian Americans think that considering race and ethnicity in admissions makes it less fair. Additionally, 36% of English-speaking Asians think that race-based admissions lead to less-qualified students. This evidence highlights just how many Asians are against AA, as well

as how they believe it hinders the number of qualified students. Despite how much higher their average scores need to be, they feel that it is still unfair and will later negatively impact them as well as the colleges. Additionally, the California Law Review, University of Berkeley's Law School journal, states that SFFA (Students for Fair Admissions) claims that Harvard's AA policy isn't the source of their negative treatment toward Asians. However, Harvard's "preferences for African-American and Hispanic applicants couldn't explain the disproportionately negative effect Harvard's admission system has on Asian Americans." Meaning to say that AA has no real reason for the "Asian penalty" placed on its students. This connects back to Harvard Law Review's quote stating how much harder it is for Asian students to get admitted because of the academic inflation put on them. Adding on, the Hoover Institute, a peer-reviewed journal aiming to improve the human condition by advancing ideas, gives reason for Harvard's discrimination that helps to give African-Americans an advantage over white and Asian applicants. It's not hard to see Harvard's African-American applicants getting the benefit during the admission process. Harvard's defense, "Because of the size of the applicant pool relative to Harvard's available slots, Harvard cannot admit all applicants who would succeed academically." However, with a constant surplus of top-tier students applying, there is no need to weigh down the school's prestige to 'bridge the academic gap' and set certain students up for failure. Lastly, whether a top university or regular college or institution, Asian Americans' achievements are being overlooked and even seen as less simply because they are always overperforming and not considered underrepresented in AA. Overall, Asian students are one of the penalized minorities that are affected by affirmative action.

Lastly, another minority group that is underrepresented due to affirmative action is Middle Easterns. According to Hannah Mesouani, with a PhD in Education for Social Justice at the University of San Diego, MENA (Middle-Eastern, North-African) is a "crisis-associated identity" becoming more common after 9/11 and its accompanying anti-terrorism language. Either planned or not, 'white coding' MENA students relieves them of the benefits given to those who are AA applicants. These Middle Eastern students are being stereotyped in admissions because of a tragedy that happened in the past. They are being excluded from AA and, therefore, overlooked in the admission process. According to the CATO Institute, a non-partisan think tank that focuses on global issues, AA gives a disadvantage to successful groups through limited benefits, e.g., admissions. Whether MENA students would benefit from these policies depends on how they compare to White applicants; if they are better educated or make more money, they would face a similar discrimination that Asians do. With the discrimination to Asians students still being treated unfairly in admissions after SFFA v. Harvard, and MENA applicants could be treated similarly if given a new category because of their high success. However, AA would still harm MENA students less than Asians; both would be at a bigger disadvantage than White applicants (Nowrasteh 3). Similarly to how Asian applicants are underrepresented, Middle Eastern students are also very well educated and would be at a disadvantage if placed into their own category. It would highlight their academic and economic success and how they have had no historical disadvantage like other races in the policy. This also raises a concern that if given their

own category, MENA students deal with the same academic penalty, similar to the "Asian penalty." This can also lead to something like the incompetence theory, which is when people are given help or an advantage, it separates them from others, and can make others think that they are less capable. And for Asian Americans and Middle Easterns, AA can be seen as a double-edged sword. On one hand, they can be given their own category and recognition in admission, but on the other, it can make them appear to need help and seem less capable. This connects to the Columbia Undergraduate Law Review, Columbia University's premier undergraduate legal publication, which states that MENA students have twice the burden: being considered white and its accompanying benefits, but also not being given their own recognized minority group, becoming excluded from affirmative action's assistance to minority students. The National Association of Iranian Americans tried to achieve recognition for Arab Americans to become a minority in the Small Business Administration's affirmative action program, but it failed. Claiming that Middle Eastern Americans aren't qualified to be minorities because they are not "socially disadvantaged groups subjected to racial or ethnic prejudice or cultural bias within American society because of their identities as members of groups." Although studies and courts have proven that Arab Americans are discriminated against as a minority group and are subjected to bias and prejudice in America, it is important to note how the Supreme Court has considered Arab Americans different from white Americans because of its bias. Even though Middle Easterners are considered a minority, it is surprising that they wouldn't be considered in AA's policy, connecting back to an underlying theme of discrimination because of past acts, as well as not being seen as disadvantaged; however, they are continually being discriminated against by these stereotypes. It's especially concerning how time and time again the Supreme Court has shown its bias and gone against its word, similarly to SFFA v. Harvard. Students and unions fight to get positive recognition, but are stuck with the original negative effects of AA. According to the Journal of Social and Political Psychology, a peer-reviewed journal publishing articles from different perspectives, A study among Arab Muslims in metropolitan Chicago found that 63% of Arabs believed that they weren't considered white, and a participant in the study shared a story on how they couldn't be considered for AA because "they said you will be considered White. But of course, in real life, we are not." The goal of this study is to show how Middle Easterns are placed into a category. Similar to Columbia Law Review, Middle Easterns are at a crossroads between being considered white and having their own category in AA. Furthermore, this connects back to how Asian Americans, despite being overly qualified, are still denied opportunities because they aren't part of affirmative action's policy. Two minorities that are outside of affirmative action's plan to increase equality are being discriminated against, and by getting rid of affirmative action altogether, Asians and Middle Easterns can be given a fair shot at getting into colleges.

The issue of inequality towards minorities not being helped by affirmative action has a simple solution: getting rid of affirmative action. According to Richard C. Atkinson, the

president emeritus of the University of California, former director of the National Science Foundation, UC's had two-tier admissions that took 40-60% of each freshman class to be

selected solely because of their grades, test scores, talents, and achievements for almost 30 years. After review and elimination of those tiers, students are chosen because of a range in their academic achievements and characteristics. They are further evaluated in how well they did in their given environment; admissions would ask questions like how they used their opportunities and how they dealt with any disadvantages to evaluate students. By determining students based on their academic standing and circumstances, what they were able to achieve, it takes away the unintended consequences put onto certain students solely because of their race. This race-blind application process will result in more students who are qualified, instead of just a percentage of each race per class. According to Denise-Marie Ordway, a Pulitzer Prize finalist and who served on the board of directors of the Education Writers Association, Stanford is starting to bring about a "holistic" approach to the application process. Aiming to admit students based on their extracurriculars, accomplishments, experiences, as well as traditional test scores, awards, and GPA. Stanford uses this approach to "understand how you, as a whole person, would grow, contribute and thrive at Stanford, and how Stanford would, in turn, be changed by you." By going back to the more traditional way of admitting students, based on merit, but also understanding who they admit, schools like Stanford set themselves and their students up for success. Connecting back to the root of the problem, since Middle Easterns are not considered to be a minority under AA's policies, having race-blind admission will shift the focus on students' achievements rather than the color of their skin, which will help fight against the application crisis. According to Hugh Ferguson, the NAASFAA Managing Editor, adopting a "percentage plan system" grants automatic admission to state universities for students who graduate within a certain top percentage of their class. This policy did show an increase in diversity at some schools, but it wasn't as effective as race-conscious admission policies. " Even though state universities guarantee automatic admission to top students, that should help students of any race (even historically disadvantaged) guarantee a higher education. However, becoming a top student is not because of race but because of drive and discipline. Yes, there are disadvantages, and that is why universities should take that into account; however, race doesn't correlate with circumstance, and each student should be able to explain how they did the best they could in the given circumstances.

Overall, affirmative action disadvantages minorities such as Asians and Middle Easterners outside of its policy. According to Samantha M. Greig, School of Psychology, University of Queensland, these affirmative action policies that are put in place to help minorities may hint that people in those groups aren't able to achieve admission because of merit. Leading to the incompetence theory, because of this, AA policies can have unintended negative consequences when put into place. The people most affected by these are Asian students, competing with other students over grade inflation. And Middle Easterns are facing

discrimination and being overlooked. Implementing race-blind policies, it helps set students apart instead of making them a minority that needs to be filled, fixing the issue of anyone outside of affirmative action's policy. By putting this into effect, it will make college admissions processes more fair and deliver higher-qualified students who can succeed. All in all, college admissions is a stressful time for everyone, and it is a chance for students to show off all of their hard work, not to be concerned about whether they are in the college's "preferred" minority group.

Works Cited

"Affirmative Action - the Policy Circle." The Policy Circle, 9 Jan. 2025, www.thepolicycircle.org/briefs/affirmative-action/ Accessed 15 Jan. 2026. Clapp, Marlene. "Affirmative Action in Higher Education | EBSCO." EBSCO Information Services, Inc. | Www.ebsco.com, 2023, www.ebsco.com/research-starters/education/affirmative-action-higher-education. Accessed 15 Jan. 2026.

Epstein, Richard. "The Affirmative-Action Showdown." Hoover Institution, 31 Jan. 2022, www.hoover.org/research/affirmative-action-showdown. Accessed 3 Feb. 2026.

Feingold, Jonathan. "SFFA v. Harvard: How Affirmative Action Myths Mask White Bonus." California Law Review, Apr. 2019, www.californialawreview.org/print/sffa-v-harvard-how-affirmative-action-myths-mask-white-bonus. Accessed 14 Jan. 2026.

Ferguson, Hugh . "NASFAA | Analysts Outline 5 Alternatives to Race-Conscious College Admission Policies ahead of SCOTUS Decision." Www.nasfaa.org, 27 June 2023, www.nasfaa.org/news-item/31017/Analysts_Outline_5_Alternatives_to_Race-Conscious_College_Admission_Policies_Ahead_of_SCOTUS_Decision. Accessed 15 Jan. 2026.

Gramlich, John. "Americans and Affirmative Action: How the Public Sees the Consideration of Race in College Admissions, Hiring." Pew Research Center, Pew Research Center, 16 June 2023, www.pewresearch.org/short-reads/2023/06/16/americans-and-affirmative-action-how-the-public-sees-the-consideration-of-race-in-college-admissions-hiring/ Accessed 15 Jan. 2026.

Greig, Samantha M, et al. "The Help That Hinders? A Meta‐Analysis of Reactions to Affirmative Action." European Journal of Social Psychology, vol. 53, no. 7, 10 Sept. 2023, https://doi.org/10.1002/ejsp.2991 Accessed 15 Jan. 2026.

Hakim, Nader H, and Nyla R Branscombe. "White, or Not Quite? Predicting Arab American Responses to Racial Categorization Forms." Journal of Social and Political Psychology, vol. 11, no. 1, 3 May 2023, pp. 182–198, https://doi.org/10.5964/jspp.5503.

Harvard Law Review. "The Harvard Plan That Failed Asian Americans." Harvard Law Review, 7 Dec. 2017, harvardlawreview.org/print/vol-131/the-harvard-plan-that-failed-asian-americans/ Accessed 14 Jan. 2026.

"NAICU - Affirmative Action." Naicu.edu, 2016,

www.naicu.edu/policy-advocacy/issue-briefs/affirmative-action/. Accessed 15 Jan. 2026.
Nowrasteh, Alex . “The Consequences of a Middle Eastern or North African (MENA) Survey Question.” Cato Institute, 28 Sept. 2023, www.cato.org/briefing-paper/consequences-middle-eastern-or-north-african-mena-survey-question?fbclid=PAAaaNSt6kIjJjEzYiBGikmTQB7Ux2YxVANzokSdBJ8jZikAOoUCh8uU3_MtQ. Accessed 14 Jan. 2026.
Ordway, Denise-Marie. “Race-Neutral Alternatives to Affirmative Action in College Admissions: Research.” The Journalist’s Resource, 29 June 2023, journalistsresource.org/education/race-neutral-alternatives-affirmative-action-college-diversity/. Accessed 15 Jan. 2026.
Pew Research Center. “Topic: Affirmative Action in College Admissions.” Pew Research Center, 20 Dec. 2024, www.pewresearch.org/fact-sheet/topic-affirmative-action-in-college-admissions/ Accessed 15 Jan. 2026.
Ruiz, Neil G., et al. “Asian Americans Hold Mixed Views around Affirmative Action.” Pew Research Center, 8 June 2023, www.pewresearch.org/race-and-ethnicity/2023/06/08/asian-americans-hold-mixed-views around-affirmative-action/. Accessed 14 Jan. 2026.
Sahawneh, Zaid. “The Lack of Arab American Categorization in the U.S. Census and Its Legal Implications for Protection against Discrimination.” Columbia Undergraduate Law Review, 19 Oct. 2022, www.culawreview.org/journal/the-lack-of-arab-american-categorization-in-the-us-census-and-its-legal-implications-for-protection-against-discrimination. Accessed 15 Jan. 2026.

A Review of Gamification in Cognitive Assessment for Dementia By Qiwen Hu

Abstract

Dementia and Alzheimer's disease (AD) is becoming a global challenge, and early cognitive screening is essential for delaying the progression of this disease. Gamification offers an alternative to traditional paper-and-pencil tests for assessing cognitive function. This review summarizes the benefits, challenges, and current status of gamified cognitive assessment for dementia screening. Gamified tools improve user motivation, support portable and home-based use, and enable automatic data collection. Key limitations include insufficient screening accuracy, difficulty measuring isolated cognitive domains, technical barriers for older adults, small sample sizes of many current researches, and lack of clinical standardization. Existing evidence confirms that gamification is effective for preliminary screening of early AD. However, it is not suitable for formal diagnosis. Future work should focus on improving accuracy, eliminating technical barriers and designing standardized gamified cognitive assessments.

Introduction

Gamification provides some benefits in both screening and training cognitive functions, which is important for dealing with the increasingly common dementia. There are currently estimated to be over 55 million people worldwide living with dementia; that number is set to rise to 139 million by 2050 (Alzheimer's Disease International). Alzheimer's Disease (AD) first presents as a set of mild cognitive symptoms referred to as mild cognitive impairment (MCI). This pre-stage of AD typically begins with short-term memory impairment, followed by declines in other cognitive domains, specifically including: language disorders (e.g., difficulty finding appropriate words), impaired episodic memory (e.g., forgetting recent conversations), executive dysfunction (e.g., difficulty completing familiar tasks), and visuospatial dysfunction (e.g., getting lost in familiar environments) (Porsteinsson et al.). Although patients at this stage may exhibit mild functional deficits, their daily lives remain relatively independent. Patients will enter the AD stage after the MCI stage: the cognitive impairment increases, and all the symptoms that appear during the MCI stage will become more severe, such as further decline in memory of recent events and decline in visuospatial function.

AD is an irreversible progressive disease, and patients gradually lose the ability to live independently (Porsteinsson et al.). Caretaking duties can be very costly and time-consuming for their family members. According to the information released by WHO in 2019, dealing with dementia costs US$1.3 trillion globally, and half of these costs are covered by informal caregivers such as patients' family members. They provide about 5 hours of care every day on average (World Health Organization). It is, therefore, important to diagnose AD as early as possible and take action to control the progression of the disease.

Dementia (symptoms, epidemiology, treatment, assessment)

Gamification provides a new approach for addressing dementia, a condition with pressing epidemiological implications. As noted, the global prevalence of dementia is estimated at over 55 million, projected to reach 139 million by 2050 (Alzheimer's Disease International). Clinically, dementia manifests through a spectrum of symptoms: starting with MCI-related short-term memory loss, language disorders, impaired episodic memory, executive dysfunction, and visuospatial dysfunction, progressing to severe cognitive impairment in AD where patients lose independent living abilities (Porsteinsson et al.). Currently, there is no cure for AD, making early detection and intervention critical to slowing progression. Traditional assessment methods, such as paper-and-pencil tests like the Mini-Mental State Examination (MMSE) and California Verbal Learning Test (CVLT), have long been used but face limitations in accessibility, participant engagement, and efficiency—gaps that gamification aims to address.

Benefits of Gamifying Cognitive Assessment

1. Increased Motivation

One of the most important benefits of gamifying cognitive assessment is to increase participants' motivation and help them finish tests or maintain training. A review by Lumsden et al. reported that studies measuring intrinsic motivation showed that gamified versions of assessments are associated with higher levels of motivation. Gamified assessments are considered to have a higher entertainment value, which increases participants' motivation (Tong et al., “A Serious Game”). One example of this is a pilot study on the game “Kitchen and Cooking,” in which participants were asked to “cook food” in a digital game. According to a self-report questionnaire, participants found the game interesting. They experienced more positive than negative emotions while playing the game and did not feel fatigued before or after the game. During the 4-week trial, participants spent an average of over 5 hours playing the game; approximately 1 hour and 21 minutes per week. Eighty-five percent of scenarios were completed at home, with over 70% of scenarios successfully passed. This sustained and autonomous participation demonstrates the continuous motivation driven by the gamification design (Manera et al.).

Another study showed a similar result: participants with neurocognitive disorders rated the game from neutral to satisfied (Wang et al.). In this study, the researchers designed a digital game named Game-Based Cognitive Assessment (GBCA) for dementia screening and compared its screening effectiveness to commonly used screening tools like the MMSE. Results showed that GBCA is an effective tool for screening cognitive degenerative diseases, and patients did not need professional instruction, which makes it a suitable tool for early screening of dementia (Wang et al.). Pérez-Rodriguez et al. introduced their Panoramix battery for MCI and AD; results showed that patients' motivation to use video games increased, suggesting this could be a valuable tool for screening. The researchers suggested that this is probably because gamification provided additional versatility,

interactivity, and personalization capabilities (Valladares-Rodriguez et al., “Episodix”). Researchers attribute the increased motivation and popularity of serious games to a sense of engagement and self-efficacy that may not be present in other aspects of their daily lives. This ability to engross users has recently begun to be leveraged for purposes beyond entertainment in the form of gamified cognitive assessments. When two versions of tests are equally repetitive and boring, the gamified versions are reported to be more interesting and enjoyable (Lumsden et al.). Some games use virtual rewards like virtual gold coins to reward users when they correctly complete tasks. The episodic memory training program “Game Show” motivates participants through virtual gold coin rewards for correct task completion and automatically adjusts the number of geometric patterns based on player performance to maintain engagement (Groznik and Sadikov). This reward-focused design raises concerns, as users may become frustrated and abandon the game due to poor scores. However, others argue that there is no evidence for this phenomenon (Lumsden et al.).

Adjusting difficulty in gamified assessments is much easier than in traditional paper-and-pencil tests. By modifying task difficulty, the game can provide an appropriate level of challenge for users: the game should be neither too difficult nor too easy, as users would likely quit playing due to boredom in both cases (Tong and Chignell). This can be easily achieved in gamified approaches by changing parameters (Tong et al., “Tracking Cognitive Decline”), which reduces the likelihood of users abandoning the game due to frustration. In the game assessment tool Whack-a-Mole, if users' performance is poor (taking too long or inaccurate hitting motions), the game system can automatically adjust the target size to help users complete the task (Tong and Chignell). Another way to assist users in completing tasks is to provide prompts after they make a certain number of mistakes (Manera et al.). Similarly, Phan Tran et al. developed an interactive system to assist participants during the game. Participants tend to abandon the game if no help is provided, but this does not happen when they receive assistance (Groznik and Sadikov).

2. Simplified Execution

Another important benefit of gamification is that it simplifies the administration of cognitive assessments. This is achieved mainly through four aspects: shortening testing time, adapting to portable devices, automatically recording data, and assisting specific groups such as older adults in completing tasks.

Gamification can reduce the duration of traditional cognitive tasks. A review by Valladares-Rodriguez et al. (“Design process and preliminary psychometric study”) mentioned that a Spanish paper-and-pencil version of the California Verbal Learning Test (CVLT) took between 65 and 75 minutes, including the time required for examiners to score the results. However, the gamified version of the CVLT, called “Episodix,” takes only about 30 minutes, significantly shortening the testing time (Valladares-Rodriguez et al., “Episodix”). “Episodix” requires participants to play the game in a virtual scene such as a park or a street. A series of objects requiring memorization (e.g., vehicles, store signs, items carried by pedestrians) are

naturally presented. The object categories and frequency of appearance follow the CVLT reference but are replaced with more visually recognizable content. These objects are presented without intentional cues, mimicking "unintentional memory" in real life. After the walking scene ends, participants are asked to select objects they saw in the game from a set of objects that did not appear. After a certain interval (e.g., 10 minutes, during which other games can be played), players are asked to recall and select target objects again to evaluate their long-term memory retention ability, and the "delayed recall accuracy rate" and "recall improvement rate after semantic clue prompts" (e.g., whether they can recall more related objects after being prompted with "transportation tools") are recorded. Finally, users are presented with "yes or no questions" (e.g., "Have you seen a red bus?") to check if participants are guessing. The game system can automatically calculate the "hit count" and "false alarm count" (misjudging unseen objects as seen), and these data can be used to generate a CVLT score.

Many gamified assessments are designed to be compatible with portable or touch devices, allowing users to play the game in familiar environments. This increases accessibility and enables screening to reach a larger population. For example, the serious game "Kitchen and Cooking" evaluates executive functions in cooking scenarios and is compatible with tablets, enabling flexible use in both home and nursing home settings (Manera et al.). In a 2014 study, the classic game Whack-a-Mole was redesigned as a gamified assessment tool. Participants need to whack moles and avoid hitting butterflies in this game, which can be played on a mobile device. The clicking motion can be replaced by tapping the mole's image with a finger (Tong and Chignell). Allowing participants to complete tests and training at home can reduce transportation costs and the time family members spend accompanying them to hospitals.

Gamified assessments can automatically record data, be completed in a shorter time, and require little to no instructor training (Groznik and Sadikov; Tong et al., "Tracking Cognitive Decline"; Wang et al.). This makes gamified assessments cheaper and more efficient than traditional paper-and-pencil tests, and games can be integrated into routine diagnosis and treatment processes (Wang et al.). Additionally, gamification allows patients to track the progression of their cognitive function decline because they can use the tools themselves at home (Groznik and Sadikov).

The portable devices also offer the added benefit of enabling older adults—who are most affected by dementia—to complete testing easily. There are already many games specifically designed for older adults (Lumsden et al.). It is important for elderly people to feel they can control their healthcare process, and this can be easily achieved through cognitive function training in the form of games (Groznik and Sadikov). However, certain adjustments may be necessary: for example, replacing fine movements that may be difficult for them with gross movements like touching screens (Tong et al., "Tracking Cognitive Decline"). Some games also include designs such as familiar cultural backgrounds, large images, and spoken instructions to assist older adults who are unfamiliar with digital

devices (Wang et al.).

Challenges

1. Limitations in Diagnostic Accuracy

Many studies adapting traditional paper-and-pencil tests report that their games' scores are highly correlated with the original tests' scores. However, several still face the challenge of accurately capturing the specific cognitive function the original test intends to measure. All the above reasons make gamified assessments difficult to use as formal diagnostic tools due to accuracy issues.

2. Challenges in Isolating Specific Cognitive Functions

The first challenge is measuring a single cognitive function without incorporating additional ones. For example, in Tong et al. ("Tracking Cognitive Decline"), the researchers found that for older adults, their game scores were positively correlated with their MMSE scores. However, some games like Whack-a-Mole do not assess a single cognitive function. In Whack-a-Mole, participants need several different cognitive functions, such as executive function (clicking the target), inhibitory function (avoiding clicking wrong targets), and attention. In this particular study, researchers collected three types of data: Median Correct Reaction Time, False Alarm Rate, and Combined Performance Metric (a score reflecting the overall situation of Median Correct Reaction Time and False Alarm Rate). All three scores require executive functions, meaning it is difficult for all interactive games to solely measure one cognitive function other than executive function, as all results of participants' cognitive processes need to be "executed" and then reflected in the game score. In Zygouris et al., participants needed to complete a shopping task in the game. More specifically, they had to find products from a given shopping list in a supermarket, select the correct products from shelves, and enter the correct total price of all selected products. Researchers collected three types of data: time-related completion variables, navigation accuracy-related variables, and attention and executive function-related variables. Time-related variables included the average completion time of a single test and the difference in completion time between different test rounds (a total of 3 rounds), which reflect learning effects. This requires an integrated performance of visual memory (recognizing products), verbal memory (remembering shopping lists), spatial navigation (positioning in the virtual supermarket), and executive function (multitask coordination: shopping + navigation + payment). Navigation accuracy-related variables included the degree of deviation from the optimal navigation trajectory (quantified by the proportion of dwell time in "green/yellow/red zones" of the virtual space) and the number of navigation errors. This task also required verbal memory (remembering shopping lists) and executive function (multitask coordination: organizing the sequence of collecting goods from shelves). Attention and executive function-related variables included the number of operational errors in shopping tasks (e.g., selecting the wrong product, entering an incorrect payment amount). This task also required verbal memory (remembering shopping lists) and visual memory (recognizing products) (Zygouris et al.). All three measured variables involve more than one cognitive function.

Another example of measuring the target function alongside the executive function is the Panoramix Suite designed by Valladares-Rodriguez et al. (“Learning to Detect Cognitive Impairment”). This test suite includes six serious games. One of the games in the Panoramix Suite is Attentix. Researchers intended this game to measure participants’ attention abilities (continuous attention, selective attention, and response inhibition ability). In this game, participants need to replicate the color sequence shown on the screen. The screen displays a sequence of colors (e.g., “red → blue → green”) sequentially at a fixed speed. The initial sequence length is one color, and the display duration of each color is fixed. After this stimulation, the user needs to click or touch the color buttons on the screen in the previously shown sequence. Since executive function refers to the cognitive ability to coordinate thoughts and achieve a specific goal (Miller and Wallis), inhibition ability (the ability to stop incorrect actions) is also part of executive function. Inhibition ability is important for completing this task because participants need to refrain from making inappropriate moves (e.g., clicking the wrong button). The Space Code game was initially thought to be a pure tool for measuring processing speed, but later studies confirmed that it has unstable correlations with multiple dimensions such as working memory and visuospatial ability, indicating that the measurement targets of gamified tasks are easily disturbed (Lumsden et al.).

To make games more engaging for users, gamified assessments often simulate daily life tasks like shopping (Zygouris et al.) or cooking (Manera et al.), and these tasks require participants to interact with the game system. This feature causes two problems. First, daily tasks often require the integration of different cognitive functions. For example, tidying a room requires visual memory (recognizing items in the room) and executive function (multitask coordination). It is difficult to find a natural daily life task that requires only a single cognitive function. Similarly, it is challenging to identify and measure all cognitive functions used in the game (Tong et al., “Tracking Cognitive Decline”). The second problem is reliance on executive functions. One key reason games are engaging is that players can interact with the game system instead of passively receiving information. However, the interaction process itself requires intact executive functions; any executive function deficits may result in lower scores, which reflect executive function dysfunction rather than deficits in the specific cognitive function being assessed. If game designers wanted to avoid this problem, they would have to eliminate interaction, which would make the game less engaging and thus negate the advantage of games over paper-and-pencil tests.

3. Accessibility and Usability Challenges for Older Adults

Gamified assessments require participants to interact with computers, and this feature means the final results can be affected by familiarity with computer operations such as using keyboards and mice. Older adults and people with AD face difficulties using many serious games: they struggle to become familiar with game technology and tools designed for the games (e.g., Wollersheim et al.; Legouverneur et al.). Participants from focus groups and pilot experiments showed that assessment results can be influenced by participants' technical proficiency (Valladares-Rodriguez et al., “Design process and preliminary psychometric study”; “Episodix”). Using keyboards and mice may introduce

confounding factors due to varying skill

levels among participants, interfering with cognitive assessment results (Valladares-Rodriguez et al., “Design process and preliminary psychometric study”; “Episodix”). Some studies mentioned that older adults prefer touch interfaces over traditional keyboards and mice. This may be because touch interfaces are more suitable for beginners than traditional mice and keyboards (Valladares-Rodriguez et al., “Design process and preliminary psychometric study”; “Episodix”). However, older adults may still face problems even if game systems are adapted to touch devices: designing games for tablets can also be challenging, as older adults may have difficulty controlling fine motor skills, and touch devices are usually small in size (Groznik and Sadikov).

Older adults’ cognition exhibits typical aging characteristics, including shorter attention spans and slower information processing speeds (Groznik and Sadikov), but gamified assessments and training often fail to take these features into account. Regarding attention, unlike younger people, older adults do not enjoy competition or collecting virtual achievements: they feel distinctly uncomfortable with game features unrelated to task goals. Such features can distract them from completing core game tasks and may cause negative emotions such as anxiety, which can subsequently reduce their motivation to complete the assessment (Katz et al.). For example, some games have virtual badge systems and real-time ranking systems. These features may increase younger participants’ motivation, but older participants may experience cognitive burdens that make it more difficult to focus on targets and complete core tasks like memory and reasoning tasks (Groznik and Sadikov).

In terms of information processing, due to the decline in older adults’ cognitive processing efficiency, their ability to adapt to complex interface layouts, operational processes, and feedback mechanisms is weaker than that of younger people. Some games are designed to be overly complex, requiring users to learn numerous operational rules or rely on technical support to play effortlessly. Such features can be challenging for individuals with reduced information processing speeds and further reduce their likelihood of participating in and accessing the game (Groznik and Sadikov). Furthermore, game designs targeting cognitively impaired groups (such as those with mild cognitive impairment or Alzheimer's disease) have not fully considered the specific characteristics of their cognitive decline, lacking appropriate difficulty gradients and timely guidance mechanisms. This leads to these users experiencing frustration due to frequent in-game errors and even abandoning participation (Tziraki et al.). Unlike neurotypical individuals or patients with mild cognitive impairment, people with more severe cognitive impairment (such as Alzheimer's disease or moderate to severe cognitive impairment) experience more negative emotions (e.g., feelings of frustration and self-denial) when they make mistakes in games. This can severely damage their motivation and may lead them to quit the game or refuse to try again (Tziraki et al.). Making mistakes can increase feelings of self-doubt rather than allowing them to view mistakes as learning opportunities. Tziraki et al. argue that games for people with dementia should incorporate “errorless learning” designs (providing cues to guide users to

complete the game and avoid mistakes). However, most existing games lack such designs and still use “trial and error” as the learning logic. The guidance provided by existing games is generally insufficient, failing to offer adequate real-time support for individuals with cognitive impairments such as memory decline and inattention (Phan Tran et al.). Individuals with cognitive impairments are likely to get stuck in games due to forgetting rules or being confused about operational processes. Simple step prompts, visual/auditory reminders, or rule reviews can effectively help them continue the task. However, most games lack such assistance mechanisms, leading to frequent errors and difficulty completing the process independently (Groznik and Sadikov). For example, the interactive system developed by Phan Tran et al. provides operational guidance through a 3D animated virtual avatar, offering game reminders when user inactivity is detected and reviewing rules when frequent errors are detected. The results showed that this assistive design significantly reduced user abandonment rates and improved task completion quality, confirming the importance of clue prompts for individuals with cognitive impairments.

These design issues directly undermine the accurate assessment of cognitive performance. User errors may stem from design adaptation problems such as operational confusion or forgetting rules, rather than reflecting their cognitive abilities. This distorts the data collected by the game, failing to accurately reflect their cognitive function (Groznik and Sadikov; Manera et al.), and their performance is affected by low technical familiarity. Furthermore, the lack of “errorless learning” and cues makes it difficult for people with cognitive impairments to gain a sense of accomplishment in games, further reducing their participation and ultimately significantly diminishing the cognitive assessment value that games should provide (Phan Tran et al.).

4. Masking Cognitive Deficits in ADHD

The process of gamifying cognitive assessments should take into account the unique characteristics of the target population. For example, Retirement Party is a cognitive assessment tool developed on a computer platform, designed to evaluate cognitive dimensions such as executive function, impulse control, and working memory in adults with attention deficit hyperactivity disorder (ADHD) (Delisle and Braun). The game uses the background of "organizing a retirement party for a colleague" to simulate completing everyday tasks. For example, participants must logically select 30 actions related to retirement party preparation within a virtual four-day period, and click an icon at a given virtual time to simulate insulin injection. The game displays task options through a fixed grid interface, indicating the virtual duration of each action, and providing a clock icon that can be clicked at any time. Participants must complete constrained actions at a fast pace while maintaining continuous attention to the virtual time (Delisle and Braun). Individuals with ADHD have poor tolerance for repetitive, unresponsive tasks. However, this game effectively activates their cognitive motivation and compensates for their poor attention time length through clear task objectives, immediate error feedback (such as a flashing red indicator panel after a missed action or timeout), and diverse task transitions (Delisle and Braun). This environmental design allows individuals with ADHD to reduce distractions through external structural constraints, masking their attention problems.

Discussion

1. General Conclusion of Current Research

This paper has explored several studies that demonstrate gamified cognitive assessment can be used for initial screening of dementia (overall or on some cognitive functions). The Virtual Supermarket Test (VST) developed by Zygouris et al. was proved to be a reliable tool for screening subjective memory complaints (SMC) and mild cognitive impairment amongst elderly people. The Game Based Cognitive Assessment (GBCA) developed by Wang et al. and serious game Whack-a-Mole developed by Tong et al. ("Tracking Cognitive Decline") were also proved effective because of the game's high correlation with classical cognitive test MMSE. The Panoramix game suite designed by Valladares-Rodriguez et al. ("Learning to Detect Cognitive Impairment") was an effective tool in predicting delirium and cognitive decline: three of its games achieved a 100% success rate in classifying healthy individuals and AD/MCI patients (Tong et al., "A Serious Game"). Tong et al. ("A Serious Game") completed a validity verification to the cognitive assessment game adapted from Whack-a-Mole, and found out that the game had a strong correlation with Montreal Cognitive Assessment (MoCA), the Mini-Mental State Examination (MMSE), and the Confusion Assessment Method (all $p < 0.001$). This proved that this game can effectively capture the cognitive function related signal. The Skill Lab game suite contains 6 mini-game. The research team designed a prediction model for users' data and can measure 8 kinds of cognitive functions (Pedersen et al.); results showed that this model can not only tell the difference between different cognitive functions, but also evaluate the overall cognitive ability, which fits the requirement of initial screening. Furthermore, the Smartkuber game developed by Boletsis and McCallum had an overall correlation of 0.81 with the MoCA test, and six of the seven assessment dimensions of the MoCA were highly correlated with the mini-games of this game. The game showed no significant learning effect, which means that it is suitable for long-term repeated screening. This meets the need for long-term monitoring of cognitive function. The gamified tool Episodix, based on a classic cognitive assessment scale, has shown in preliminary validation its ability to assess the cognitive status of older adults, predict mild cognitive impairment and the onset of Alzheimer's disease. It can quickly collect cognitive-related data, meeting the convenience and predictive requirements of a primary screening tool (Valladares-Rodriguez et al., "Design process and preliminary psychometric study"). Episodix was psychometrically validated and demonstrates the ability to differentiate between healthy individuals, patients with mild cognitive impairment, and Alzheimer's disease patients with high accuracy. It can also quickly complete cognitive function screening in large populations, meeting the efficiency and accuracy requirements of primary screening (Valladares-Rodriguez et al., "Episodix"). The serious game "Kitchen and cooking" developed by Manera et al. was designed to assess and stimulate executive function and motor skills. The study clearly pointed out that the game can only be used as a supplement instead of a replacement to traditional assessment methods (Manera et al.).

Taken together, these studies show that gamified cognitive assessments are effective in screening cognitive decline in patients, and have many advantages that traditional tests lack. Gamifying cognitive assessment can provide reliable reference and hint for early identification of cognitive impairment, and have a high efficiency for initial screening.

However, gamified cognitive assessments cannot be used as a formal diagnosis tool yet, a point that was highlighted in many of the studies mentioned above. For example, GBCA developed by Wang et al. is defined only as a screening tool, and can only show probable cognitive dysfunction, and cannot be used as a diagnosing tool. The abnormal results are just for showing risk of having cognitive impairment and advising users to have further examination in hospital. Similarly, although VST can effectively classify MCI and SMC from healthy people, the researchers still claimed that this tool has not moved beyond research or clinical settings and still require professional examiners' help when examining (Zygouris et al.). Although the Panoramix game kit achieves 100% differentiation between healthy individuals and those with Alzheimer's disease (AD)/MCI, the researchers mentioned directly that the aim of this game suite is just early detection of dementia (Valladares-Rodriguez et al., "Learning to Detect Cognitive Impairment"). Serious games' core value is to motivate users to do the test regularly and catch signals of dysfunctions on time, not directly give the diagnosing conclusion (Groznik and Sadikov).The Episodix study clearly points out that its resolution in identifying specific cognitive impairments is insufficient. Its psychometric properties have not been fully validated, and it lacks comprehensive data support. Therefore, it cannot provide the accurate diagnostic criteria required for a definitive diagnosis (Valladares-Rodriguez et al., "Design process and preliminary psychometric study"). In summary, the core function of existing gamified cognitive screening tools is to provide initial screening and risk warnings, and they cannot be used as a standalone diagnostic tool for cognitive impairment.

2. Limitation of Current Research

There is already a large number of studies focused on different gamified assessments. However, there are still some limitations in the literature. The first important limitation is that many studies that test the effectiveness of a particular gamified assessment usually have a very small participant number, and because of sampling convenience, participants usually come from the same region. In the research conducted by Paletta et al., researchers only had 15 participants for testing MIRA and 8 participants for MMA. In the Virtual Supermarket Test experiment, the sample included a total of 95 people, including 48 healthy older adults with subjective memory complaints (SMC) and 47 patients with mild cognitive impairment (MCI) (Zygouris et al.). The samples were from the day care center of the Greek Association of Alzheimer's Disease and Related Disorders (Zygouris et al.). In the Whack-a-mole game-related experiments, the sample included 24 healthy volunteers aged 21-51 years, including 7 women and 17 men (Tong and Chignell). The samples came from

the University of Toronto (Tong and Chignell). In the Panoramix game kit-related experiment, the sample consisted of 16 elderly individuals, including 8 without cognitive impairment (HC), 5 with Alzheimer's disease (AD), and 3 with mild cognitive impairment (MCI), comprising 12 women and 4 men (Valladares-Rodriguez et al., “Learning to Detect Cognitive Impairment”). The samples were taken from local communities and related patient groups in Galicia, Spain (Valladares-Rodriguez et al., “Learning to Detect Cognitive Impairment”). In the Game-Based Cognitive Assessment experiment, the sample consisted of 124 people, including 67 patients with neurocognitive impairment (NCD) and 57 healthy controls (HCs), aged 55-90 years (Wang et al.). Samples were collected from the neurology and geriatric psychiatry outpatient departments of Tainan Hospital and local communities (Wang et al.). These experiments all had relatively small sample sizes and sampling concentrate to a relatively small region. This may add bias to their conclusions and make generalising to a general population more challenging.

In existing studies, the classification of cognitive status in experiments is simple, mostly simply dividing into healthy groups and cognitive impairment groups (MCI/AD), without further subdividing the specific subtypes and severity of cognitive impairment, which makes it difficult to meet the needs of accurate screening and differential diagnosis. In the study by Zygouris et al., the experimental group included healthy elderly individuals and those with neurocognitive impairment (MCI). In the Wang et al. study, the experimental group was divided into a Neurocognitive Impairment Group (NCD) and a healthy control group. In the Valladares-Rodriguez et al. study (“Episodix”), the experimental group included a healthy control group, MCI patients, and AD patients. In the study by Manera et al., the experimental groups were divided into a mild cognitive impairment (MCI) group and an Alzheimer's disease (AD) group. In the study by Groznik and Sadikov, the experimental groups were divided into a healthy control group (HC), a mild cognitive impairment (MCI) group, and an Alzheimer's disease (AD) group. The group separations are usually gross and can only tell the difference between neurotypicals and people with cognitive dysfunction. Only a few of them can make the distinction between a user having MCI or AD, which is not adequate.

Existing research lacks standardized clinical procedures, often remaining in laboratory settings or small-scale pilot programs. It lacks unified operational guidelines, and outcome interpretation standards, making it difficult to integrate into routine clinical practice. Furthermore, since most existing studies are single-center investigations, results are easily influenced by region, medical resources, and evaluators. Their stability and reliability require cross-regional replication and validation.

3. Directions for Further Research

In the future, research on gamified cognitive assessment and training tools needs to focus on accuracy and practicality. For different cognitive impairments, more games targeting specific disease screening need to be developed. Also, cross-device data collection standards need to be established. By unifying the operating logic and data recording specifications of different devices such as mobile phones and tablets, serious games can eliminate measurement biases caused by device differences and increase cognitive tests’

accessibility. For the elderly, interactive design should be optimized, special functions such as voice-assisted navigation, simplified touch operation logic, and enlarged interface elements to lower the technical threshold should be added to improve accessibility for elderly users. Gamified screening should be incorporated into the routine primary healthcare cognitive screening process. A simplified operation module for general practitioners should be developed, along with automated report generation. This characteristic can lower the bar of using the screening tool. When an abnormal result is obtained from gamified screening, The game should give the user a tip that they should go to hospital for further examination.

Conclusion

Overall, this review has demonstrated that gamified cognitive assessments can be an effective and valuable tool in the screening of several cognitive impairments, but at present lack the specificity to be used as a formal diagnosis tool.

Works Cited

Alzheimer's Disease International. "Dementia facts & figures." Alzheimer's Disease International, 2024, www.alzint.org/about/dementia-facts-figures/.

Boletsis, Costas, and Simon McCallum. "Smartkuber: A Serious Game for Cognitive Health Screening of Elderly Players." Games for Health Journal, vol. 5, no. 4, 2016, pp. 230–38, https://doi.org/10.1089/g4h.2015.0107.

Delisle, Josianne, and Claude M. J. Braun. "A context for normalizing impulsiveness at work for adults with attention deficit/hyperactivity disorder (combined type)." Archives of Clinical Neuropsychology, vol. 26, no. 7, 2011, pp. 602–13, https://doi.org/10.1093/arclin/acr043.

Groznik, Vida, and Aleksander Sadikov. "Gamification in Cognitive Assessment and Cognitive Training for Mild Cognitive Impairment." Augmented Reality Games II, edited by Vladimir Geroimenko, Springer Nature Switzerland, 2019, pp. 179–204, https://doi.org/10.1007/978-3-030-15620-6_8.

Katz, Benjamin, et al. "Differential effect of motivational features on training improvements in school-based cognitive training." Frontiers in Human Neuroscience, vol. 8, 2014, article 242, https://doi.org/10.3389/fnhum.2014.00242.

Legouverneur, Grégory, et al. "Wii sports, a usability study with MCI and Alzheimer's patients." Alzheimer's & Dementia, vol. 7, no. Suppl., 2011, pp. S500–S501, https://doi.org/10.1016/j.jalz.2011.05.2398.

Lumsden, Jim, et al. "Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy." JMIR Serious Games, vol. 4, no. 2, 2016, article e11, https://doi.org/10.2196/games.5888.

Manera, Valeria, et al. "'Kitchen and cooking,' a serious game for mild cognitive impairment and Alzheimer's disease: A pilot study." Frontiers in Aging Neuroscience, vol. 7, 2015, article 24, https://doi.org/10.3389/fnagi.2015.00024.

Miller, Earl K., and Jonathan D. Wallis. "Executive Function and Higher-Order

Cognition: Definition and Neural Substrates." Encyclopedia of Neuroscience, edited by Larry R. Squire, vol. 4, Academic Press, 2009, pp. 99–104.
Paletta, Lucas, et al. "Towards Decision Support with Assessment of Neuropsychological Profiles in Alzheimer's Dementia Using Playful Tablet-Based Multimodal Activation." Advances in Human Factors in Healthcare and Medical Devices, edited by Tareq Ahram et al., Springer Nature Switzerland, 2021, pp. 469–77, https://doi.org/10.1007/978-3-030-80285-1_54.
Pedersen, Martin K., et al. "Measuring Cognitive Abilities in the Wild: Validating a Population-Scale Game-Based Cognitive Assessment." Cognitive Science, vol. 47, no. 6, 2023, article e13308, https://doi.org/10.1111/cogs.13308.
Pérez-Rodríguez, Rubén, et al. "Learning to Detect Cognitive Impairment through Digital Games and Machine Learning Techniques." Methods of Information in Medicine, vol. 57, no. 4, 2018, pp. 197–207, https://doi.org/10.3414/ME17-02-0011.

Phan Tran, Minh Khue, et al. "Assistance for Older Adults in Serious Game Using an Interactive System." Games and Learning Alliance, edited by Alessandro De Gloria and Remco Veltkamp, Springer International Publishing, 2016, pp. 286–91, https://doi.org/10.1007/978-3-319-40216-1_30.
Porsteinsson, Anton P., et al. "Diagnosis of Early Alzheimer's Disease: Clinical Practice in 2021." Journal of Prevention of Alzheimer's Disease, vol. 8, no. 3, 2021, pp. 371–86, https://doi.org/10.14283/jpad.2021.23.
Tong, Tiffany, and Mark Chignell. "Developing a serious game for cognitive assessment: choosing settings and measuring performance." Proceedings of the 2014 Chinese CHI International Symposium, Association for Computing Machinery, 2014, pp. 70–79, https://doi.org/10.1145/2592235.2592246.
Tong, Tiffany, et al. "A Serious Game for Clinical Assessment of Cognitive Status: Validation Study." JMIR Serious Games, vol. 4, no. 1, 2016, article e7,https://doi.org/10.2196/games.5006.
Tong, Tiffany, et al. "Tracking Cognitive Decline with a Serious Game: Benchmarking Against the Mini-Mental State Examination." Proceedings of the 2020 HFES 64th International Annual Meeting, 2021, pp. 6–10, https://doi.org/10.1177/1071181320641002.
Tziraki, Chariklia, et al. "Designing Serious Computer Games for People With Moderate and Advanced Dementia: Interdisciplinary Theory-Driven Pilot Study." JMIR Serious Games, vol. 5, no. 3, 2017, article e16, https://doi.org/10.2196/games.6514.
Valladares-Rodriguez, Sonia, et al. "Design process and preliminary psychometric study of a video game to detect cognitive impairment in senior adults." PeerJ, vol. 5, 2017, article e3508, https://doi.org/10.7717/peerj.3508.
Valladares-Rodriguez, Sonia, et al. "Episodix: a serious game to detect cognitive impairment in senior adults. A psychometric study." PeerJ, vol. 6, 2018, article e5478, https://doi.org/10.7717/peerj.5478.
Valladares-Rodriguez, Sonia, et al. "Learning to Detect Cognitive Impairment through

Digital Games and Machine Learning Techniques." Methods of Information in Medicine, vol. 57, no. 4, 2018, pp. 197–207, https://doi.org/10.3414/ME17-02-0011.
Wang, Ching-Sen M., et al. "Using Self-Administered Game-Based Cognitive Assessment to Screen for Degenerative Dementia: A Pilot Study." Journal of Alzheimer's Disease, vol. 86, no. 3, 2022, pp. 877–90, https://doi.org/10.3233/JAD-215142.
Wollersheim, Debbie, et al. "Physical and psychosocial effects of Wii videogame use among older women." Australian Journal of Emerging Technologies and Society, vol. 8, no. 2, 2010, pp. 85–98.
World Health Organization. "Dementia." World Health Organization, 2025, www.who.int/news-room/fact-sheets/detail/dementia.
Zygouris, Stelios, et al. "Detection of Mild Cognitive Impairment in an At-Risk Group of Older Adults: Can a Novel Self-Administered Serious Game-Based Screening Test Improve Diagnostic Accuracy?" Journal of Alzheimer's Disease, vol. 78, no. 1, 2020, pp. 405–12, https://doi.org/10.3233/JAD-200880.

EcoFiber - biodegradable packaging based on oyster mushroom mycelium and filamentous algae grown in a photobioreactor

Compiled by
Assylzhan Sailau
Semyon Chistyakov
Najmudin Adikhanov
Mansur Kakharman
Rudik Ivan
Aldabergenov Amirali
Nazarbayev Intellectual School in Nura Astana
Nazarbayev Intellectual School IB

2026

Abstract

Plastic pollution is one of the most global problems on our planet. Plastic is a chemical compound of synthetic polymers that can take several centuries to decompose. Because of this peculiarity, plastic can accumulate in different parts of the world, polluting ecosystems, thus harming not only animals but also people.

To prevent further deterioration of the situation, scientists are trying to find alternatives to plastic, but all the currently proposed substitutes do not match and do not have the same qualities as plastic.

Our project is aimed at creating a biodegradable material that can replace plastic. The basis for the development is mycelium of oyster mushrooms and filamentous algae. The material is created by colonizing the substrate with mycelium. The mycelium acts as a bioclay, forming strong and flexible structures with the algae. The filamentous algae add additional strength characteristics due to their special tine structure.

For efficient algae production we use a photobioreactor. This device creates ideal conditions for algae growth by controlling light, temperature and nutrient concentrations. This approach allows us to scale the production of environmentally friendly material and reduce the cost of its manufacture.

We are confident that the material developed can replace plastic in packaging and other areas, reducing pollution and making the environment cleaner. This is an important step towards a more responsible and environmentally sustainable future.

Keywords: plastic pollution, biodegradable material, mycelium of oyster mushrooms, filamentous algae, photobioreactor.

Chapter 1

Introduction

Plastic pollution is a global and widespread problem: 7 out of the 9.2 billion tonnes of plastic produced between 1950 and 2017 have become plastic waste, causing harm to our planet. The most significant downside of plastic is that it takes a long time to decompose. Plastic takes between 20 and 500 years to decompose; it does not break down completely, but breaks down into microplastics, which subsequently end up in water and soil, threatening ecosystems, and in the human body, posing a threat to health. Every year, 19–23 million tonnes of plastic harm aquatic ecosystems, and it is also known that around 2,100 animal species have been affected by plastic. [1,2]

There are several factors contributing to the problem of plastic pollution, such as insufficient recycling and improper disposal, but these are all consequences of the mass production and consumption of plastic – over 460 million tonnes of plastic are produced worldwide every year. At the same time, alternative materials are often significantly more expensive and less convenient, which is why the scale of the problem continues to grow. [3]

Every year, 602,740 tonnes of plastic waste are generated in Kazakhstan, which constitutes the plastic pollution in our country. [4] However, there is no suitable alternative to plastic, as the material must be biodegradable in order to prevent any pollution. As the problem of plastic pollution requires an immediate solution, it is clear from the root cause of the problem that we need to develop a material that is as convenient as plastic, yet does not cause such significant damage to the environment. In this regard, the development of bio-packaging, as proposed in this project—bio-packaging based on oyster mushroom mycelium and filamentous algae, using a photobioreactor to cultivate the algae in order to enable the efficient mass production of the bio-packaging, is a highly pressing task for saving our planet

from plastic pollution.
Practical significance: the material produced serves as an alternative to plastic; unlike plastic, however, it is entirely environmentally friendly and biodegradable
Hypothesis: this research will identify a method for the rapid and efficient production of bioplastic packaging that serves as an alternative to plastic
Project objective: to create environmentally and economically viable bioplastic packaging based on filamentous algae and oyster mushroom mycelium, using an innovative algae cultivation technology developed in the form of a photobioreactor
Research objectives:
1) To analyse the growth of Pleurotus ostreatus mycelium on a sterile substrate made from filamentous algae, in comparison with traditional types of lignocellulosic raw materials.
2) To develop a photobioreactor for the efficient and accelerated production of biomass from filamentous algae.
3) To assess the growth rates of algae in the photobioreactor.
4) Conduct comparative testing of the resulting biomaterials for various properties (strength, degradability, water absorption).
5) Compare the developed product with commercial equivalents.
6) Produce a pilot batch of the developed biopackaging.

Chapter 2

Literature Review

1. Plastic Pollution as a Global Environmental Crisis

Plastic waste is one of the most severe environmental issues of today. As per UNEP [1], an entire garbage truck worth of plastic gets into aquatic systems every day, polluting lakes, rivers, and oceans across the globe. As stated by Parker [2], the amount of plastic production from the 1950s to date is estimated at 9.2 billion tonnes, out of which about 7 billion tonnes were wasted. Synthetic polymers are known for their long-term degradation, with the decomposition of plastics taking anywhere from 20 to 500 years, breaking down into microplastics, which contaminate soil, water bodies, and even organisms. As per UNEP [1], each year 19–23 million tonnes of plastics find way into aquatic systems, impacting 2,100 different species [1].

According to the report on RBC Trends [3], an ignored player in terms of pollution is the textiles sector, producing significant quantities of synthetic fibre waste. In the case of Kazakhstan, according to the United Nations Development Programme [4], the country produces around 602,740 tonnes of plastic waste annually. The issue is not confined only to developed countries; there is no better solution, so the problem is becoming even bigger.

2. Mycelium-Based Biocomposite Materials

The vegetative tissue of fungi called mycelium [5] has gained significant scientific interest lately as a renewable and biodegradable base for composite materials. Attias et al. [6] discussed various possibilities related to the use of mycelium biocomposites in design and architecture, providing examples of their potential applicability in different cases. Popyvanov and Shirokhikh [8] found out that the mycelium of basidiomycetes produces interesting structural biopolymers where hyphal strands naturally bind together the organic substrates in solid blocks. This research was expanded by Silverman, Cao and Cobb [9] who successfully used mycelium bioblocks in footwear manufacturing proving their sufficient durability and malleability. Joshi, Meher and Poluri [10] studied the process of mycelium

bioblock formation on the basis of lignocellulose agricultural waste and revealed that mycelium could work as a natural adhesive without using any artificial binders.

3. Filamentous Algae as a Reinforcing Component

Filamentous green algae belonging to the genus Cladophora [7] have highly crystalline cellulose cell walls arranged in a multilayered structure, making them quite stiff in terms of tension resistance for their biological mass. At the same time, unrestricted proliferation of Cladophora can cause HABs in eutrophic ecosystems, negatively affecting oxygen concentrations and species diversity [11]. Thus, the cultivation of Cladophora algae in a controlled environment provides an opportunity to achieve two objectives simultaneously: obtaining a useful material to reinforce composites and minimizing ecological harm from excessive proliferation of algae in nature.

4. Photobioreactor Technology for Scalable Algae Production

Photobioreactor (PBR) is a closed growth system capable of controlling light, temperature, CO2, and nutrients for photosynthetic organisms [12]. Chemodanov et al. [13] showed the possibility of commercial-scale production of macroalgae using PBRs for biorefineries. The analytical characterisation of the composite is conducted via FTIR spectroscopy according to the methodology proposed by Tarasevich [15].

5. Synthesis and Research Gap

From the above review of the literature, it can be concluded that:
- plastic pollution is a pressing, measurable problem facing the globe today [1-4]
- mycelium composites are an established scientific category of biodegradable materials [6, 8-10];
- filamentous algae contribute to reinforcement owing to the crystalline structure of cellulose contained therein [7, 11], and;
- technology allows for algal mass production in photobioreactors [12, 13].

However, to date, there have been no attempts to integrate Cladophora algae cultivated in a PBR and oyster mushroom mycelium (Pleurotus ostreatus) into a composite material for use as an alternative to plastic.

Chapter 3

Research Methods

In this chapter, we provide information on how the experiments were performed for: 1) the preparation and inoculation of substrates with the mycelium of oyster mushrooms; 2) the characterization of the mechanical and chemical properties of the generated biomaterials; and 3) the construction, assembly, and functioning of the photo bioreactor for filamentous algae. All experiments were performed at the Plant Biotechnology Laboratory of the Department of Biology, Plant Protection and Quarantine, KATIU.

3.1 Biological Materials and Equipment

The biological materials utilized during this research included the following:

- Mother culture of oyster mushroom (Pleurotus ostreatus), strain K-80
- Algal species Cladophora glomerata, isolated from a pond at Astana botanical garden

• Lignocellulose materials (six in total): Cladophora glomerata, wood shavings (mix of oak, beech and ash), wheat straw, cardboard cuttings, sunflower husks, and cotton fibers

Some of the important equipment that was used includes: autoclave VK-35 for substrate sterilization; laminar flow biosafety cabinet (Lamsystems BAV p01 "LaminarS"-1.2) for inoculation [14]; thermostat (manufacturer Kazakh Medical Equipment Plant); drying chamber Snol; balances; IR spectrum analyzer (FTIR spectrometer, IRSpirit-X series); spring balance bezmen for mechanical characteristics determination; photobioreactor.

3.2 Algae Collection and Preparation

The Cladophora glomerata samples were obtained from the pond located in the Astana Botanical Garden. The species identification was performed using light microscope through the comparison of the structure of cell walls and positioning of chloroplasts with reference information [7]. Obtained samples were washed to eliminate any contaminants followed by drying under normal room conditions for 24 hours on the pieces of parchment paper to eliminate extra water.

3.3 Substrate Preparation and Mycelium Inoculation

Six substrate types were prepared for comparative analysis, with three replicate jars (80 ml glass) per substrate. The moisture ratios used are presented in Table 1.

Table 1. Substrate compositions used for mycelium colonisation trials

Substrate	Dry mass (g)	Distilled water (g)
Cladophora glomerata	10	10
Sawdust (oak/beech/ash)	10	10
Wheat straw	8	8
Shredded cardboard	10	10
Sunflower husk	12	12
Cotton fibre	14	14

Each jar was sealed with a foam plug in the lid to allow gas exchange while preventing contamination. Jars were sterilised in an autoclave at 123°C for 30 minutes. After cooling, inoculation with P. ostreatus mother culture (strain K-80) was performed inside the laminar airflow cabinet following standard biosafety protocols [14]. Inoculated jars were incubated in a thermostat at 24–26°C for 13 days. Mycelial growth depth (cm) was recorded on days 3, 7, 10, and 13 post-inoculation across all three replicates per substrate.

3.4 Biocomposite Fabrication

The substrates fully colonized were converted to planar biocomposites. The fully colonized substrates

were compressed in a petri dish mold using the flat iron plates to uniformly apply pressure using 10 kilograms of pressure for seven days, allowing further mycelial growth and fixing of the structure. Dehydration of the final product took place in a drying chamber at 100°C for 12 hours to stop further mycelial growth and geometry fixation. This process resulted in three different biocomposites, namely, algae-mycelium, cotton-mycelium, and straw-mycelium (refer to Appendix E).

3.5 Characterisation of Biocomposite Properties

3.5.1 Mass. Each specimen was weighed on analytical balances. Results from three replicates per material were averaged.

3.5.2 Chemical structure (FTIR). Chemical composition was analysed by FTIR spectroscopy following the pressed-disc method described by Tarasevich [15], using a 1:10 sample-to-KBr ratio. Spectra were acquired on the IRSpirit-X Series FTIR spectrophotometer. Full spectra for all three materials are provided in Appendix G.

3.5.3 Tensile strength. A spring balance was used to measure the force at failure. Specimens were perforated 1.1 cm from the edge (3 holes), a hook was attached, and the balance was pulled horizontally while the opposite end was held fixed. Three trials per material were performed; tests were video-recorded for accuracy.

3.5.4 Flexural strength. The same spring balance was used. The specimen was clamped so that 3.5 cm extended beyond the table edge, a weight held the fixed end, and the balance was pulled vertically downward until fracture. Three trials per material were conducted.

3.6 Photobioreactor Design and Assembly

Following a study of tubular and flat-panel designs of photobioreactors [12], a unique design that incorporates the benefits of both designs was designed in that it provides efficient circulation while exposing the algal biomass to sufficient light. A 3D model of the photobioreactor was created before the assembly (see Appendix A).

In its final form, the photobioreactor contains two tanks of 15 liters each that are linked by (i) a peripheral tube that encircles both tanks and (ii) a direct link between the two tanks. Submersible pumps are used to circulate the growth medium in the system, ensuring consistent exposure to light. The illumination is provided through an RGB LED Strip within 400-700 nm blue spectrum (see Appendix A).

The control system comprises the ESP32 WROOM microcontroller board along with (i) a carbon dioxide sensor (MQ-135), (ii) humidity sensor (DHT22), (iii) light intensity sensor (GY-30), and a two-channel relay module for controlling the pump activity. Remote control capabilities are achieved using a custom-built Telegram bot using Arduino IDE (see Appendix D and C).

3.7 Algae Cultivation and Growth Measurement

The initial biomass of algae (47.07 g) was equally distributed into three flasks; two of which were photobioreactors (each containing 15.69 g of algae) while one flask served as the control with 15.69 g of algae in 500 ml of pure distilled water without any nutrients. The biomass was measured every day through seven days starting from 14 October 2024, and the daily increase in mass (g/day) of each vessel was obtained.

Chapter 4

Results and Discussion

4.1. Efficiency of the Developed Photobioreactor and Algae Cultivation

The photobioreactor, which was automatically controlled by the ESP32 controller, provided optimal conditions for the cultivation of filamentous algae (Cladophora glomerata) by adjusting such parameters as light intensity (using GY-30), humidity/temperature (using DHT22), and CO2 concentration (MQ-135).

The cultivation experiment started on October 14, 2024, and showed significant differences in the amount of accumulated biomass for the experimental photobioreactor vessels and control. The initial biomass weight was 15.69 g for all vessels.

The results of 7-day monitoring include the following changes in biomass:

Photobioreactor Vessel №1 and №2: The biomass steadily and exponentially increased during seven days since the automation and circulation through submersible pumps and the use of RGB LEDs (400–700 nm) helped avoid biomass stagnation and increase the efficiency of photosynthesis. Thus, the biomass weight increased significantly by the end of the day (for detailed biomass growth rate see Table 1, Appendix F).

Control vessel (500 ml distilled water without nutrients): No biomass was produced initially; then the process stabilized and resulted in some degradation due to the lack of nutrient medium and light illumination.

Daily increment of biomass proves that two tanks' system with circulation of liquid provides enough lighting and ventilation to increase production of biomass needed for biodegradable packaging.

4.2. Synthesis and Structural Characteristics of Mycelium-Based Composites

The stability of the designed EcoFiber is provided by the biological relationship between the mycelium of the oyster mushroom (Pleurotus ostreatus) and the substrate used. There have been evaluated three types of substrates in the conducted research:

1. Composite of straw and mycelium;
1. Composite of cotton and mycelium;
1. Composite of filamentous algae (Cladophora glomerata) and mycelium.

From the visual and mechanical evaluation results, it can be concluded that each composite sample had different structural characteristics (Appendix D, Image 61):

- Straw composite – made a rather rigid but brittle structure due to the coarse nature of the fibers; the straw did not provide a full colonization of its fibers by the mycelium microstructures.
- Cotton composite – gave a flexible and light-weight material without required compressive resistance and structure density for packaging applications.
- Algae composite (EcoFiber) – was the most optimal in terms of structure density and homogeneity. Unique tines of filamentous algae allowed for effective interaction with the fungi. Mycelium served as an excellent bioclay, bonding algae fibers together.

4.3. FTIR Spectroscopy Analysis

In order to examine the bonding nature and molecular interaction in the composite material, Fourier-transform infrared (FTIR) spectroscopy was carried out. FTIR spectra of the initial compounds were analyzed and compared to the FTIR spectra of the resultant composites (full FTIR graphs are included in Appendix H).

The following observations have been made regarding the functional groups responsible for the stabilization of the EcoFiber matrix:
- The broad peak in the range of 3200-3400 cm^{-1} is characteristic of stretching vibrations of hydroxyl (-OH) groups, which suggests that the composite contains cellulose extracted from algae and glucans from the fungus walls.
- Peaks in the range of 1630-1650 cm^{-1} represent Amide I bands; this indicates the existence of chitin in mycelium, which serves as a reinforcing agent.
- Intensified peaks in the fingerprint region (1000-1150 cm^{-1}) are characteristic of hydrogen bonds formation between algal cellulose fibers and fungal chitin-glucan complex.
This cross-linking is one of the factors that results in the improved mechanical characteristics of the algae-mycelium composite in comparison with conventional agricultural wastes.

4.4 Discussion

The findings clearly indicate that by merging filamentous algae and fungi mycelia, an alternative to synthetic polymers has been found. Regular bioplastics, such as corn starch or PLA, require extensive agricultural land resources for manufacturing. The EcoFiber design is based on a closed loop photobioreactor which is not dependent on any agricultural lands.
A novel ESP32 control system resolves the main issue connected with algae growing, namely, sedimentation and uneven illumination of dense cultures. Monitoring the culture via the dedicated Telegram bot has helped to maintain biomass production.
As regards materials science, the combination of Cladophora glomerata and Pleurotus ostreatus is indeed a bio-composite. Mycelia produce natural enzymes that make some modifications in the surface of the substrate, thus promoting the adhesion. Given the fact that filamentous algae are characterized by a high aspect ratio and strength, they function as fiberglass fibers while mycelia serve as a resin, making the algae-mycelium mixture much more uniform compared to the samples made from straw and cotton.

Chapter 5

Conclusion

The study showed that development of the eco-friendly packing material, called EcoFiber, using the oyster mushroom mycelium and filamentous algae under laboratory conditions is a feasible process.
Based on the results of the current work, the following major conclusions can be made:

1. Performance of Photobioreactor: The photobioreactor system, which combines two tanks with ESP32 control system and pumps, provides optimal conditions for the growth of Cladophora glomerata. Spectral RGB illumination and pump-driven circulation produce fast and steady daily biomass growth compared to conventional static cultivation.
1. Optimization of Material: Out of all tested substrates (hay, cotton, and algae), the best results were achieved when mixing filamentous algae and mycelium in the form of composite material. Intertwining of fungal hyphae and structures of the algae provides dense matrix structure.
1. Molecular-Level Study: FTIR analysis revealed strong molecular interactions and hydrogen bond formation within the complex of algal cellulose and fungal chitin-glucan system.

Future Prospects and Recommendations:
Future versions of this project will involve determining precise mechanical parameters such as tensile strength, water hydrophobicity, and precise rate of decomposition when subjected to different soil conditions. In addition, by improving the control loop of the ESP32 system to incorporate automatic delivery of nutrients, the cultivation system can be scaled for commercial purposes and become a sustainable alternative to artificial plastic products.

References

1. United Nations Environment Programme. (2024). *Plastic pollution*. https://www.unep.org/plastic-pollution#:~:text=Every%20day%2C%20the%20equivalent%20of,polluting%20lakes%2C%20rivers%20and%20seas
2. Parker, L. (2024). *Explaining the global plastic pollution crisis*. National Geographic. https://www.nationalgeographic.com/environment/article/plastic-pollution
3. RBC Trends. (2024). *Uchenye uznali, skol'ko plastikovykh otkhodov proizvodit shveynaya industriya* [Scientists found out how much plastic waste the garment industry produces]. https://trends.rbc.ru/trends/green/66c73bb49a79473907d1ad72
4. United Nations Development Programme Kazakhstan. (2023). *Boy plastiku!* [Fight against plastic!]. https://www.undp.org/ru/kazakhstan/stories/boy-plastiku
5. Wikipedia. (2023). *Mitseliy* [Mycelium]. Wikipedia, The Free Encyclopedia. https://ru.wikipedia.org/wiki/%D0%9C%D0%B8%D1%86%D0%B5%D0%BB%D0%B8%D0%B9
6. Attias, N., Danai, O., Ezov, N., Tarazi, E., & Grobman, J. Y. (2017). *Developing novel applications of mycelium based biocomposite materials for design and architecture*. Paper presented at the Conference on Building with Bio-Based Materials: Best Practice and Performance Specification, Zagreb. https://www.researchgate.net/publication/319901570_Developing_novel_applications_of_mycelium_based_bio-composite_materials_for_design_and_architecture
7. Wikipedia. (2024). *Kladofora* [Cladophora]. Wikipedia, The Free Encyclopedia. https://ru.wikipedia.org/wiki/%D0%9A%D0%BB%D0%B0%D0%B4%D0%BE%D1%84%D0%BE%D1%80%D0%B0
8. Popyvanov, D. V., & Shirokikh, A. A. (2021). Sozdanie biopolimernykh materialov na osnove mitseliya bazidial'nykh gribov [Creation of biopolymer materials based on the mycelium of basidiomycetes]. In *Ekologicheskaya bezopasnost' v usloviyakh antropogennoy transformatsii prirodnoy sredy: sbornik materialov vserossiyskoy shkoly-seminara* (pp. 413–416). Perm State University.
9. Silverman, J., Cao, H., & Cobb, K. (2020). Development of mushroom mycelium composites for footwear products. *Clothing and Textiles Research Journal*, *38*(2), 119–133. https://doi.org/10.1177/0887302x19890006
10. Joshi, K., Meher, M. K., & Poluri, K. M. (2020). Fabrication and characterization of bioblocks from agricultural waste using fungal mycelium for renewable and sustainable applications. *ACS Applied Bio Materials*, *3*(4), 1884–1892. https://doi.org/10.1021/acsabm.9b01047
11. Burkholder, J. M. (2009). Harmful algal blooms. In *Encyclopedia of Inland Waters* (pp. 264–285). Elsevier. https://doi.org/10.1007/springerreference_308797
12. Labinstruments LLC. (2022). *Fotobioreaktor: chto eto, vidy i sfery primeneniya* [Photobioreactor: What it is, types and areas of application]. https://labinstruments.ru/stati/fotobioreaktor
13. Chemodanov, A., Robin, A., & Golberg, A. (2017). Design of marine macroalgae photobioreactor integrated into building to support seagriculture for biorefinery and bioeconomy. *Bioresource Technology*, *241*, 1084–1093. https://doi.org/10.1016/j.biortech.2017.06.061
14. Thermo Fisher Scientific. (n.d.). *Instruktsiya po ekspluatatsii Laminarnyy boks biologicheskoy bezopasnosti MSC-Advantage*™ [Operating instructions Biological safety cabinet MSC-Advantage™]. Thermo Electron LED GmbH. https://assets.thermofisher.com/TFS-Assets/LPD/manuals/MSCAdvantage_50163668_RevD.pdf
15. Tarasevich, B. N. (2012). *Osnovy IK spektroskopii s preobrazovaniem Fur'e. Podgotovka prob v IK spektroskopii: Posobie k spetspraktikumu* [Fundamentals of FTIR spectroscopy. Sample preparation

in IR spectroscopy: A manual for a special workshop]. Lomonosov Moscow State University.

Appendix A
3D Model and Technical Specifications of the Photobioreactor

Figure A1: 3D model of the photobioreactor

A B

Figure A2: assembled photobioreactor
Appendix B
Microscopic Analysis of *Cladophora glomerata*

A B
Figure B1: Cladophora glomerata morphology under 40x magnification (A) and 100x magnification (B)
Appendix C
Automation Circuit and ESP32 Microcontroller Architecture

Figure C1: Wiring diagram for connecting sensors (MQ-135, DHT22, GY-30) and a relay to the ESP32 board**Appendix D**
Arduino IDE Source Code for Photobioreactor Automation

Figure D1: Code in the Arduino IDE for controlling electronics via a Telegram bot**Appendix E**
Comparative Morphological and Physical Analysis of Cultivated Composites

Figure E1: a biomaterial based on cotton and mycelium

Figure E2: a biomaterial based on straw and mycelium

Figure E3: a biomaterial based on filamentous algae and mycelium

Table E1: masses of biomaterials

Copy No.	Straw-based (grams)	Cotton-based (grams)	Based on filamentous algae (grams)
1	6,5919	10,5409	8,5319
2	6,5920	10,5308	6,5857
3	6,5995	10,6734	7,6490
Averag	6,5944	10,5817	7,5888

e mass			

Table E2: Results of the tensile strength test

Biomaterial	Test 1 (H)	Test 2 (H)	Test 3 (H)	Average value (H)
Based on filamentous algae	0,491	3,532	6,622	3,548
Made from straw	2,354	1,472	1,275	1,700
Made from cotton	1,766	1,717	2,305	1,929

Table E3: Results of the bending strength test

Biomaterial	Test 1 (H)	Test 2 (H)	Test 3 (H)	Average value (H)
Algae-based	10,938	6,916	9,418	9,091
Made from straw	4,905	11,282	1,619	5,935
Made from cotton	2,452	2,649	2,158	2,420

Appendix F

Daily Biomass Growth Dynamics

Table F1: Algal growth in different environments

Days	No. 1 photobioreactor vessel (grams)	No. 2 photobioreactor vessel (grams)	A container holding 500 ml of water (grams)

15.10.2024	15,68	16,24	15,69
16.10.2024	16,00	16,74	15,69
17.10.2024	16,89	18,04	15,69
18.10.2024	18,25	19,86	15,69
19.10.2024	18,96	20,30	15,69

20.10.2024	20,08	20,44	15,69
21.10.2024	20,83	21,71	15,69

Appendix G
FTIR Spectroscopy Graphs

A

B

C
Figure G1: Graphs (spectrograms) of results from an IR Fourier-transform spectrometer; (A) results of the biomaterial based on straw and mycelium; (B) results of the cotton- and mycelium-based biomaterial; (C) results of the biomaterial based on filamentous algae and mycelium

Can a Dual-Reporter Senolytic Biosensor Visualize and Eliminate Senescent Cells Driving Tau and Amyloid-Beta Pathology in Alzheimer's Disease? By Alinur Zaitkhan

Abstract

Alzheimer's Disease (AD), affecting 50 million globally, is a neurodegenerative disease (ND) characterized by cognitive decline and high memory loss, imposing massive financial and psychological burdens is expected to triple by 2050 without intervention (Peng et al., 2023). Currently, treatments toward AD are ineffective long-term, targeting anti-tau/amyloid for temporary symptom benefits. Here, we propose a dual-report system combining p16 driven mCherry fluorescent reporter with the CFP/YFP-FRET tau biosensor in a *INK-ATTAC* transgene to identify, locate, and eliminate senescent cells within AD brain organoids in real time that drive tau and amyloid-beta (Aβ) deposition. The final design operates inside cells (Image 1.1).

Introduction

AD is one of the leading causes of cognitive decline and death in diseases, impacting millions globally. 60-80% of dementia cases are AD. The projected burden of AD in the US alone is predicted to double from 6.9 million to 13.8 million by 2060 and over $360 billion annually is spent on dementia-related care (Alzheimer's Association, 2024). While AD is the most common ND, its pathophysiological basis involves abnormal deposition of Aβ and tau protein phosphorylation. No current drugs have been able to fully reverse AD, due to tau and Aβ damage appearing 10-20 years too late after AD has become irreversible(Peng et al., 2023). While current drugs have been focusing on directly targeting anti-tau and anti-amyloid, senescent cell marker *p16INK4a* (*p16*) has been proved to be a significant contributor toward both tau-dependent neurodegeneration (Bussian et al., 2018) and Aβ accumulating diseases(Zhang et al., 2019). However, current senolytic treatment, termination of senescent cells, haven't yet pinpointed which senescent cells need to be removed and where they are(Gonzales et al., 2023). This proposal aims to develop a dual-reporter system that focuses on finding and eliminating certain senescent cells before the onset of AD via a p16-driven mCherry fluorescent reporter protein exclusive to p16-positive senescent cells, a CFP/YFP-FRET biosensor producing signal by CFP and YFP interacting with each other whenever tau starts aggregating, and a modified *INK-ATTAC* model holding FKBP-Casp8 protein that eliminates the certain senescent cells.

Literature Review

Current methods to stop or cure AD have an overall 99.6% failure rate that lead to temporary memory boosts, symptomatic treatment, or are ineffective (Djurovic-Topalovic et al., 2025). However, senolytic therapy is the most promising lead toward effective treatment (Riessland & Orr, 2023). Cellular senescence drives AD due to leaking senescence-associated secretory phenotype (SASP), contributing to Aβ and tau deposition(Wang et al., 2024). Consequently, Baker et al.'s (2011) previously developed transgene *INK-ATTAC*, utilizing a biomarker for senescence, *p16*, to successfully induce apoptosis onto p16-positive senescent

cells using an FKBP-Casp8 fusion protein with a bind-site space for AP20187. Ogrodnik et al. (2021) used the model to conclude senescent cell clearance alleviates brain inflammation and age-related cognitive impairment. However, Baker et al. 's (2011) transgene cannot display where the certain cells are, if the elimination post-clearance was completely successful, and how it affected the deposition of tau tangles and Aβ. To address this, Haston et al. 's (2023) mCherry, a protein producing a red fluorescent signal in p16-positive cells, successfully pinpointed and tracked senescent cells in living mice using a *p16Ink4a* promoter. Moreover, Holmes et al. (2014) engineered a biosensor using CFP and YFP that produces a FRET signal when tau starts to aggregate, months before traditional methods could detect pathology. Maina, Smet-Nocca, & Bitan (2024) further established the FRET signal to evaluate through integrated FRET density (FRET-positive cell count multiplied by signal intensity), providing a precise quantitative amount of tau aggregating across the area. Using the existing advances, an efficient dual-fluorescent biosensor senolytic of p16-positive cells and tau aggregation is possible.

Methodology

This study will create a dual-fluorescent senolytic biosensor able to eliminate and monitor senescent cells driving tau and Aβ accumulation. The system integrates three components: *INK-ATTAC* transgene, mCherry protein, and CFP/YFP biosensor. Our study aims to: Construct a validated design, map senescent cells alongside monitoring tau aggregation, and administer AP20187 for selective senescent cell elimination to assess tau response.

Senolytic Construction & Design

Prior to implementation into organoid, each component must be validated independently by lentiviral transductions (genetic delivery via virus) into HEK293T (human test cells) for functionality and restricted specificity confirmation.

Building on Baker et al. ’s (2011) *INK-ATTAC* transgene, a *p16Ink4a* promoter will be coupled to the myristoylated FKBP-Casp8 fusion alongside a p16-driven mCherry fluorescence reporter (Haston et al., 2023; Liu et al., 2019) through a lentiviral vector. CFP/YFP-FRET biosensor is introduced via a separate lentiviral vector to monitor real-time responses to tau deposition, quantified through FRET density measurement when tau aggregates (Holmes et al., 2014; Maina, Nocca, & Bitan, 2024). mCherry is chosen over EGFP as a fluorescent reporter to locate senescent cells due to its red signal ensuring distinct color spectrum from the cyan and yellow CFP/YFP-FRET signals – allowing dual-channel visualizations without interference. Utilizing public softwares (e.g. Benchling) to construct an *in silico* design and synthesize through Twist Bioscience, the completed dual-report biosensing senolytic will be deposited into Addgene, allowing global researchers to build upon the system. Healthy p16-negative HEK293T cells will be used to confirm neither mCherry nor Caspase 8 activate in non-senescent cells, while HEK293T cells will be H_2O_2-treated to force senescence and validate activation. Outcome will result in

mCherry signal active in senescent HEK293T cells, CFP/YFP FRET signal activating upon introduced tau seeds, and FKBP-Casp8 allowing for a selective and contained elimination – confirmed by the disappearance of a mCherry signal.

Brain Organoid Implementation

Prior to clinical trials, the dual-system will be introduced to iPSC brain organoids, as they replicate tau and senescent astrocytes and microglia, which mouse models cannot (Abud et al., 2027; Mayhew & Singhania, 2022; Fernandes et al., 2024; Aguado et al., 2023). The HEK293T validated construct will be introduced to the organoids through lentiviral transduction and mCherry-positive senescent cells will be mapped through confocal imaging using red signals across the tissue. Post-mapping, AP20187 is introduced into the organoid's culture media and binds to FKBP-Casp8 components, triggering apoptosis toward mCherry detected senescent cells through FKBP-Casp8 dimerization (Baker et al., 2011; Bussian et al., 2018). Simultaneously, CFP/YFP-FRET biosensor will detect and assess tau aggregation using FRET density measurement to compare pre and post AP20187 administration. mCherry fluorescent signals will cease as confirmation of elimination, while reduced FRET density will confirm a significant decrease of tau aggregation in response to senescent cell clearance.

Data Analysis

Data analysis results will be analyzed for mCherry signal intensity reduction per area through signal pinpointing, changes in tau aggregation using FRET density, spatial proximity (measurement of distance between mCherry-positive clusters and high FRET density zones) to evaluate senescent cell's contribution to tau damage sites. Comparing the tests across three organoids will evaluate mCherry and FRET density pre and post AP20187 being <5% chance. Overall data would provide awareness on the construct's viability, and benefits.

Conclusion

The proposal details a dual-report senolytic tool utilizing an INK-ATTAC transgene with P16INK4A-driven mCherry and CFP/YFP-FRET biosensors. If viable, this system confirms senescent cells drive tau and Aβ accumulation, providing a foundation for future senolytic use AD treatment. Treating this specific function could track and halt the spread of AD

ADDENDUM

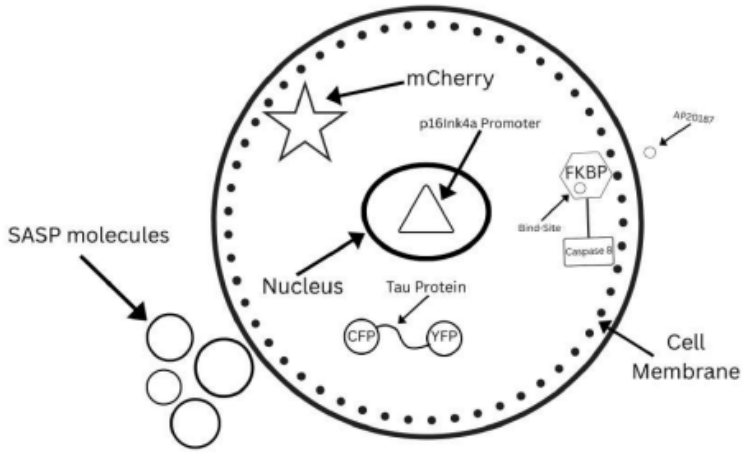


Image 1.1: Simple visualization of components inside of a p-16 positive senescent

Works Cited

1.Alzheimer’s Association. (2024, April 30). 2024 Alzheimer's disease facts and figures. Alzheimer’s & Dementia, 20(5), 3708-3821. https://doi.org/10.1002/alz.13809 2.Abud, E. M., Ramirez, R. N., Martinez, E. S., Healy, L. M., Nguyen, C. H. H., Newman, S. A., Yeromin, A. V., Scarfone, V. M., Marsh, S. E., Fimbres, C., Caraway, C. A., Fote, G. M., Madany, A. M., Agrawal, A., Kayed, R., Gylys, K. H., Cahalan, M. D., Cummings, B. J., Antel, J. P., Mortazavi, A., Carson, M. J., Poon, W. W., & Blurton-Jones. (2017, April 19). iPSC-Derived Human Microglia-like Cells to Study Neurological Diseases. Neuron, 94(2), 278-293. https://doi.org/10.1016/j.neuron.2017.03.042

3.Aguado, J., Amarilla, A. A., Taherian Fard, A., Albornoz, E. A., Tyshkovskiy, A., Schwabenland, M., Chaggar, H. K., Modhiran, N., Gómez-Inclán, C., Javed, I., Baradar, A. A., Liang, B., Peng, L., Dharmaratne, M., Pietrogrande, G., Padmanabhan, P., Freney, M. E., Parry, R., Sng, J. D. J., Isaacs, A., Khromykh, A. A., Valenzuela Nieto, G., Rojas-Fernandez, A., Davis, T. P., Prinz, M., Bengsch, B., Gladyshev, V. N., Woodruff, T. M., Mar, J. C., Watterson, D., & Wolvetang, E. J. (2023, November 13). Senolytic therapy alleviates physiological human brain aging and COVID-19 neuropathology. Nature Aging

3, 1561–1575. https://doi.org/10.1038/s43587-023-00519-6
4.Bussian, T. J., Aziz, A., Meyer, C. F., Swenson, B. L., van Deursen, J. M., & Baker, D. J. (2018, September 19). Clearance of senescent glial cells prevents tau-dependent pathology and cognitive decline. Nature, 562(7728), 578–582. https://doi.org/10.1038/s41586-018-0543-y
5.Baker, D. J., Wijshake, T., Tchkonia, T., LeBrasseur, N. K., Childs, B. G., van de Sluis, B., Kirkland, J. L., & van Deursen, J. M. (2011, November 2). Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. Nature, 479(7372), 232–236. https://doi.org/10.1038/nature10600
6.Djurovic-Topalovic, A., Horgan, N. G., & Fortin, J. S. (2025, June 18). Updates on Tau-Related Drug Targets and Potential Disease-Modifying Therapies for Alzheimer Disease. The American Journal of Pathology, 195(11), 1975-1987. https://doi.org/10.1016/j.ajpath.2025.05.020
7.Fernandes, S., Revanna, J., Pratt, J., Hayes, N., Marchetto, M. C., & Hage, F. H. (2024, July 31). Modeling Alzheimer's disease using human cell derived brain organoids and 3D models. Frontiers in Neuroscience, 18. https://doi.org/10.3389/fnins.2024.1434945
8.Hatson, S., Gonzalez-Gualda, E., Morsli, S., Ge, J., Reen, V., Calderwood, A., Moutsopoulos, I., Panousopoulos, L., Deletic, P., Carreno, G., Guiho, R., Manshaei, S., Gonzalez-Meljem, J. M., Lim, H. Y., Simpson, D. J., Birch, J., Pallikonda, H. A., Chandra, T., Macais, D., Doherty, G. J., Rassl, D. M., Rintoul, R. C., Signore, M., Mohorianu, I., Akbar, A. N., Gil, J., Muñoz-Espín, D., & Martinez-Barbera, J. P. (2023, July 10). Clearance of senescent macrophages ameliorates tumorigenesis in KRAS-driven lung cancer. Cancer Cell, 41(7), 1242-1260.e6. https://doi.org/10.1016/j.ccell.2023.05.004
9.Holmes, B. B., Furman, J. L., Mahan, J. E., Yamasaki, T. R., Mirbaha, H., Eades, W. C., Belaygorod, L., Cairns, N. J., Holtzman, D. M., & Diamond, M. I. (2014, September 26). Proteopathic tau seeding predicts tauopathy in vivo. Proceedings of the National

Academy of Sciences of the United States of America, 111(41), E4376-E4385. https://doi.org/10.1073/pnas.1411649111
10.Liu, J., Souroullas, G. P., Diekman, B. O., Krishnamurthy, J., Hall, B. M., Sorrentino, J. A., Parker, J. S., Sessions, G. A., Gudkov, A. V., & Sharpless, N. E. (2019, January 25). Cells exhibiting strong p16INK4a promoter activation in vivo display features of senescence. Proceedings of the National Academy of Sciences of the United States of America, 116(7), 2603–2611. https://doi.org/10.1073/pnas.1818313116
11.Mayhew, C. N., & Singhania, R. (2023, March 17). A review of protocols for brain organoids and applications for disease modeling. STAR Protocols, 4(1), 101860. https://doi.org/10.1016/j.xpro.2022.101860
12.Maina, K. N., Smet-Nocca, S., & Bitan, G. (2022, October 31). Using FRET-Based Biosensor Cells to Study the Seeding Activity of Tau and α-Synuclein. Protein Aggregation, 125-145. https://doi.org/10.1007/978-1-0716-2597-2_10
13.Ogrodnik, M., Evans, S. A., Fielder, E., Victorelli, S., Kruger, P., Salmonowicz, H.,

Weigand, B. M., Patel, A. D., Pirtskhalava, T., Inman, C. L., Johnson, K. O., Dickinson, S. L., Rocha, A., Schafer, M. J., Zhu, Y., Allison, D. B., von Zglinicki, T., Lebrasseur, N. K., Tchkonia, T., Neretti, N., Passos, J. F., Kirkland, J. L., & Jurk, D. (2021, January 20). Whole-body senescent cell clearance alleviates age-related brain inflammation and cognitive impairment in mice. Aging Cell, 20(2). https://doi.org/10.1111/acel.13296

14.Orr, M. E, Gonzales, M. M., Garvarino, V. R., Kauts, T., Palavicini, J. P., Lopez-Cruzan, M., Dehkordi, S. K., Mathews, J., Zare, H., Xu, P., Zhang, B., Franklin, C., Habes, M., Craft, S., Peterson, R. C., Tchkonia, T., Kirkland, J., Salardini, A., Seshadri, S., & Musi, N. (2023, September 7). Senolytic therapy to modulate the progression of Alzheimer's Disease (SToMP-AD) – Outcomes from the first clinical trial of senolytic therapy for Alzheimer's disease. Research Square. https://doi.org/10.21203/rs.3.rs-2809973/v1

15.Peng, Y., Jin, H., Xue, Y., Chen, Q., Yao, S., Du, M., & Liu, S. (2023, August 3). Current and future therapeutic strategies for Alzheimer's disease: an overview of drug development bottlenecks. Frontiers in Aging Neuroscience, 15. https://doi.org/10.3389/fnagi.2023.1206572

16.Riessland, M., & Orr, M. E. (2023, October 24). Translating the Biology of Aging into New Therapeutics for Alzheimer's Disease: Senolytics. The Journal of Prevention of Alzheimer's Disease. https://doi.org/10.14283/jpad.2023.104

17.Wang, Y., Kuca, K., You, L., Nepovimova, E., Heger, Z., Valko, M., Adam, V., Wu, Q., & Jomova, K. (2024, May 15). The role of cellular senescence in neurodegenerative diseases. Archives of Toxicology. https://doi.org/10.1007/s00204-024-03768-5

18.Zhand, P., Kishimoto, Y., Grammatikakis, I., Gottimukkala, K., Cutler, R. G., Zhang, S., Abdelmohsen, K., Bohr, V., Sen, J. M., Gorospe, M., & Mattson, M. P. (2019, April 1). Senolytic therapy alleviates Aβ-associated oligodendrocyte progenitor cell senescence and cognitive deficits in an Alzheimer's disease model. Nature Neuroscience, 22(5), 719–728. https://doi.org/10.1038/s41593-019-0372-9

The Correlation Between *Raphanus sativus* L. Seed Germination and the Exposure to Zinc Oxide By Kaiden Kim

Introduction

In modern times, the environmental consequences of sunscreen usage have gained significant attention as evidence regarding the toxicity of its chemical ingredients has surfaced. While sunscreen protects human skin from ultraviolet (UV) radiation through its UV filters, these filters often damage the environment. Specifically, UV filters have been classified as pseudo-persistent pollutants as they are continuously released into the environment and resist degradation (Amini). Therefore, the pollution of sunscreen compounds continues to persist throughout the environment. Steinhoff offers evidence to directly support this idea in his article as he finds that there is "an estimated 6,000 to 14,000 tonnes of UV filters released annually into coral reef zones alone" (par. 3). As a result, the magnitude and consistency of UV filter pollution make the topic of sunscreen pollution particularly relevant as a rising issue in the modern world.

While a variety of sunscreens exist, oxybenzone, octocrylene, and zinc oxide (ZnO) are among the most widely used UV filters, each associated with their own environmental risks. Oxybenzone has largely been proven to be a damaging agent to environments through a variety of measures (Periago and Moz). The reports of negative environmental effects have become so numerous that "some governments have already taken action" as "Hawaii, the US Virgin Islands and Palau have enacted bans on the sale of sunscreens containing oxybenzone" (Staff, par. 8). Meanwhile, zinc oxide has largely escaped criticism due to its common commercial label as an environmentally safe sunscreen. Specifically, Ginzburg et al. find "sunscreens containing TiO_2 and ZnO are increasingly marketed as safer alternatives to small-molecule ('chemical') sunscreens" (1274). However, research demonstrates "many sunscreen labels boast that their contents are 'eco-friendly' or 'reef-safe'" while "there are no regulations defining what an 'eco-friendly' sunscreen actually is" (Periago and Moz, par. 8). Therefore, zinc oxide does not completely escape criticism either as there is a lack of regulation on its commercialized environmental safety. The culmination of this lack of regulation and definition results in what Wadyka describes as "some tricky choices and no definitive answer to the question of what's safe" for consumers (par. 6). As a result, research on the effects of ZnO on the environment has become particularly relevant. Specifically, seeds are an important component of the environment as pollutants in soil often exhibit a threat to seed germination, posing a hazard to food supply around the globe (Luo et al.). Therefore, to address these various uncertainties and topics, the research question asks: What is the correlation between exposure to realistic concentrations of zinc oxide in freshwater and the germination rate of *Raphanus sativus* L. seed?

Literature Review

UV Filters as Environmental Pollutants

The classification of sunscreen chemicals as pseudo-persistent pollutants arises from their widespread use and continual release into aquatic environments. Balmer et al. find that this continual release was primarily originating from "two principal input pathways, namely, 'direct inputs' from recreational activities (release from skin during bathing and swimming) and 'indirect inputs' via wastewater treatment plants" (953). This proves their status as environmental pollutants. However, existing research has mostly covered the release of UV filters into saltwater environments primarily due to the high frequency of recreational activities at beaches. Examples include Downs et al., who find in their research that an investigation of "contamination of Hanauma Bay saw water-column oxybenzone concentrations of more than 1600 ng/L" (2). Meanwhile, Goksøyr et al. find "the surface microlayer samples contained only low levels of analytes when they were at all detected" (4788). Despite their differences in findings, both perspectives demonstrate the release of UV pollutants into oceanic ecosystems has long been established. However, research has also shown that UV pollutants reach bodies of freshwater. For example, Poiger et al. discover "UV filters were detected at low concentrations in water samples from Lake Zurich and Huttnersee" where "the highest concentrations (up to 125 ng l^{-1}) were found at Lake Huttnersee" (962). Although the study mentions that the concentration in freshwater is low, the continued addition of sunscreen pollutants to water systems only increases the chance of making potential issues more prevalent. However, as concentrations of UV filters in freshwater are often found to be low, the environmental impact of UV filters such as ZnO in freshwater systems is understudied compared to those in oceanic ecosystems. As a result, the contamination of ocean water and freshwater by UV filters demonstrates some of the many ways in which sunscreen enters the environment. With the potential for freshwater systems to reach crops, it becomes increasingly relevant to analyze the possible implications of UV filter release on agricultural factors such as seeds.

Review of Raphanus sativus L.

When examining the ecological effects of UV filters, such as ZnO, it is important to test a model that is reproducible and agriculturally relevant. The plant *Raphanus sativus* L. (radish) meets this standard as it is widely recognized in the field of environmental toxicology as a bioindicator species; this classification of species revolves around organisms that are used to measure the general ecological toxicity of a stimulus through observable biological responses, including inhibited seed germination or reduced root growth (Yavuz et al.). As a result, due to the classification of radish as a bioindicator species, its responses to toxic pollutants can indicate the health of a broader ecosystem. However, the implications of testing on radish go beyond the health of a general ecosystem. Specifically, radish is also used as a model species as its responses can reflect the reactions of other plants across species. For example, comparative studies demonstrate that the magnitude of germination inhibition in radish parallels that of other common species such

as lettuce, wheat, and barley when exposed to similar pollutants (Di Salvatore et al.; Oleszczuk). These cross-species experiments demonstrate that radish is a plant model that is capable of predicting how a variety of crops across different species could respond to environmental conditions. Overall, existing research proves radishes can both indicate the general health of an ecosystem and reflect the responses to pollutants seen in a variety of different plant species. Therefore, the relevancy of testing the effects of ZnO on radish is further solidified.

While radish in general is a satisfactory species to use in testing ecological toxicity, radish seeds were specifically chosen to be tested due to their rapid germination rate and high sensitivity to environmental contaminants. These characteristics make them highly responsive in short-term experiments concerning pollutants such as ZnO. Radish seeds are further relevant in phytotoxicity experiments as their germination and root elongation can be accurately measured, creating reliable standards for identifying pollutant stress (Wierzbicka and Obidzińska). Although the levels of ZnO filters naturally found in the environment may be low, studies have already demonstrated that radish seeds can exhibit responses to pollutants even at low concentrations. For example, Yavuz et al. find that exposing radish seeds to trace amounts of petroleum by-products significantly reduces radish seed germination, confirming their sensitivity to low pollutant concentrations. As a result, the seeds of radish are an appropriate model for researching the effects of ZnO due to their rapid germination rates and extreme sensitivity to stimuli.

Review of Zinc Oxide

Zinc oxide has often been perceived as a safer, mineral-based UV filter alternative to chemical-based filters such as oxybenzone. This argument is primarily based on the existing literature that views zinc oxide as an organic compound that promotes plant growth. Specifically, Geremew et al. find ZnO supplements "are effective seed priming agents for *Amaranthus tricolor*, significantly enhancing germination traits, seedling vigor, and antioxidant enzyme activities" (12). While this may be true, this type of research exposes plants to zinc oxide at unrealistically low concentrations. As the supplementation of zinc oxide yields positive results, many perceive zinc oxide to be a safe, environmentally friendly alternative to UV filters such as oxybenzone. As a result, a substantial amount of research around the effects of zinc oxide as a UV filter pollutant on the environment is missing. Miller et al. highlight this concern as they find "the substantial unequal treatment of organic and inorganic UV filters in political decision-making, as well as in the 'reef safe' sunscreen market is not scientifically justified" leading them to "highly encourage a risk-based approach for ... both, organic and inorganic UV filters" (10). Therefore, zinc oxide—an organic UV filter—and its effects on the environment lack the same scrutiny in research that oxybenzone often does. As a result, it becomes increasingly relevant to analyze the effects of water polluted by zinc oxide-based filters on environmental conditions. The little research that does cover the environmental effects of ZnO contradicts findings such as those of Geremew et al. who report positive effects. For example, by examining zinc oxide as a pollutant in water, Hazeem et al. note that ZnO particles "were found to have a negative effect on algal growth and chlorophyll concentration" (2829). As a

result, ZnO demonstrates its debatable effects on environmental conditions. While controlled studies may find the ZnO particles support plant growth, studies of the compound as a pollutant reveal largely different results. Therefore, research on the effects of ZnO at environmentally realistic concentrations remains largely uncovered with most research analyzing zinc oxide as a plant supplement.

Gap and Hypothesis

Upon reviewing the existing literature, research demonstrates that ZnO has largely been proven to have a positive effect on the environment at extremely low concentrations through enhanced plant growth. However, contrasting arguments find that extreme concentrations of ZnO lead to negative environmental effects; this has led researchers to encourage analyzing the environmental effects of the UV filter ZnO with greater scrutiny. Specifically, this uncertainty arises from the lack of literature addressing the effects of ZnO at concentrations regularly found in the environment. As a result, the gap in research is made clear: realistic simulations on the effects of ZnO in the environment remain absent. Within this limited literature, the impact of ZnO on seeds also remains understudied. Meanwhile, it is clear that ZnO and other UV filters are introduced into the environment through freshwater pathways that may reach crops. Therefore, the topic of research for this paper is situated: the goal is to find the correlation between realistic concentrations of ZnO in freshwater and its effects on radish seed germination. It is hypothesized that ZnO at low environmentally representative concentrations will correlate with normal radish seed growth. Meanwhile, simulations of ZnO exposure at moderate and extreme concentrations will correlate with inhibited radish seed growth as the amount of ZnO approaches levels toxic to the plant.

Method

General Design

This study used a germination bioassay with *Raphanus sativus* L. (radish) that aimed to correlate exposure to ZnO concentrations with seed germination. A germination bioassay is a type of study that measures the biological toxicity of a substance in plant growth by observing seed germination. Specifically, radish seeds were chosen, and their significance was established earlier in the literature review. This experiment largely followed the set-up of a standardized Petri dish and germination paper method established by the International Seed Testing Association which outlines the rules for germination bioassays (International Seed Testing Association). The guidelines explain that Petri dishes are used instead of soil as they allow for more precise control of factors such as temperature, moisture, and oxygen compared to soil. As a result, they minimize the variability of factors affecting seed germination, consequently creating accurate models of seed development. Therefore, this study replicated the design as seeds exposed to the stimulus were kept in Petri dishes during growth.

In the study, the stimulus was ZnO concentrations at environmentally representative levels of 50, 150, and 500 nanograms per liter to represent ZnO UV filters released

into the environment. These values were taken from the measured levels of ZnO in freshwater systems worldwide. According to Dumont et al., concentrations of ZnO UV filters in surface waters could range from 1 nanogram per liter to hundreds of nanograms per liter as the most exposed rivers had concentrations "exceed[ing] 300 ng L^{-1} for nano ZnO" (347). These ranges of ZnO concentration were also supported by Gottschalk et al., who reported similar data on higher concentrations. Therefore, the selected range of ZnO concentrations reflect real environmental conditions reported by the existing literature. As a result, the study was designed with three exposure groups at 50, 150, and 500 nanograms per liter of ZnO.

Lastly, two controls were included: a negative control group of deionized (DE) water and a dispersant control group of DE water with 0.01% unscented dish soap. This dispersant group was created to isolate any of the effects the soap used to disperse ZnO particles had on the seeds. Existing studies such as Cheng et al. also used control groups of water and dispersant in germination involving tomato seeds, further reinforcing their value to this research paper.

Pilot Study

To further justify the chosen concentration of ZnO, it is important to mention a pilot study conducted prior to the experimental design. Specifically, the concentrations of ZnO were not only chosen to represent environmentally realistic data but also chosen as they were biologically responsive. In further detail, a range of nanogram per liter ZnO solutions were exposed to radish seeds under identical conditions and identical durations as used in the actual experiment. After 7 days, there were notable variations in germination in the higher ranges of ZnO concentration starting at 40 nanograms per liter while there was little variation below this range. Therefore, as the radish seeds showed varied germination starting at around 40 nanograms, the range of 50, 150, and 500 nanogram per liter solutions was chosen as they would likely lead to a notable statistical trend while reflecting environmentally realistic concentrations.

Solution Preparation

To create the solutions, 10 milligrams of reagent grade ZnO powder was measured and added to a 100-milliliter volumetric flask. Then the dispersant solution of 100 milliliters of deionized water with 0.01% dish soap was added to the volumetric flask to fill to the 100-milliliter mark. This created a 100 milligram per liter stock solution (Stock 1) of ZnO that was transferred to an amber bottle and shaken to prevent particles from settling. Studies such as Yin et al. followed similar procedures of solution preparation and dilution but sonicated the solution to disperse particles rather than using physical force. However, a sonicator could not be obtained for this research. To further dilute the solution, a 1 milligram per liter solution of ZnO was created by pipetting 1 milliliter of Stock 1 into 99 milliliters of the dispersant solution in a separate 100 milliliter amber bottle, creating Stock 2. Then, 1 milliliter of Stock 2 was added to 99 milliliters of dispersant to create the 10 nanogram per liter solution of ZnO (Stock 3). Lastly, 10 milliliters of Stock 3 was added to 90 milliliters of dispersant to create the final solution of 1 nanogram per liter ZnO solution

(Stock 4). To perform these series of dilutions, the equation

$C_1V_1 = C_2V_2$(1)
was used where C_1 and V_1 are the initial concentrations and initial volumes respectively. Meanwhile, C_2 and V_2 were the concentrations and volumes needed respectively. For example, to make Stock 4, the volume of Stock 3 needed was found by rearranging (1) to show V_1 = (1 ng/L × 100 mL) / (10 ng/L) = 10 mL(2)

Next, the solutions used for the three stimulus groups were created using Stock 4 as the source. For all testing solutions, the calculations were made using (1) again. For the 50 nanogram per liter solution, 95 milliliters of dispersant solution was added to 5 milliliters of Stock 4 in a 100-milliliter amber bottle. For the 150 nanogram per liter solution, 85 milliliters of the dispersant solution was combined with 15 milliliters of Stock 4. Lastly, for the 500 nanogram per liter solution 50 milliliters of dispersant solutions was combined with 50 milliliters of Stock 4. This concluded the solution preparation.

Experimental Setup

Each treatment and control group had four replicate Petri dishes containing 20 seeds each. Zhu et al. provided guidelines for this setup as they also used four replicates per group and around 20 seeds per replicate in their study of rain frequency and seed germination. To expose the seeds to ZnO, each Petri dish was lined with germination paper and moistened with 5 milliliters of the appropriate chosen solution. Jiang et al. used identical volumes of solutions to moisten the papers in the same 90-millimeter Petri dishes used for this research. The Petri dishes were placed in a growth chamber at 23 degrees Celsius with humidity at 60% and a photoperiod of 16 hours of light and 8 hours of darkness, following the setup of ideal conditions for radish growth used by Kanjevac et al. Additionally, the placement of dishes was randomized within the chamber and rotated daily to minimize any unforeseen variability in the growth chamber. Lastly, following the timeline of Maroufi and Farahani who also researched radish germination, the results were collected on day seven and are included in Appendix A.

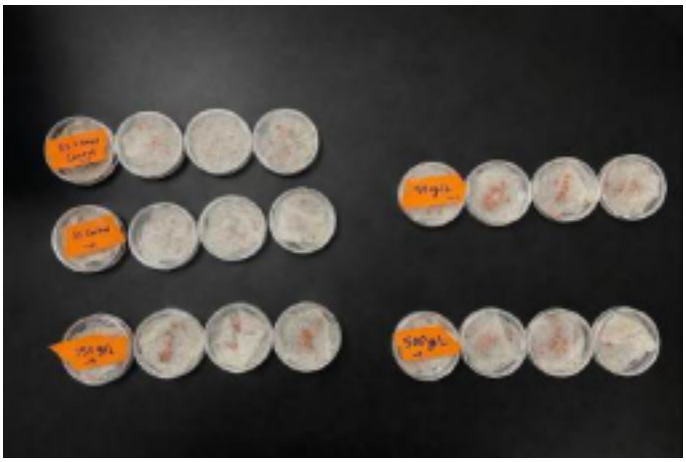


Fig. 1. Image of Petri dish preparation.

Data and Analysis

Results

For this study, the percentage of seeds germinated per Petri dish and the average radicle (root) length of seeds per Petri dish was measured, reflecting existing studies in germination such as Toscano et al. and Di Salvatore et al. who also collected this data. Following the guidelines of Zhu et al., seeds were considered germinated when the radicle grew one millimeter from the seed coat (3). These measurements were collected to create a model of linear regression that demonstrates the correlation between realistic ZnO levels and the germination success of the seeds. Pouvreau et al. use a similar method as they also create a linear regression model that displayed the percent germination and radicle length of seeds. This display of data establishes the correlational relationship between seed germination and exposure to the stimulus. From this data, the Pearson correlation coefficient and *p*-value are taken to evaluate the strength of the correlation, reflecting the analysis of Zhang et al. Overall, the purpose and use of the data collected for this paper is based on existing studies on germination.

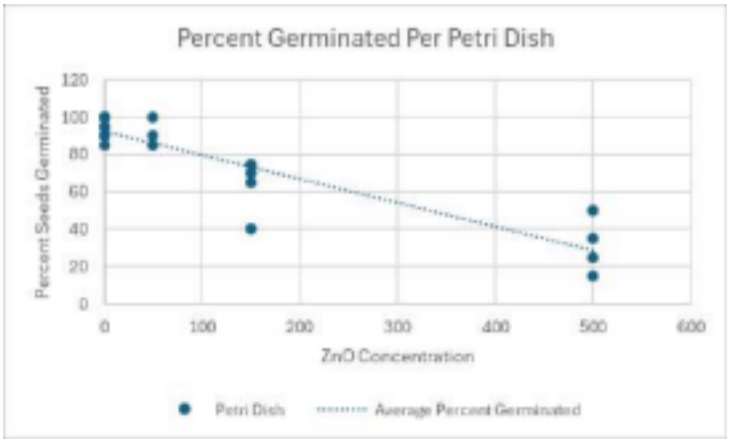


Fig. 2. Percent seeds germinated per Petri dish.

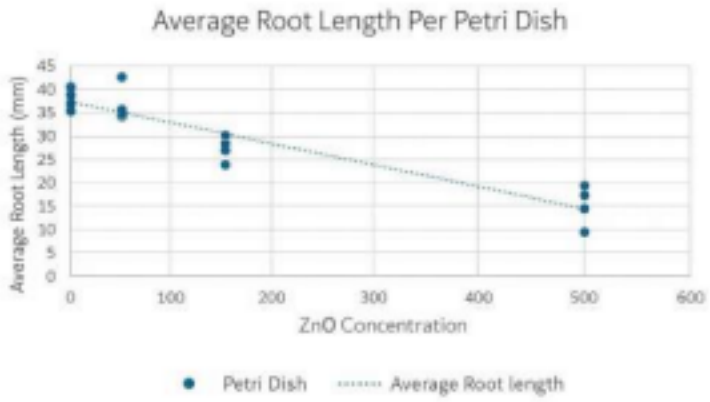


Fig. 3. Average root length per Petri dish.

The data collected is displayed on two separate models of linear regression with Petri dishes representing each data point. The x-axis for each model is the ZnO concentration, and the y-axis is the measurement taken. Petri dish data is only available at 50 ng/L, 150 ng/L, and 500 ng/L ZnO concentrations as these were the only stimulus groups tested. However, a line of best fit is graphed to demonstrate the general strength and direction of the correlation between the exposure groups. For Figure 2, the Pearson's coefficient and *p*-value are –0.912 and 2.22×10^{-8} respectively. Meanwhile, for Figure 3, the Pearson's coefficient and *p*-value are –0.933 and 2.11×10^{-9} respectively. These models are created, and these values are found using the functions within Microsoft Excel.

Evaluation and Discussion

As seen in Figure 2, the percent germination demonstrates a clear negative trend as ZnO concentration increases. The germination percentage for the deionized water control and the dispersant control are extremely close in value, indicating that the dispersant used in ZnO solutions did not impact seed viability. Therefore, the correlation between ZnO concentration and seedling success can be isolated to the ZnO stimulus. Meanwhile, the seeds exposed to ZnO solutions show progressively lower germination percentages in correlation with increasing concentrations, particularly at 150 ng/L and 500 ng/L. Figure 3 demonstrates a similar trend as the radicle length in both the control groups remained high relative to the seeds exposed to ZnO solutions. Meanwhile, the average radicle length per Petri dish decreases in correlation to increasing ZnO concentrations, indicating increased levels of stress. The strength of these correlations is also proven by the Pearson correlation coefficients as Figure 2 has a value of –0.912, indicating there is a strong negative correlation between percent germination and ZnO concentrations; this means the variables measured change inversely proportional to each other throughout all groups. Figure 3 has a similar Pearson correlation coefficient of –0.933, indicating a slightly stronger negative correlation between average radicle length and ZnO concentrations. Additionally, *p*-values

for Figure 2 and Figure 3 are 2.22×10^{-8} and 2.11×10^{-9} respectively, demonstrating that the Pearson correlation coefficients are statistically significant, and the correlation is not based on chance. While there is a strong negative correlation as a whole, percent germination and radicle length both drastically decrease in proportion to ZnO concentrations at 150 ng/L and 500 ng/L, demonstrating ZnO causes stress to radish seed growth at medium to high concentrations. Therefore, the hypothesis is validated as radish seeds are scarcely impacted at low environmentally representative concentrations of 50 ng/L while medium and high concentrations of 150 ng/L and 500 ng/L correlate to greater seedling stress. Regardless, the overarching theme shows that increased ZnO concentrations strongly correlate to decreased germination percentage and shorter radicle length in radish seeds.

Conclusion

This study demonstrates a correlational relationship between ZnO levels and early radish seed development. At low environmentally representative concentrations, the seedling response correlates with no seedling stress as there was little change in the percent germination or the radicle length. Meanwhile, the medium and high environmentally representative concentrations correlate with decreasing radicle length and percent germination in relation to increasing ZnO concentration. These results support the hypothesis that increasing levels of ZnO in freshwater systems correlates with the inhibition of radish seed development. As there has been limited research in the broader body of literature on the impacts of ZnO release, this research provides new insight on how the release of ZnO-based UV filter pollutants relates to the development of crop seeds such as radish. However, as this study establishes correlation and not causation between ZnO levels and germination inhibition, the results should purely be interpreted as support for the hypothesis of this study rather than a definitive measure of ZnO toxicity. Overall, the study demonstrates that higher ZnO concentrations regularly found in the environment can correlate to decreased radish seed germination.

Implications

First, the results of this study contribute to the ongoing debate on the true validity of “ecologically safe” sunscreens that are commonly marketed to consumers to be harmless for the environment. Specifically, ZnO is primarily released into the environment through the use of “ecologically safe” sunscreens that contain ZnO UV filters. Contrasting this label, this study has shown a correlation between ZnO exposure and germination inhibition. Therefore, the trends developed in this study imply that the common labeling of ZnO sunscreens as “ecologically safe” is challenged, reflecting the need for reevaluation expressed by Miller et al. in the literature review. This concern is emphasized as the literature review establishes radishes as a bioindicator species, meaning that their responses to stimuli—for this paper ZnO—can show the general toxicity the stimuli may have to a broader ecosystem. As a result, the correlation established in this study implies that the “ecologically safe” status of ZnO sunscreens commonly marketed towards consumers and environmental analysts alike must be reevaluated. By changing this marketing label, the use of ZnO sunscreens by consumers could decrease at a global scale, leading to less ZnO

release into the environment. Overall, this could lead to decreased revenue for companies that sell ZnO sunscreen.

Beyond changing customer perception, reevaluating the labeling of ZnO sunscreen may also lead to changes in environmental policies. Currently, the use of sunscreens containing oxybenzone has become highly restricted in several coastal countries due to its heavily researched negative environmental impacts (Staff). However, "ecologically safe" ZnO sunscreens have not been heavily researched or faced the same regulations due to their labeling, leading to their continuous release into the environment (Miller et al.); this demonstrates the gap within the literature around the environmental impacts of ZnO, and the consequential lack of regulation with ZnO sunscreens that stem from the "ecologically safe" label. As the results of the study challenge this label, this study may spark preliminary discussions that encourage future environmental researchers and policymakers to expand the current knowledge of ZnO's environmental effects. This could be achieved through testing the effects of ZnO exposure on a variety of plants to inform future regulations on the scale of environmental impact from ZnO release. Eventually, the expanded research could lead to policies that recategorize ZnO sunscreens as environmentally harmful, further clarifying the implications of ZnO sunscreen use to consumers.

Limitations

Although this study presents evidence of a relationship between ZnO concentration and inhibited germination, some limitations restrict how far the implications of the conclusion can extend. First, radish seeds are a bioindicator species, meaning their responses to stimuli can give insight into the potential ecological stress the stimuli may cause. However, specific responses seen in individual species may differ due to components such as seed composition and nutritional demands (Jain et al.). Therefore, the correlation established in this study cannot be generalized to specific species of other crops.

Next, as this study utilized a correlational relationship, the results cannot identify the specific internal mechanisms ZnO may trigger in radish seeds and the reason behind the exhibited effects. As a result, this research can only support the idea that increasing ZnO concentration correlates to decreased radish seed germination, rather than the notion that ZnO directly causes radish seeds to germinate to a lesser extent.

Lastly, the setting and concentrations of ZnO tested were restricted for this study. Consequently, a natural setting with soil and biotic factors may alter ZnO behavior, meaning that the correlation from this study cannot be generalized to all radish seeds grown in the natural environment. Meanwhile, the restriction of ZnO concentrations means that the correlation exists between the low to high environmentally realistic concentrations of 50 to 500 ng/L. As the literature review establishes that UV filters such as ZnO are considered pseudo-persistent pollutants and continue to accumulate in the environment, the correlation established in this study will no longer represent the potential ecological

impacts of ZnO once the concentrations in the environment extend beyond the range used in this study.

Future Directions

Future research could expand the ecological scope of the findings by researching the correlation between ZnO exposure and the germination of additional plant species. These species may have seeds that vary in structure and tolerance to environmental stressors, helping determine whether the responses seen in radish seeds are unique to the species or reflect broader plant responses. With this additional research, the true impacts of ZnO as a pollutant would extend the research to general terrestrial plants. Additionally, future research could continue the methods for an extended period of time to determine how ZnO correlates to later stages of plant development; this includes measures such as root development, biomass accumulation, and crop yield. With this difference in methods, the findings could inform future researchers of the impact ZnO may have on plant development beyond initial germination. Furthermore, differences in methods that change the growth conditions could be implemented such as mimicking microbial communities

in soil or fluctuating moisture levels. These changes may fundamentally alter the correlation seen between seed growth and ZnO exposure, expanding the implications to reflect more environmentally realistic conditions. Lastly, the current research could be expanded to specifically examine the effects ZnO has on the mechanisms of seedling growth. This would illuminate why ZnO may disrupt early developmental processes, creating a causal relationship between ZnO concentration and seedling growth.

Appendix A: Radish Measurements

Table A1. Measurements of germination percentage and average radicle length per Petri dish.

Treatment	**ZnO Concentration**				**Average Root Length**
DIWa	0				41.2 mm

ter					
DI Water	0				39.5 mm
DI Water	0				37.3 mm
DI Water	0				35.6 mm
Dispersant Control	0				40.8 mm
Dispersant Control	0				38.4 mm

DispersantControl	0				35.7 mm
DispersantControl	0				35.8 mm
ZnO50ng/L	50				42.5 mm
ZnO50ng/L	50				35.3 mm

ZnO 50 ng/L 50 3 17 85% 34.8 mm ZnO 50 ng/L 50 4 17 85% 34.6 mm ZnO 150 ng/L 150 1 15 75% 30.3 mm ZnO 150 ng/L 150 2 14 70% 28.7 mm ZnO 150 ng/L 150 3 13 65% 26.9 mm ZnO 150 ng/L 150 4 8 40% 24.1 mm ZnO 500 ng/L 500 1 10 50% 19.9 mm ZnO 500 ng/L 500 2 7 35% 18.4 mm

ZnO 500 ng/L 500 3 5 25% 15.1 mm ZnO 500 ng/L 500 4 3 15% 10.2 mm

Works Cited

Amini, M. “Sunscreen’s Impact on Marine Life Needs Urgent Investigation, Study Finds.” The Guardian, 18 Feb. 2025, www.theguardian.com/environment/2025/feb/18/sunscreens-impact-on-marine-life-needs-urgent-investigation.

Balmer, M. E., et al. “Occurrence of Some Organic UV Filters in Wastewater, in Surface Waters, and in Fish from Swiss Lakes.” Environmental Science and Technology, vol. 39, no. 4, 2005, pp. 953–62, doi.org/10.1021/es040055r.

Cheng, F., et al. “The Garlic Allelochemical Diallyl Disulfide Affects Tomato Root Growth by Influencing Cell Division, Phytohormone Balance and Expansin Gene Expression.” Frontiers in Plant Science, vol. 7, 2016, pp. 1–16, doi.org/10.3389/fpls.2016.01199.

Di Salvatore, M., et al. “Assessment of Heavy Metals Phytotoxicity Using Seed Germination and Root Elongation Tests: A Comparison of Two Growth Substrates.” Chemosphere, vol. 73, no. 9, 2008, pp. 1461–64, doi.org/10.1016/j.chemosphere.2008.07.061.

Downs, C. A., et al. “Oxybenzone Contamination from Sunscreen Pollution and Its Ecological Threat to Hanauma Bay, Oahu, Hawaii, U.S.A.” Chemosphere, vol. 291, 2022, pp. 1–18, doi.org/10.1016/j.chemosphere.2021.132880.

Dumont, E., et al. “Nano Silver and Nano Zinc-Oxide in Surface Waters: Exposure Estimation for Europe at High Spatial and Temporal Resolution.” Environmental Pollution, vol. 196, 2015, pp. 156–64, nora.nerc.ac.uk/id/eprint/508857/1/N508857JA.pdf.

Geremew, A., et al. “Nanopriming with Zinc Oxide: A Novel Approach to Enhance Germination and Antioxidant Systems in Amaranth.” Frontiers in Plant Science, vol. 16, 2025, pp. 1–15, doi.org/10.3389/fpls.2025.1599192.

Ginzburg, A. L., et al. “Zinc Oxide-Induced Changes to Sunscreen Ingredient Efficacy and Toxicity under UV Irradiation.” Photochemical & Photobiological Sciences, vol. 20, no. 10, 2021, pp. 1273–85, doi.org/10.1007/s43630-021-00101-2.

Goksøyr, A., et al. “Balsa Raft Crossing the Pacific Finds Low Contaminant Levels.” Environmental Science and Technology, vol. 43, no. 13, 2009, pp. 4783–90, doi.org/10.1021/es900154h.

Gottschalk, F., et al. “Modeling Flows and Concentrations of Nine Engineered Nanomaterials in the Danish Environment.” International Journal of Environmental Research and Public Health, vol. 12, no. 5, 2015, pp. 5581–5602, pdfs.semanticscholar.org/0f98/d380b4a1651ce8b49d5d6f53b31f5a2fc167.pdf. Hazeem, L. J., et al. “Cumulative Effect of Zinc Oxide and Titanium Oxide Nanoparticles on Growth and Chlorophyll a Content of Picochlorum sp.” Environmental Science and Pollution Research, vol. 23, no. 3, 2015, pp. 2821–30, doi.org/10.1007/s11356-015-5493-4.

International Seed Testing Association. International Rules for Seed Testing: Annexe to Chapter 7—Validated Seed Health Testing Methods. ISTA, 2022, www.fito2009.com/fitop/7-004.pdf.
Jain, N., et al. “Does Seed Size and Surface Anatomy Play a Role in Combating Phytotoxicity of Nanoparticles?” Ecotoxicology, vol. 26, no. 2, 2017, pp. 238–49, doi.org/10.1007/s10646-017-1758-7.
Jiang, L., et al. “PVC Inhibits Radish (Raphanus sativus L.) Seedling Growth by Interfering with Plant Hormone Signal Transduction and Phenylpropanoid Biosynthesis.” Horticulturae, vol. 11, no. 8, 2025, p. 896, doi.org/10.3390/horticulturae11080896.
Kanjevac, M., et al. “Improvement of Germination and Early Growth of Radish (Raphanus sativus L.) through Modulation of Seed Metabolic Processes.” Plants, vol. 11, no. 6, 2022, p. 757, doi.org/10.3390/plants11060757.
Luo, C., et al. “Environmental Risk Substances in Soil on Seed Germination: Chemical Species, Inhibition Performance, and Mechanisms.” Journal of Hazardous Materials, vol. 472, 2024, p. 134518, doi.org/10.1016/j.jhazmat.2024.134518.
Maroufi, K., and H. A. Farahani. “Increasing Germination by Hydropriming Method in Radish (Raphanus sativus L.).” Advances in Environmental Biology, vol. 5, no. 10, 2011, pp. 3440–43, www.aensiweb.com/old/aeb/2011/3440-3443.pdf.
Miller, I. B., et al. “Toxic Effects of UV Filters from Sunscreens on Coral Reefs Revisited: Regulatory Aspects for ‘Reef Safe’ Products.” Environmental Sciences Europe, vol. 33, no. 74, 2021, pp. 1–13, doi.org/10.1186/s12302-021-00515-w.
Oleszczuk, P. “Phytotoxicity of Municipal Sewage Sludge Composts Related to Physico-Chemical Properties, PAHs and Heavy Metals.” Ecotoxicology and Environmental Safety, vol. 69, no. 3, 2008, pp. 496–505, doi.org/10.1016/j.ecoenv.2007.04.006.
Periago, R. P., and C. Moz. “The Not-So-Sunny-Side of Sunscreen.” Revolve, 9 July 2025, revolve.media/features/impact-sunscreen-environment-pollution.
Poiger, T., et al. “Occurrence of UV Filter Compounds from Sunscreens in Surface Waters: Regional Mass Balance in Two Swiss Lakes.” Chemosphere, vol. 55, no. 7, 2004, pp. 951–63, doi.org/10.1016/j.chemosphere.2004.01.012.
Pouvreau, J. B., et al. “A High-Throughput Seed Germination Assay for Root Parasitic Plants.” Plant Methods, vol. 9, no. 1, 2013, p. 32, doi.org/10.1186/1746-4811-9-32. Staff, D. “Scientists Warn of Growing Threat from Sunscreen to Marine Life, Call for Urgent Investigation.” Down to Earth, 20 Feb. 2025, www.downtoearth.org.in/science-technology/scientists-warn-of-growing-threat-from-sunscreen-to-marine-life-call-for-urgent-investigation.
Steinhoff, N. “Impact of Sunscreens on Marine Life ‘Needs Urgent Investigation.’” Oceanographic, 18 Feb. 2025, oceanographicmagazine.com/news/impact-of-sunscreens-on-marine-life-needs-urgent-investigation/.
Toscano, S., et al. “Annual Garden Rocket and Radish as Microgreens: Seed Germination Response to Thermal and Salt Stress.” Agronomy, vol. 15, no. 2, 2025,

p. 361, doi.org/10.3390/agronomy15020361.
Wadyka, S. “The Truth About ‘Reef Safe’ Sunscreen.” Consumer Reports, 10 July 2018, www.consumerreports.org/health/sunscreens/the-truth-about-reef-safe-sunscreen-a3578637894/.
Wierzbicka, M., and J. Obidzińska. “The Effect of Lead on Seed Imbibition and Germination in Different Plant Species.” Plant Science, vol. 137, no. 2, 1998, pp. 155–71, doi.org/10.1016/S0168-9452(98)00138-1.
Yavuz, B., et al. “Using Radish (Raphanus sativus L.) Germination to Establish a Benchmark Dose for the Toxicity of Ozonated-Petroleum Byproducts in Soil.” Chemosphere, vol. 313, 2023, p. 137382, doi.org/10.1016/j.chemosphere.2022.137382.
Yin, L., et al. “Effects of Silver Nanoparticle Exposure on Germination and Early Growth of Eleven Wetland Plants.” PLoS ONE, vol. 7, no. 10, 2012, pp. 1–7, doi.org/10.1371/journal.pone.0047674.
Zhang, J., et al. “Differences in Seed Properties and Germination between Native and Introduced Populations of Triadica sebifera.” Journal of Plant Ecology, vol. 13, no. 1, 2020, pp. 70–79, doi.org/10.1093/jpe/rtz048.
Zhu, Y., et al. “Response of Seed Germination and Seedling Emergence of Haloxylon ammodendron to Rain Frequency and Temperature Change from Four Desert Ecosystems, Northwest China.” AoB PLANTS, vol. 15, no. 2, 2022, pp. 1–9, doi.org/10.1093/aobpla/plac048.

Project topic:
“Tenzo” – adaptive tensegrity exoskeleton
Compiled by:
Semyon Chistyakov
Najmudin Adikhanov
Assylzhan Sailau
Jaleel Gabitov
Yerkenur Zhanabekova
Arnur Togambayev
Scientific supervisor:
Moldabayeva Z.S.Kazakhstan, 2026

ABSTRACT
The purpose of the research project is to develop, experimentally test, and economically validate an adaptive tensegrity exoskeleton for rehabilitation of the musculoskeletal system, ensuring ease of construction, biomimetics, user comfort, and disruptive market affordability. Hypothesis: the use of tensegrity structures that closely mimic the mechanics of the human body will make it possible to create a lightweight, adaptive exoskeleton that surpasses existing rigid analogues in functional efficiency while drastically reducing manufacturing costs and lowering barriers to market entry. Stages and procedure of the study:Analysis of existing exoskeletons (rigid, soft, hybrid), identifying both their technical shortcomings and high-cost market barriers. Study of the principles of biotensegrity and examples of applications in robotics. Mathematical modeling of kinematics and statics of a structure. CAD model design and selection of cost-effective, optimized production materials. Development of exoskeleton management software. Prototype assembly, experimental testing, benchmarking against traditional models, and conducting a bottom-up unit economics analysis. The experimental and financial methodology included structural modeling, practical testing of the physical prototype (measuring angles, moments, cable tension, building force- angle dependencies, assessing stability and ergonomics), alongside a comprehensive Bill of Materials (BOM) calculation and target market sizing (TAM, SAM, SOM). The novelty of the research lies in the application of the tensegrity principle to create an elbow exoskeleton for medical rehabilitation that is explicitly optimized for low-cost mass production. The degree of independence of the project is high: the authors conducted a complete theoretical analysis, structural modeling, physical prototype manufacturing with dynamic testing, and a comprehensive startup viability assessment. The results of the work showed that the tensegrity exoskeleton has low mass and rigidity, adaptively distributes operational loads, and preserves natural joint

micro-movements. Concurrently, the financial analysis demonstrated that replacing heavy rigid links with a polymer-cable matrix achieves a production Cost of Goods Sold (COGS) of approximately $250, offering a 99% cost reduction compared to legacycommercial analogues that cost up to $330,000. At the same time, technical limitations were revealed—instability at low tensions and high sensitivity to assembly accuracy.

1. Introduction

Musculoskeletal disorders are a significant problem that anyone can face. Currently, they affect approximately 1.7 billion people worldwide. Treatment and prevention are also time-consuming and expensive. Exoskeletons devices that mimic human biomechanics are now being
used to more effectively treat musculoskeletal disorders. Current exoskeleton models have numerous drawbacks, including their weight and high cost advanced models can reach $330,000 per unit. From an entrepreneurial perspective, this astronomical pricing model creates a severe market failure. While the Total Addressable Market (TAM) spans 1.7 billion potential users globally, the prohibitive Cost of Goods Sold (COGS) of incumbent solutions restricts advanced rehabilitation exclusively to top-tier, high-budget medical institutions. This leaves the massive segment of mid-tier regional clinics and direct-to-consumer (B2C) home rehabilitation completely unserved, presenting a highly lucrative market entry opportunity for a disruptive, low-cost alternative. Among their greatest drawbacks are their inability to accommodate bone displacements and their inaccurate replication of human biomechanics.

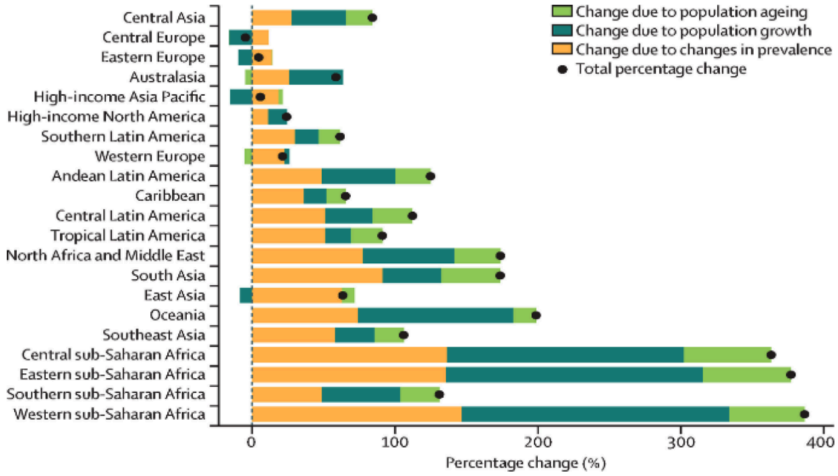


Fig. 1 Breakdown of projected changes in the incidence of other musculoskeletal disorders by region from 2020 to 2050.

An effective and convenient solution to this problem is the use of tensegrity structures in exoskeletons. Tensegrity is a design principle for structures made of rods and cables that

combines lightness, discontinuity, high flexibility, and superior biomimetic properties. Tensegrity exoskeletons could become a comfortable, flexible, affordable, and lightweight replacement for current models, potentially bringing breakthroughs in musculoskeletal treatments and greatly benefiting people. Commercializing this technology allows the Tenzo project to establish a powerful Unique Value Proposition (UVP). By replacing heavy, precision-machined rigid metallic links with an optimized structural web of cables and 3D-printed materials, we fundamentally alter the cost structure of medical robotics. This enables a highly scalable business model that dramatically reduces the capital expenditure (CapEx) for healthcare providers, making multi-axis robotic therapy accessible to the Serviceable, Addressable Market (SAM) of local rehabilitation centers. *Hypothesis:* Leveraging adaptive tensegrity topologies enables the manufacturing of a structurally compliant, medical-grade upper-limbrehabilitation exoskeleton at a fraction of the cost of traditional rigid analogues, thereby validating both its technical efficacy and its financial scalability within the healthcare equipment market, and assessing the economic viability and commercialization strategy for the developed platform.

Goals: Identifying the shortcomings of existing exoskeletons, designing and developing a tensegrity structure, creating an exoskeleton based on the structure, programming the exoskeleton, testing and comparing the results

Tasks:

1. Study existing exoskeletons (rigid, soft, hybrid) and identify their shortcomings. Consider the advantages of tensegrity structures.
2. Study the current use of tensegrity structures in exoskeletons
3. Study the loads on joints and muscles during movement.
4. Develop a model of exoskeleton-human interaction.
5. Consult with professors and engineers
6. Model an adaptive tensegrity structure that mimics the dynamics of
7. human movement.
8. Determine optimal materials
9. Compare different configurations of joints and tension elements.
10. Develop a program to control the exoskeleton.
11. Assemble and test a prototype
12. Compare the prototype with existing exoskeletons
13. Calculate the economic and technical feasibility of implementation
14. Collaborate with medical centers
15. Calculate the estimated Bill of Materials (BOM) and projected Cost of Goods Sold (COGS) to prove the venture's unit economics
16. Analyze target market sizing (TAM, SAM, SOM) and outline a sustainable B2B commercialization framework for clinic integration.

2. Literature Review

Modern exoskeletons face profound market limitations: they are heavy, prohibitively expensive (up to $330,000 per unit) , and fail to accurately replicate natural human biomechanics. According to a systematic analysis published in The Lancet Rheumatology (2023), the global socioeconomic burden of musculoskeletal disorders is acceleratingrapidly. This growing patient volume means that current high-cost rehabilitation models are financially unsustainable for healthcare systems, creating a critical market demand for affordable, lightweight wearable robotic

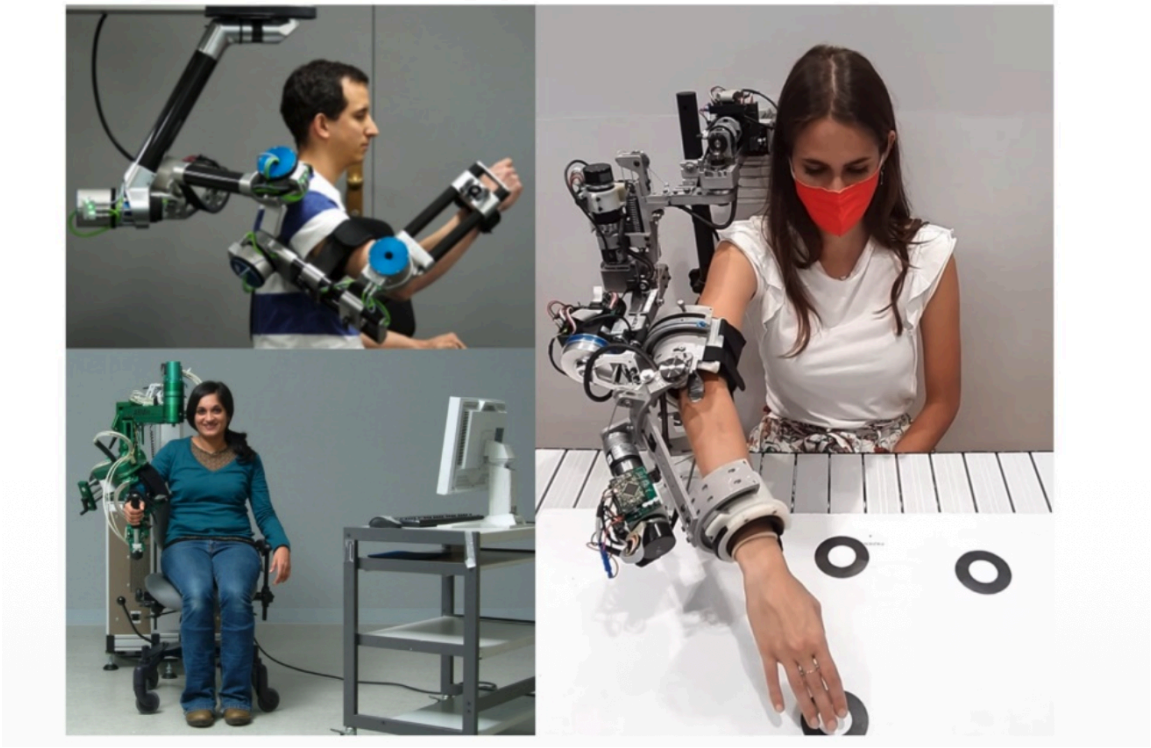


Fig 2 Modern exoskeletons for elbow joint rehabilitation

Tensegrity structures are models based on the tension of rods and cables. These structures closely mimic human biomechanics and are extremely lightweight. Research shows that the human body can be viewed as a tensegrity structure Ligaments, tendons, and fascia act as tension elements; the overall system distributes loads, ensuring strength with minimal weight.

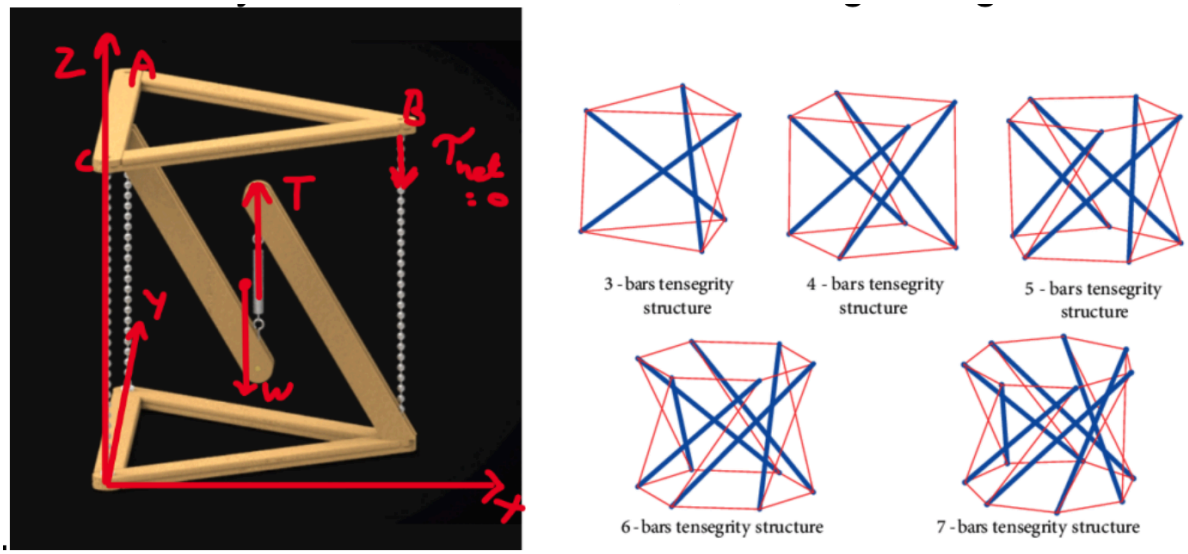


Fig.3 Tensegrity structures

The work of T. Flemons and more recent studies (Bones of Tensegrity) demonstrate that tensegrity models effectively explain joint micromotion and dynamic stress distribution. This makes this approach a natural candidate for the design of orthoses and exoskeletons. Tensegrity structures have already proven successful in various fields, and robotics has become a key area of their development. In robotics, tensegrity is used to create lightweight, adaptive systems: NASA uses tensegrity robots for planetary exploration (SUPERball projects).

Academic research has developed prototypes of tensegrity manipulators and joints that demonstrate multi-axis motion with low weight.

Advantages: variable stiffness (via adjustable cable tension), impact resistance, light weight, and modularity.

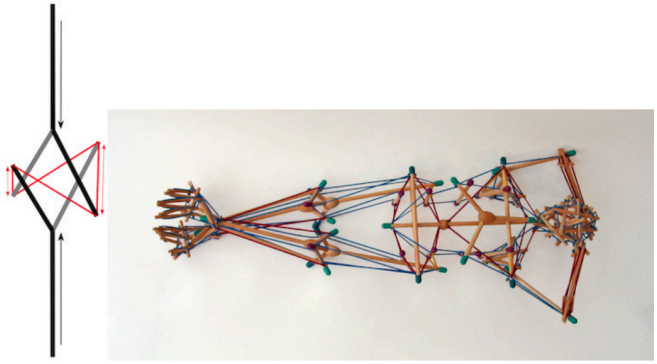


Fig.4 T.Flemons “Bones of tensegrity”

Research has already been conducted on the use of tensegrity structures to simulate human body parts. The paper "Design of a Movable Tensegrity Arm with Springs" presents a prototype robotic arm based on a tensegrity structure. Six aluminum rods (approximately 200 mm long) were used, connected by elastic elements (springs with a stiffness coefficient of $k \approx 1.5–2.0$ N/mm). The assembled module weighs approximately 1.2 kg. The work implements two degrees of freedom, the maximum flexion angle is 70°, and the force required to move the structure is 8–12 N.

Functionality:
Two degrees of freedom (flexion/extension and rotation) are implemented; maximum flexion angle is ~70°; the force required to move the structure is on average 8-12 N.

The study demonstrated high biomechanical adaptability:
• the load is distributed throughout the structure, reducing local stress in the joints;
• when applying a lateral force of up to 15 N, the structure remained stable and returned to its working position (self-stabilization).

The elbow joint plays a key role in everyday movements (lifting, self-care). After injuries, strokes, or degenerative diseases, support and rehabilitation are necessary. Conventional rigid orthoses restrict natural
micromovements, which slows recovery.

A tensegrity exoskeleton will:
• support the joint without restricting bone micromovement;
• distribute the load more physiologically;
• regulate stiffness and angle of motion during recovery;
• ensure lightness and comfort during long-term wear.

Despite research on tensegrity robots and the biotensegrity concept, the scientific literature contains very few implemented exoskeleton prototypes for specific human joints, especially for the elbow.

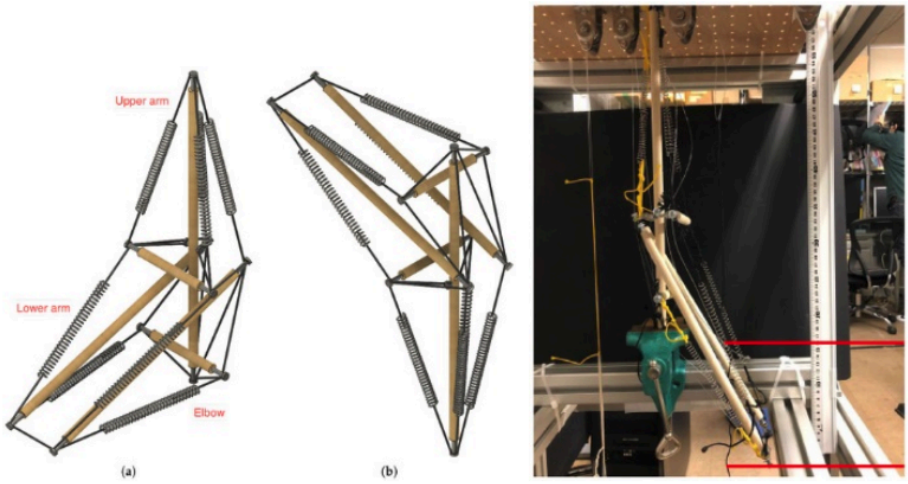


Fig.5 Tensegrity arm

In one of the few studies, "Design and Control of an Upper Limb Bionic Exoskeleton Rehabilitation Device Based on Tensegrity Structure" (Peng Ni et al.), a prototype upper limb exoskeleton was developed using a tensegrity structure. The authors simplified the anatomical model of the arm and constructed a two-rod, four-cable topology, where the rods simulated the bones and the cables simulated the flexors and extensors.

A hinge and rack-and-pinion mechanism were additionally used to compensate for elbow axis misalignment. A control algorithm based on an impedance controller was developed, the parameters of which were adapted using a BP neural network and PSO optimization. Fuzzy logic was used to assess the patient's condition.

Experiments demonstrated the stability of the design, the alignment of movement trajectories with arm movement, wearer comfort, and the ability to assist movement with loads up to 2 kg.

Key disadvantages:
• Limitation to one primary degree of freedom (elbow flexion/extension);
• Simplification of the anatomical model (radius and ulna are combined into a single element);
• Requirement for rigid constraints, which partially reduces the advantages of tensegrity;
• Lack of detailed quantitative error metrics and long- term testing;
• Limited data on materials, weight, and practical reliability.

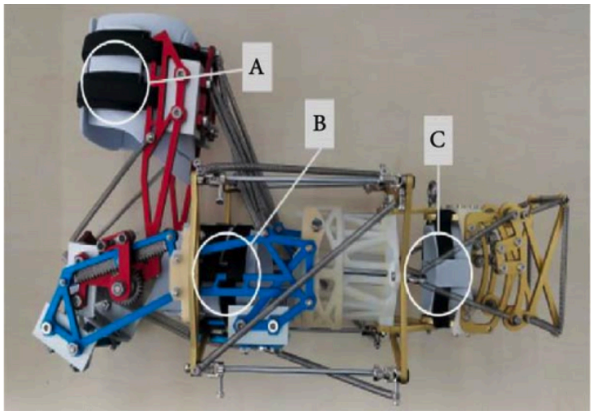


Fig.6 Tensegrity exoskeleton for arm

The main challenges in creating an exoskeleton based on a tensegrity structure are:
insufficient data on how to design the geometry of a tensegrity structure for the anatomy of the hand; few models linking human biomechanics and a controlled tensegrity structure; limited experiments on rehabilitation applications.

Using a tensegrity structure for an elbow exoskeleton will allow for the creation of:
• a lightweight and energy-efficient exoskeleton;
• a system that mimics the natural mechanics of the elbow joint;
• adjustable and safe support for patients and users.

3. Research methods

2.1 Selecting a tensegrity architecture for the elbow joint
A tensegrity structure consists of rods (rigid elements under compression) and ropes/cables (tension element). Wikipedia+2MDPI+2. The topology must provide Degrees of Freedom (DOF) for flexion/ extension, as well as some small lateral displacements (micro- displacements). A structure with a single module or a chain of modules connected by internal cables (similar to tensegrity robot structures) can be used.

The paper, "A light-weight, multi-axis compliant tensegrity joint," presents an approach to a tensegrity-based multi-axis joint that could serve as a prototype for an elbow with additional degrees of freedom. arXiv It is required that the rigid rods do not touch each other (discrete compression) - this is a characteristic feature of tensegrity. Wikipedia+1 Pre-tensioning of the cables must be provided to ensure structural stability and shape control. PMC+2PubMed+2

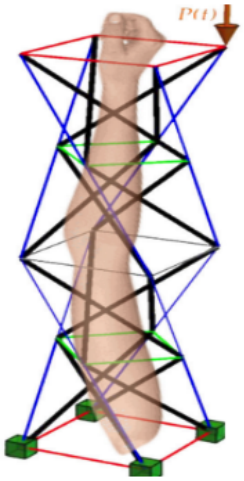


Fig.7 Tensegrity structure on arm



Fig.8 Thread tension testing

The first prototype of the single-line compression system was developed and tested. The prototype was constructed from cuffs, a motor, and a string. The effectiveness of the elbow flexion tension was demonstrated.



Fig 9 The first prototype of an exoskeleton

2.2 Mathematical Modeling: Statics and Kinematics Static Equilibrium:

ns - number of rods
nc - number of cables
ui - direction vector (unity normalized) for cable i
Ti - tension in cable i
fext - external forces (e.g., moment, load)
rj - forces in rod j, assuming the rods are in compression
The equilibrium equation at the nodes (in nodal form) can be written as:

$$\sum_{i=1}^{n_c} T_i \mathbf{u}_i + \sum_{j=1}^{n_s} \mathbf{r}_j + \mathbf{f}_{ext} = \mathbf{0}.$$

In addition, taking into account the rigidity of the cables (linear dependence of tension on elongation):

$$T_i = k_i(l_i - l_{i0})$$

where ki is the cable stiffness (modulus), li is the current length, li0 is the nominal (tensed) length.

Matrix formulation:
Let us denote the matrix of connections as A, where each row corresponds to a node, columns to cables, and the values ±1\pm 1 indicate the direction.

Then:

$$\mathbf{A}\,\mathbf{T} + \mathbf{R} + \mathbf{F}_{ext} = \mathbf{0}$$

The solution to this system gives the tension T. However, it is important that the solution satisfies the constraint Ti≥ 0 (the cable does not work in compression).

Kinematics/geometric dependencies:

If we set the configuration (bending angle), then the lengths of the cables li can be expressed through the coordinates of the nodes (depending on the angle). Let us assume that the structure is defined by nodes xk(θ), where θ is the elbow flexion angle.

Then:

$$l_i(\theta) = \|\mathbf{x}_{p(i)}(\theta) - \mathbf{x}_{q(i)}(\theta)\|$$

where the cable connects nodes p(i)p(i) and q(i)q(i).
Then:

$$T_i(\theta) = k_i(l_i(\theta) - l_{i0})$$

Stiffness (thin model):
The local rigidity of the structure along the bending axis (modular rigidity) can be calculated by differentiation:

$$K_\theta = \frac{dM}{d\theta}$$

where M is the total moment caused by the tension in the cables. Or through the second derivative of the potential energy:

$$K_\theta = \frac{d^2U}{d\theta^2}$$

Potential energy of the system:

$$U = \frac{1}{2}\sum_{i=1}^{n_c} k_i(l_i(\theta) - l_{i0})^2$$

The first and second derivatives give the forces and stiffnesses. Accounting for friction, residuals, and dynamics: If the movement is rapid or the load changes, damping and inertia must be taken into account.

Motion model:

$$I\ddot{\theta} + C\dot{\theta} + K_\theta(\theta - \theta_0) = M_{\text{actuator}}(\theta)$$

where Mactuator is the control torque from the drive (motors + cables).

3.3 Techno-Economic and Market Feasibility Methodology

To validate the business venture scalability of the Tenzo exoskeleton alongside its mechanical performance, a techno-economic assessment methodology was integrated into the research framework: Cost of Goods Sold (COGS) Framework: A bottom-up accounting method was implemented to calculate the complete Bill of Materials (BOM), aggregating the costs of structural polymers, electronic microcontrollers, actuation motors, and tension cables to determine the baseline manufacturing cost. Market Sizing Estimation: Utilizing top-down market analysis driven by global healthcare statistics from open-source epidemiological data, the target market was segmented into Total Addressable Market (TAM), Serviceable Addressable Market (SAM), and Serviceable Obtainable Market (SOM).

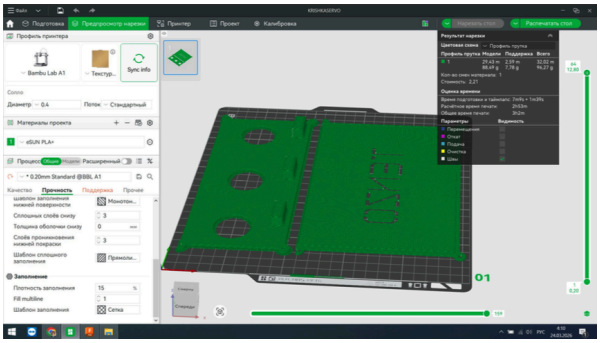
Competitive Pricing Matrix: The projected commercial price of the Tenzo device was benchmarked against existing industry-standard rigidorthoses to evaluate the financial return on investment (ROI) and cost- advantage metrics for medical clinics.

4. Results and Discussion

As a result, an exoskeleton was developed and tested. It was lightweight approximately 500 grams. The exoskeleton was evaluated in a rehabilitation center, and its cost was also estimated.







Component Category	Description / Specifications	Baseline Unit Cost (USD)
Structural Matrix	**High-strength FDM printed polymers**	**$40.00**

	(PLA/PETG optimization)	
Actuation System	**High-torque servo motors, driver modules, and pulleys**	**$120.00**
Tensile Network	**High-tensile braided steel/synthetic cables and ergonomic cuffs**	**$30.00**
Sensory Hardware	**Microcontroller unit (MCU), strain gauge sensors, and IMU tracking**	**$60.00**
Total Production COGS	**Complete manufacturing cost per unit**	**$250.00**

5. Conclusion

The following conclusions can be drawn from the results of the conducted research:

1. Tensegrity structures are ideal for creating exoskeletons: they mimic the biomechanics of the human body and are capable of simulating micro-displacements of joints.

2. Tensegrity exoskeletons provide an ideal user experience with good lifting force while maintaining low weight. The study demonstrated that using tensegrity structures as the basis for an elbow joint exoskeleton is a promising approach, offering a combination of lightweight construction, adaptability, and biomimetic motion. The developed prototype confirmed the ability to reproduce natural kinematics while maintaining structural integrity and redistributing loads through a system of tensioned cables. Unlike traditional rigid exoskeletons, the proposed design exhibits greater compliance to micro-displacements and lower weight, reducing energy consumption during operation and improving user comfort. However, limitations related to instability at low tensions, the influence of friction, and assembly accuracy have been identified, requiring refinement at both the material and control algorithm levels. Thus, the obtained results confirm the scientific and practical significance of applying tensegrity principles in rehabilitation and assistive robotics. Promising development areas include expanding the design's functionality to accommodate multi-axis movements, using intelligent tension control systems, and implementing more advanced sensors to improve precision and reliability. This will

enable the creation of a new generation of lightweight, efficient, and ergonomic exoskeletons for medical and domestic applications.

REFERENCES

1. Global, regional, and national burden of other musculoskeletal disorders, 1990–2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021 [Электронный ресурс]. - The Lancet Rheumatology, 2023. URL: https://www.thelancet.com/journals/lanrhe/article/PIIS2665-9913%2823%2900232-1/fulltext2. Technology Acceptance Model for Exoskeletons for

Rehabilitation of the Upper Limbs from Therapists' Perspectives [Электронный ресурс]. - Sensors, 2023. URL: https:// www.mdpi.com/1424-8220/23/3/1721

3. Force Density Relation and Lightweight Modeling of Single Layer Tensegrity Structures [Электронный ресурс]. - ResearchGate, 2020. URL: https://www.researchgate.net/ publication/342549805_Force_Density_Relation_and_Lightweight_Modeling_of_Single_Layer_Tensegrity_Structures

4. The forces acting on the tensegrity structure [Электронный ресурс]. - Physics StackExchange, 2024. URL: https://physics.stackexchange.com/questions/802657/the-forces-acting-on-the-tensegrity-structure

5. Bones of Tensegrity [Электронный ресурс]. - Intension Designs, 2022. URL: https://intensiondesigns.ca/bones- of-tensegrity/

6. Design of a Movable Tensegrity Arm with Springs Modeling an Upper and Lower Arm [Электронный ресурс]. - Applied Sciences, 2021. URL: https://www.mdpi.com/ 2076-0825/12/1/18

7. Design and Analysis of Tensegrity-Inspired Manipulators [Электронный ресурс]. - ResearchGate, 2024. URL: https://www.researchgate.net/publication/ 379406146_Design_and_Analysis_of_Tensegrity-Inspired_Manipulators

8. Design of Wearable Tensegrity Structures Focusing on the Tension Propagation Function Throughout the Body [Электронный ресурс]. - Proceedings of the Design Society, 2021. URL: https://www.cambridge.org/core/journals/proceedingsof-the-design-society/article/design-of-wearable-tensegrity-structures-focusing-on-the-tension-propagation-function-throughout-the-body/1273B027BC79CCC4AB0CA81A6D3D03E6

How Can Artificial Intelligence and Machine Learning Be Applied to Recruiting Athletes at the College Level, in Particular in Football? By Jonah Khoshabe

Abstract

As data and technology play an increasingly important role in sports, artificial intelligence (AI) and machine learning (ML) have become prominent tools in college athlete recruiting. This paper examines how AI and ML can be applied to the college recruiting process, with a primary focus on college football. The paper begins by explaining the fundamental concepts behind AI and ML, followed by an overview of their existing applications in sports analytics and talent identification. It then describes the traditional college football recruiting process and analyzes how AI-based tools can improve its efficiency and effectiveness. These tools include video analysis for evaluating game film, predictive models for assessing player performance, and matching systems that connect athletes with programs that best fit their skills and preferences. Various modeling approaches—including neural networks, logistic regression, support vector machines, clustering methods, and random forests—are discussed to illustrate how data-driven techniques can support recruiting decisions. The recruiting process is further framed as a two-sided matching problem, in which both athletes and colleges seek optimal outcomes. Within this framework, the paper argues that AI and ML can reduce inefficiencies, identify overlooked talent, and improve the quality of matches between players and programs.

Introduction

Each year, college football programs may evaluate well over one hundred prospective athletes, despite having only a limited number of available scholarships—often around twenty-five per recruiting class. This imbalance creates significant time pressure and workload for coaches, scouts, and recruiting staff. Coaches must identify positional needs, assess which athletes fit their tactical systems, and determine which prospects merit active recruitment, all within a highly competitive environment.

This paper examines how artificial intelligence (AI) and machine learning (ML) can be used to improve the college football recruiting process. While recruiting has long been a central component of college athletics, recent advances in data analytics have provided coaches and universities with new tools to evaluate and identify talent more efficiently. AI- and ML-based methods have the potential to streamline recruiting workflows, expand the pool of evaluated athletes, and support more informed decision-making. At the same time, the increasing use of automated systems raises concerns about whether the recruiting process may become overly mechanized, potentially reducing the human judgment traditionally involved in talent evaluation.

This research is motivated by an interest in both college sports and artificial intelligence. College athletics often depend on identifying strong matches between athletes and programs, where both parties can develop and succeed together. At the same time, advances in AI and ML demonstrate how computational systems can learn from data and support complex

decision-making across a wide range of applications. These developments motivate an examination of how similar techniques might be applied to college recruiting. Building on these interests, this paper addresses the following research question: How can artificial intelligence (AI) and machine learning (ML) be applied to the recruitment of college athletes, particularly in football? To answer this question, the paper proceeds as follows. The first section provides an overview of AI and ML concepts. The second section reviews existing literature on the use of AI and ML in sports analytics and talent identification. The third section analyzes how these methods can be applied to the college football recruiting process, alongside a discussion of traditional recruiting practices.

Overview of Artificial Intelligence and Machine Learning

Although artificial intelligence was first developed in the mid-1950s, it has become a more prominent part of everyday life in the 2020s. While many individuals interact with AI through electronic devices on a daily basis, the underlying concepts behind these systems are often not well understood. Artificial intelligence (AI) refers to computational systems that are capable of performing tasks that typically require human intelligence, such as reasoning, learning from experience, and problem-solving. In this sense, AI enables machines to replicate certain aspects of human cognitive processes. AI systems can take many forms, with different models designed for specific tasks and applications.

One important category of AI is the large language model (LLM). These models are trained on vast collections of text drawn from books, websites, and other written sources, allowing them to understand and generate human language. LLMs can perform a range of language-based tasks, including answering questions, generating written content, and engaging in conversational interactions.

A major subfield of AI research is natural language processing (NLP), which focuses on enabling computers to analyze, interpret, and generate human language. Applications such as ChatGPT and Google Gemini are examples of NLP-based systems, as they are designed to process user inputs and generate coherent, contextually appropriate responses.

Machine learning (ML) is a subset of artificial intelligence that relies on data-driven algorithms to identify patterns and make predictions or decisions. Rather than being explicitly programmed with fixed rules, ML systems learn from data and improve their performance as they are exposed to additional examples.

Machine learning methods are commonly divided into supervised and unsupervised learning. In supervised learning, models are trained on labeled datasets, meaning that the correct outcomes are known in advance and used to guide the learning process. In contrast, unsupervised learning operates on unlabeled data and seeks to uncover underlying patterns or groupings without predefined targets.

A foundational supervised learning technique is linear regression, which models an

outcome variable as a linear function of input variables (e.g., $Y = \beta 0 + \beta 1X1 + \ldots + \beta NXN + \varepsilon$). Extensions such as Lasso, Ridge, and Elastic Net regression introduce regularization to reduce

overfitting. More flexible supervised models include nonlinear methods such as logistic regression and support vector machines (SVMs), which can capture more complex relationships between variables. Classification algorithms assign observations to discrete categories based on decision rules; decision trees provide a simple implementation, while random forests combine many trees to improve predictive performance. Neural networks represent another class of supervised models, inspired by the structure of the human brain, in which data are processed through interconnected input, hidden, and output layers.
An important unsupervised learning approach relevant to this topic is clustering, which groups similar observations into clusters based on shared characteristics.

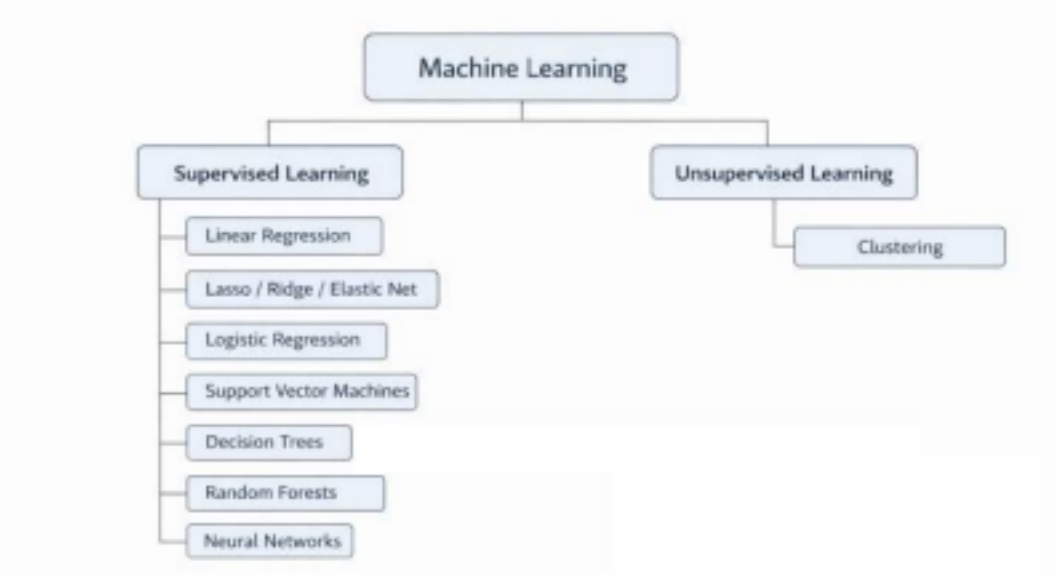


Fig 1: Overview of Machine Learning Models (Author's own compilation)

Application of AI and ML to Sports and Talent Recruitment

This section reviews existing academic literature on the application of artificial intelligence (AI) and machine learning (ML) to sports analytics and talent recruitment. The studies discussed below provide both empirical evidence and conceptual frameworks that inform how AI- and ML-based methods can be applied to college athlete recruiting.
Allal-Chérif et al. examine how AI and digital tools enhance recruitment processes across three stages: identifying, selecting, and retaining talent. Using multiple case studies—including LinkedIn, L'Oréal's *Reveal* platform, the Ari chatbot, and Randstad.tech -- the authors demonstrate how machine learning, semantic analysis, and natural language processing (NLP) improve matching between candidates and organizations. AI systems

automate applicant screening, assess soft skills through game-based evaluations, and predict candidate–organization fit. While the study focuses on corporate recruitment, its framework closely parallels college athlete recruiting, where performance data and behavioral indicators can be used to improve
matching between athletes and programs. The authors also caution against algorithmic bias and excessive reliance on data-driven profiling, which may overlook unconventional talent. Beal et al. analyze the use of AI and ML across team sports, identifying five primary application areas: match outcome prediction, tactical decision-making, player recruitment, fantasy sports, and injury prediction. The authors document the use of statistical models (e.g., Poisson and Bayesian regressions), support vector machines (SVMs), decision trees, random forests, and neural networks trained on player- and team-level performance data. Player recruitment and team formation are framed as optimization problems, in which teams use data from scouting, sensors, and performance tracking to identify undervalued players and construct lineups that maximize winning probability. The paper also highlights the use of deep learning and imitation models to study player movement patterns and simulate tactical scenarios, such as “ghost teams” in basketball and football. These findings are directly relevant to college recruiting, where similar data-driven methods can improve evaluation accuracy and recruiting efficiency.

Chmait and Westerbeek provide an accessible overview of how AI and ML have transformed sports analytics across both performance-related and business applications. The paper reviews AI use cases in game analytics, talent identification, training optimization, injury prediction, and fan engagement. Common techniques include artificial neural networks, Bayesian models, reinforcement learning, and clustering algorithms. The authors emphasize a broader shift from traditional rule-based analysis toward data-driven modeling approaches that learn patterns directly from large datasets. Importantly, the paper also highlights ethical and legal concerns related to data ownership, privacy, and the influence of algorithms on athletes and fans. The authors conclude that AI is unlikely to replace human decision-making but will increasingly serve as a complementary tool in coaching, athlete development, and sports management.

Pietraszewski et al. review how AI and ML are transforming talent identification and athlete development in team sports such as football, basketball, and soccer. The authors categorize AI applications into three areas: data collection and processing (e.g., wearable sensors and motion tracking), predictive modeling of player performance and potential, and decision-support tools for coaches and scouts. Common ML methods include neural networks, support vector machines, random forests, and deep reinforcement learning, trained on large datasets containing movement, physiological, and match statistics. The study highlights how AI can reduce human bias in scouting by providing objective, data-driven assessments, while also emphasizing emerging ethical issues related to fairness, transparency, and data privacy.

Scopinho focuses specifically on the use of artificial neural networks (ANNs) in athlete recruitment. ANNs are modeled after the structure of the human brain and consist of input, hidden, and output layers that learn from data through backpropagation. The paper reviews three empirical studies across different sports. In Australian football, ANNs ranked

under-18 players in the AFL Draft with accuracy comparable to professional recruiters. In cricket, ANNs outperformed human selectors in predicting player performance by up to 35 percent. In swimming, ANNs predicted Olympic 200-meter backstroke times within 0.05 seconds of actual

results. The author concludes that while ANNs are valuable decision-support tools, they cannot replace human judgment in evaluating non-quantifiable attributes such as leadership, teamwork, and chemistry.

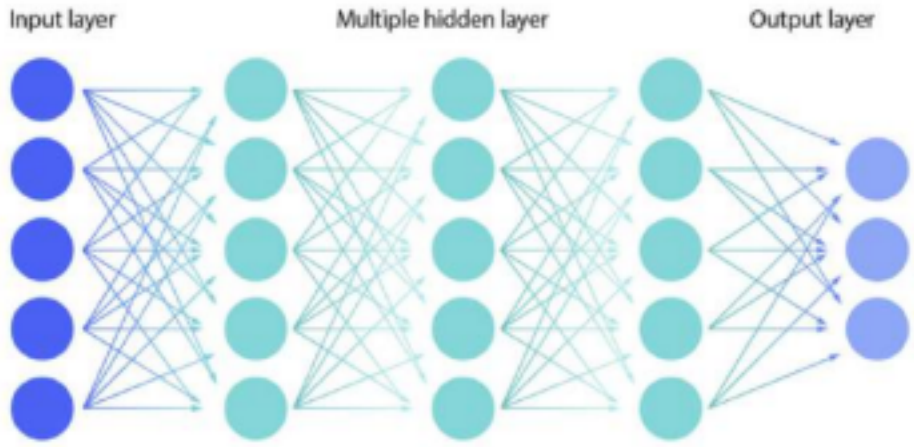


Fig 2: An Illustration of Artificial Neural Networks (ANN). In an artificial neural network, data goes through input layers, then through multiple hidden layers, and then through output layers. The outputs are the predictions (Source: IBM)

Finally, Cheremukhin explains the Mortensen–Pissarides (1994) search-and-matching model from labor economics, which describes how workers and firms form matches in the presence of search frictions and imperfect information. The model includes unemployed workers, vacant jobs, and a matching function that determines how efficiently new matches are formed. Frictions such as limited information, recruiting costs, and mismatched expectations lead to both unemployment and unfilled vacancies. This framework maps naturally onto college athlete recruiting, where athletes resemble job seekers, programs resemble firms, and recruiting operates as a two-sided matching market with uncertainty and search costs. AI and ML tools can improve matching efficiency by predicting player potential, identifying program fit, and reducing bias and search time—effectively increasing the efficiency of the matching process. This economic framework provides a theoretical foundation for analyzing how technology can improve recruiting outcomes in college sports.

Application of Artificial Intelligence and Machine Learning to College Athlete Recruiting This section first outlines how the traditional college football recruiting process operates and then discusses how artificial intelligence (AI) and machine learning (ML) methods can improve each stage of that process. Specific modeling approaches are highlighted, and the section concludes by connecting recruiting to the

search-and-matching framework from labor economics. While the focus is on college football, many of the mechanisms discussed are applicable to other sports with similar recruiting structures.

College football recruiting typically begins during an athlete's high school career and follows a multi-stage evaluation process. Athletes attend recruiting camps where university coaches and scouts assess performance, often alongside hundreds of other prospects. Recruiting platforms such as 247Sports rank athletes on a five-star scale derived from numerical evaluations, which frequently determine which players receive continued attention.

Following initial exposure, coaches study game film to evaluate technical skills, physical attributes, and decision-making. Coaches and scouts then contact selected athletes directly, while athletes may also reach out to programs of interest. In-person evaluations play an important role at this stage, with coaches traveling to high schools to observe athletes in live competition. Information on leadership, work ethic, academic habits, and personal background is often gathered through conversations with athletes, families, and high school coaches.

The final stage of recruiting involves competition among programs for top prospects through campus visits and scholarship offers. This stage typically culminates before or during an athlete's senior year. In recent years, Name, Image, and Likeness (NIL) opportunities have further intensified competition, as programs attempt to differentiate themselves through financial incentives.

Despite its importance, the recruiting process is highly time-intensive and prone to inefficiencies. Coaches and scouts must evaluate large numbers of athletes under time constraints, often relying on limited information from brief camps or isolated performances. This can lead to misclassification and unstable commitments. For example, Winn documents that approximately 16.5 percent of recruits in a sample decommitted at least once, highlighting the difficulty of identifying durable matches between athletes and programs.

AI and ML tools offer several opportunities to improve recruiting efficiency. One application is the enhancement of camp evaluations. Motion-tracking algorithms, automated speed and acceleration measurement, and skeletal tracking systems can provide standardized, objective assessments of athletic performance. These tools reduce human error and labor costs while allowing a broader pool of athletes to be evaluated consistently.

AI can also optimize communication and scouting logistics. Machine-learning–based prioritization and scheduling systems can help coaches allocate time and travel resources more efficiently, ensuring that attention is focused on the most promising prospects while reducing unnecessary scouting trips.

Game-film analysis represents another area where AI can add value. Computer vision techniques, pose estimation, and action-recognition models can automatically extract metrics related to movement, footwork, speed, and decision-making from video data. These tools reduce manual review time and generate objective performance indicators that complement traditional film evaluation.

Although soft skills such as leadership and work ethic are difficult to quantify, AI-based tools can provide partial insights through natural language processing (NLP), sentiment analysis, and structured behavioral assessments. Similar methods are already used in corporate hiring and may be adapted for athlete recruiting.

One of the most challenging aspects of recruiting is identifying strong matches between athletes and programs. Machine-learning–based recommendation systems can address this problem by comparing athlete characteristics—such as position, playing style, academic profile, and performance metrics—with program-specific preferences and historical recruiting patterns. Clustering methods and predictive models can identify which athletes are most likely to succeed within a given system.

From a modeling perspective, recruiting decisions can be framed using input–output structures. Inputs may include athletic statistics, game-film features, sensor or GPS data, and physical measurements, while outputs may take the form of binary recruiting decisions, star ratings, or continuous scores representing projected potential. Supervised models, such as logistic regression, can be used to predict recruit-or-not decisions or estimate commitment probabilities, offering interpretable results that highlight the importance of individual features.

More flexible models, such as support vector machines (SVMs), decision trees, random forests, and neural networks, allow for the identification of nonlinear patterns in complex datasets. Neural networks are particularly well-suited for video and image analysis, where they can identify performance patterns that may be difficult for human evaluators to detect. At the same time, overreliance on automated systems raises concerns, as many important athlete attributes—such as leadership, resilience, and teamwork—are not easily captured in data.

The recruiting process can be further understood through the Mortensen–Pissarides search-and-matching model, which shows that match quality improves when information becomes more accurate and search frictions are reduced. In this framework, athletes and programs operate in a two-sided market characterized by uncertainty and costly search. AI and ML tools reduce these frictions by improving evaluation accuracy, standardizing information, and facilitating better matches. In economic terms, AI increases the efficiency of the matching function, leading to more stable and mutually beneficial athlete–school relationships.

Conclusion

This paper examined the question: how can artificial intelligence (AI) and machine learning (ML) be applied to the recruitment of college athletes, particularly in football? By integrating insights from computer science, sports analytics, and labor economics, the paper demonstrated how these technologies can modernize and improve the traditional recruiting process.

The analysis showed that AI and ML have already transformed multiple areas of sports, including performance evaluation, injury prediction, tactical decision-making, and talent identification. The literature review highlighted that modern recruiting increasingly relies

on large datasets and predictive modeling to support decision-making.

Applying these insights to college football recruiting, the paper argued that AI and ML can improve nearly every stage of the recruiting pipeline—from evaluating athletes at camps, to analyzing game film using computer vision, to assessing soft attributes, and to matching players with programs based on historical data.

Specific modeling approaches, including neural networks, random forests, support vector machines, clustering algorithms, and recommendation systems, can help coaches make more accurate and objective recruiting decisions. These tools reduce search costs, mitigate certain forms of human bias, and allow programs to evaluate a larger pool of prospective athletes more efficiently.

The recruiting process was further framed using the Mortensen–Pissarides search-and-matching model, which characterizes recruiting as a two-sided market with significant frictions. Within this framework, AI and ML technologies increase matching efficiency, leading to more stable and mutually beneficial outcomes for athletes and programs.

The broader implications of integrating AI and ML into college recruiting include potential effects on competitive balance, athlete development, and resource allocation across programs. While early adopters of data-driven recruiting may gain temporary advantages, widespread adoption could eventually promote greater efficiency and fairness across the recruiting landscape.

Future research could examine how AI-based recruiting tools perform across different sports, how algorithmic bias can be minimized in athlete evaluation, and how data privacy and governance—particularly in light of NCAA Name, Image, and Likeness (NIL) regulations—should evolve as AI becomes more prevalent. Additional studies could also explore long-term outcomes, such as whether AI-assisted recruiting leads to higher retention rates, improved athlete–program fit, or stronger on-field performance over time. Understanding these dynamics will be essential as AI becomes an increasingly central component of college sports recruiting.

Works Cited

Adams, Anna. “247Sports Football Recruiting Rankings: What the Ratings Mean, When They Are Released, What Matters, and Who Is on the Team.” 247Sports, www.247sports.com. Allal-Chérif, Oihab, et al. “Intelligent Recruitment: How to Identify, Select, and Retain Talents from around the World Using Artificial Intelligence.” Technological Forecasting and Social Change, vol. 169, 2021, p. 120822.

Beal, Ryan, et al. “Artificial Intelligence for Team Sports: A Survey.” The Knowledge Engineering Review, vol. 34, 2019, p. e28.

Cheremukhin, Anton A. “Labor Matching: Putting the Pieces Together.” Federal Reserve Bank of Dallas Research Department Working Paper, no. 1102, 2011.

Chmait, Nader, and Hans Westerbeek. “Artificial Intelligence and Machine Learning in Sport Research: An Introduction for Non-data Scientists.” Frontiers in Sports and Active Living, vol. 3, 2021, p. 682287.

Pietraszewski, Przemysław, et al. “The Role of Artificial Intelligence in Sports Analytics: A Systematic Review and Meta-Analysis of Performance Trends.” Applied Sciences, vol. 15, no. 13, 2025, p. 7254.

Scopinho, Tales. “Artificial Neural Networks on Recruiting Athletes.” STEM Fellowship Journal, vol. 6, no. 1, 2021, pp. 76–80.

Winn, Luke. “The Commitment Project: A Study of Top-100 Recruit Behavior.” Sports Illustrated, 5 Aug. 2011.

Adolescent Depression and School Based Mental Health Interventions By
Roxanna A. Abdollahi, Hannah L. Dunn

Abstract
Adolescence is a period of time in which an individual undergoes rapid physical and emotional changes. According to the National Institute of Mental Health, in 2021, approximately 5 million adolescents between twelve and seventeen years old (20.1%) experienced at least one major depressive episode in the U.S.. There are significant disparities between depression rates and treatment rates, thus many adolescents with depression are not treated. Untreated depression can present severe impediments: lower performance in academic settings, suicidal thoughts or actions, and poor relationships. The purpose of this paper is to explore trends in depression among high school adolescents, the effects of school-based mental health interventions, and the perspectives and experiences of school providers on these issues. The aim of the present pilot study is to gather information that could potentially inform the development of more effective and accessible school-based strategies to support adolescent mental health. To collect data, an eleven-question survey was distributed to staff at a school in Northern California. The survey included a mix of long-answer questions and multiple-choice responses. All participants noted the prevalence of depression in high academic pressure environments, especially schools known for good academics. Certain groups of students, like the LGBTQ+ community and students of color were identified as having a higher susceptibility to depression by the participants of the study. Social media was identified as a main factor in the increase of depression rates.

Keywords Adolescence, Depression, School-Based Interventions, teachers & mental health providers

Introduction

Adolescence is a period of time in which an individual undergoes rapid physical and emotional changes. The major developmental tasks of adolescence include biological maturation, the development of personal identity, new cognitive and social skills, the establishment of independence, and psychological growth (Christie & Viner). Adolescence is not only a significant period of development, but also a vulnerable one, during which psychological disorders, such as depression, can emerge. This study will look at school-based interventions for adolescent depression in a high school setting, using data collected through a survey of school personnel.

Symptoms of Depression

The general symptoms of depression, as outlined by the Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition; DSM-5), are sad, empty, or irritable mood, along with physical and cognitive changes that affect an individual's ability to function and last for a specified period of time—two weeks for Major Depressive Disorder, and 6 months for Persistent

Depressive Disorder, for example (American Psychiatric Association). When comparing depression in adults to depression in adolescents, there are a few key differences. Adolescents generally have more anxiety, disruptive behavior, and somatic symptoms than adults. Their depressed mood may also manifest as crankiness or irritability. These symptoms are shared with other mental disorders, including hypothyroidism, anemia, sleep apnea, and other chronic diseases (Riglin et al.), which makes diagnosis substantially harder for clinicians. In addition, because adolescents are already undergoing intense emotional, hormonal, and physical changes that can lead to changes in sleep, appetite, or energy levels, it can be difficult to separate the difference between typical teenage behavior and a mood disorder. Understanding how adolescent depression presents is essential for identifying teens at risk and developing treatments accordingly, and thus, is an essential factor to consider as we think about school based mental health interventions.

Prevalence of Depression

In 2021, approximately 5 million adolescents between twelve and seventeen years old (20.1%) experienced at least one major depressive episode in the U.S. (National Institute of Mental Health, n.d.). Additionally, there is a clear discrepancy between depression rates and treatment rates, meaning many adolescents with depression are not treated: only 45% of adolescent females received treatment, while only 33% of adolescent males received treatment (Wilson & Dumornay). Given the importance of early intervention and accessible mental health resources for adolescents, it is important to reduce the treatment gaps. Untreated depression can present severe impediments: lower performance in academic settings, suicidal thoughts or actions, and poor relationships. The increase in the number of depression cases may be due to greater awareness of mental health issues and reduced stigma around seeking help, although an actual increase in the number of cases cannot be ruled out.

Causes of Depression

An individual's psychological health is shaped by their genetics and their environment. A prominent environmental factor that is correlated with depression is poverty. Studies have shown that individuals living in impoverished areas are 10-11% more likely to have poor mental health. This connection is complex because poverty can both contribute to and result from mental

illness. For example, depression can limit an individual's work opportunities, which can keep them trapped in a cycle of poverty. Poverty also puts children at greater risk for physical, behavioral, emotional, and social difficulties, which can affect their development and increase the chance of mental health problems later in life (Knifton & Inglis).

Parental depression has also posed another environmental risk factor for the development

of depression in adolescence. Depression in parents has been linked to increased frustration, hostility, and withdrawal while parenting (National Research Council (US) & Institute of Medicine (US) Committee on Depression, Parenting Practices, and the Healthy Development of Children). For example, mothers experiencing depression may find it harder to be attentive or emotionally available to their newborns, which can influence early bonding. Parental depression is also linked to cognitive and social delays and violent behavior in the child. Furthermore, preteen children of depressed mothers are more likely to blame themselves for negative events compared to children of non-depressed mothers, and adolescents with depressed parents often have lower self-worth and self-esteem (Myers et al.). Additionally, the U.S. National Research Council and the U.S. Institute of Medicine report that "parental depression has been consistently associated with a number of behavior problems and psychopathology in children, including higher rates of depression, earlier onset, longer duration, greater functional impairment, higher likelihood of recurrence, and elevated levels of anxiety, and both internalizing and externalizing symptoms" (139). Overall, individuals raised by depressed parents are more likely to have cognitive and social delays, develop violent behavior, and develop depression.

Parental depression also poses a genetic risk. Research indicates that depression can run in families due to inherited genetic factors that affect brain chemistry and function (Myers et al.). In particular, genes involved in the regulation of dopamine, glutamate, and stress hormones contribute to depression risk. Lohoff reported that "twin studies suggest a heritability of 40% to 50%, and family studies indicate a twofold to threefold increase in lifetime risk of developing MDD among first-degree relatives" (1). Thus, depression is a combination of genetic and environmental interactions, where inherited genetic risks increase a person's sensitivity to life stressors, such as trauma or difficult experiences. These factors work together to influence whether someone develops depression, confirming the interplay of nature and nurture.

School-Based Treatments for Adolescent Depression

Adolescents spend the majority of their time in school, which is why school-based interventions for depression are necessary. School is also one of the causes of poor mental health in adolescents, due to demanding school work and social interactions.

There are several different types of interventions for depression, including both targeted and universal interventions. Targeted interventions are when the intervention only involves students who have been found to have depression, such as therapy sessions during school hours. On the other hand, universal interventions are more of a preventative measure, provided to a large number of students who may or may not have depression. An example of a universal intervention is a lesson on mindfulness or having a better growth mindset. These interventions can be delivered by different personnel within a school community, such as teachers or mental health professionals. There is no clear consensus on whether teachers or clinicians are more effective at administering single-session interventions. Research suggests that both can be effective, and the optimal approach may depend on factors like the specific intervention, the target population, and the support environment (Zhang et al.).

When a cluster-randomized control trial was conducted across six high-performing schools, the results revealed that students in schools with wellness centers showed significant reductions in depression and anxiety symptoms compared to control schools (Yang).

Study Purpose & Research Question

Depression rates are increasing while the number of adolescents receiving treatment is lagging behind. Given that adolescence is a critical time in development when significant changes are occurring, it is imperative to address these challenges through improved prevention, diagnosis, and intervention efforts. The purpose of this paper is to explore trends in depression among high school adolescents, the effects of school-based mental health interventions, and the perspectives and experiences of school providers on these issues. The aim of the present pilot study is therefore to gather information that could potentially inform the development of more effective and accessible school-based strategies to support adolescent mental health and reduce the current treatment gap.

Methods

Participants

Three school therapists and one teacher from a public high school in Northern California consented to complete the survey and have their de-identified responses used for the purposes of this study. The high school has approximately 2,100 students enrolled, and a strong academic record - students score in the top percentiles on statewide exams, and the majority go on to attend four-year colleges across the country.

The six staff therapists at the school provide a range of mental health services to students in the school, including individual therapy sessions, drop-in group-based activities (such as mindfulness and art-based sessions), and crisis management, all within the school's on-site Wellness Center. Students are eligible for up to five free therapy sessions, and appointments are initiated through a confidential referral process, typically made by a teacher or school staff member. In addition to mental health services delivered by professional mental health providers, the Wellness Center is also home to peer-to-peer education programs aimed at promoting a positive campus climate and encouraging healthy choices.

Measures

To collect data, an eleven-question survey was distributed to school staff in a high school in northern California. The survey included a mix of long-answer questions and multiple-choice responses. The questionnaire was designed on Google Forms and administered via an email link to participants.

The questionnaire was developed based on a review of the literature on adolescent depression and school-based mental health interventions. An initial pool of 30 questions

was consolidated to 11, organized around key areas of focus for this paper (i.e., trends in depression rates, identification of depression, impact of social media, and programs addressing depression). Participants were asked to share their observations and perspectives on a range of issues, including the prevalence of student depression, contributing factors, identification and signs of

depression, changes in how mental health is discussed, and school-based intervention strategies. See *Appendix A* for a list of the questions presented to participants.

Procedure

The lead researcher first contacted the counseling center, and visited in person to gauge interest in participation. For those who expressed interest, a link to the questionnaire was sent via email. Participants who completed the survey and provided consent for de-identified responses are included in the analysis and presented in the paper.

Data Analysis Plan

All responses were reviewed by the lead author. Emergent themes were identified through a thematic qualitative analysis process, and responses were organized according to these categories.

Results

Results are presented below according to theme.

Theme 1: Rates of Depression & Certain Groups of Students

In response to questions about prevalence of depression, all participants (therapists and the teacher) reported observing an increase in the rates of depression among adolescents of all high school ages over the last several years. Therapists attribute the increase in observed rates of depression to a number of factors. One reason stated is that there is better psychoeducation and resources available to students, as well as the de-stigmatization of mental health disorders, which contributes to personal awareness and willingness to seek help, as therapist #1 says: *"Mental health is slowly becoming less taboo of a subject, which has positively impacted students' ability to seek help."* Additionally, a teacher noted: *"When I first started teaching 30+ years ago, the saying was "just get over it". That has all but diminished, and now we have a Wellness Center and therapists on site every day for support."* There is nuance to rates of depression and there may be an increase in depression among adolescents due to other factors as well.

The therapists and teacher surveyed in this study also noted that LGBTQ+ students and students of color exhibit higher rates, or are more vulnerable to depression across all grade levels (9th through 12th): "*we see an increased rate of depression in students who are part*

of any minority population or marginalized groups" (Therapist #3). One therapist attributed higher depression rates among minority students to the social political climate: "*rates of anxiety and depression has increased within LGBTQ+ and BIPOC* [Black, Indigenous, and People of Color] *folks due to feelings of uncertainty regarding their rights.*" Another therapist noted the same: *"The cultural-political climate can lead to an increase in depression, particularly for vulnerable populations, including but not limited to the LGBTQ+, BIPOC, undocumented, and SPED communities"* (Therapist #2). This same therapist elaborated on this comment:

"I notice that anti-LGBTQ+ bias, racism, and ableism greatly increases risk for depression and safety concerns for LGBTQ+, students of color, and students with disabilities, respectively. I notice an increase in depression for students with disabilities, particularly learning disabilities, Autism Spectrum Disorder, ADHD, emotional disturbance, and anxiety-related disorders related to challenges with accessing the curriculum without proper accommodations."

This data indicates that more specialized resources need to be devoted to these students, as they are more vulnerable to depression.

Theme 2: Academic Pressure & Social Media As Contributing Factors to Depression

All participants discussed academic pressure and social media as key factors in contributing to student depression. In terms of academic pressure, participants noted the particularly rigorous school environment: *"Academic pressure has been very prominent, especially in the region we live in"* (therapist #3). Another therapist (#1) noted that: *"Academic pressure and cultural components of academic success can create significant stress/anxiety which then leads to feelings of failure, exhaustion, and depression when standards are not met"* (therapist #1). Furthermore, therapist #2 described how academic pressure can stem from internal, and external resources, and can lead to detrimental outcomes: *"Academic pressure, which often comes from a mixture of students, parents, and school culture/expectations, often increases stress and can lead to disengagement in self-care, which can lead to an increase in depression symptoms if not properly managed."* A teacher described how pressure to gain college acceptance can drive student anxiety and can also be inflated by social media: *"The competition for college acceptance has blown up by posts on social media."* Additionally, social media was noted as a contributing factor to the increasing rates of depression across all participants' answers. Each participant noted different aspects of social media that can play a role. For instance, therapist #1 described how social media inundates students and can impact how they see the world and their sense of identity at particularly vulnerable stage of development:

"Social media may play an important role in adolescent depression. Students' brains are developing while they take in non-stop (and not-so-regulated) information being fed to them through multiple platforms. Social media may be impacting their world views, self-esteem, and sense of self at a different pace and rate than other

generations."

Another therapist (#2) described how social media can open the door for poor coping and cyber-bullying:*"Students often use social media as a maladaptive coping skill, limiting access to more evidence-based emotion regulation, distress tolerance skills, etc. Social media can also increase vulnerability to cyberbullying."*

A teacher also noted how social media impacts adolescent's communication and relationships:
"Generally speaking, students talk to each other MUCH less than years ago - before social media. I used to have to discipline so much monkey-business during class! Now I have to remove their phones to get them to talk to each other (a bit of hyperbole, but not really!)."
Through the testimonials given by school staff, academics and social media are adding stress, leading to a need for more support related to these integral aspects of students' lives.

Theme 3: Interventions
All therapists and the teacher surveys described looking for similar warning signs (isolative behavior, acting out) to determine when a student might need intervention. The teacher explained that a sign a student might be struggling is when, *"They do not engage with others, or*
they cover it by being overly outgoing. Any "extreme" behavior is a red flag for me."
Therapist #3 stated, *"A student may begin to isolate themselves, present as withdrawn, decrease in motivation."*
All the surveyed participants noted that the Wellness Center has been an important development in the dissemination and accessibility of mental health resources and interventions for students: *"there have been some improvements in recognizing the impact mental health has on students over the last four years with the introduction of the Wellness Center" (therapist #2)*. In particular, the teacher surveyed described specific aspects of services offered at the Wellness Center that have been effective so far in her observation: *"The Wellness Center, licensed therapists for counseling, and of course, the therapy dogs!"*
Outside of the Wellness Center, teachers and therapists described state-wide policies, school based policies, as well as a "culture of support" among staff and students that have all served as effective intervention and prevention efforts. For instance, the surveyed teacher explained how she sees a new California state law as being beneficial by reducing student cell phone access in school and thereby promoting healthy relationships and interactions among high schoolers:
"I appreciate the new CA law (effective the 2026-27 school year) of banning cell phones during class time. At least that requires adolescents to be off of social media and

interacting with their peers for most of the school day."

` Additionally, all participants described the benefits of their efforts to foster a school climate that emphasizes open, honest conversations among faculty and with students, as well as regular check-ins. One therapist explained how these kinds of positive faculty-student interactions are essential in offering support to high schoolers: *"Have honest conversations with those they're concerned about, emphasize that you're coming from a place of love and concern."* (therapist #1). Similarly, a teacher described the effectiveness of this approach: *"A one-on-one check-in with students on a regular basis would be prudent. Just letting the student know that someone is paying attention and cares about them goes a long way!"* (teacher). Therapist #3 also emphasized the need for students to feel that staff members are accessible and approachable: staff: *"I think all staff can do their best to be approachable and continue to build rapport with their students so students feel comfortable to be vulnerable."* In a similar vein, therapist #2 described how these conversations between faculty and students can also incorporate essential psychoeducation: *"By providing them with psychoeducation regarding the impact screen time has on their mental health and by teaching them more appropriate skills, such as DBT skills, that could support them in better managing their mental health symptoms."* Together, these insights highlight the importance of consistent, caring relationships and open communication between staff and students as foundational to early intervention and mental health support in schools. See *Appendix B* for a table of themes and main points.

Discussion

Overall, the participants noted the prevalence of depression in a high academic pressure environment, especially in schools known for good academics. The expectations set by their peers, teachers, and parents can make students feel like they are inferior to others. Certain groups of students, like the LGBTQ+ community and students of color were identified as having a higher susceptibility to depression by the participants of the study. Social media was identified as a main factor in the increase of depression rates: participants described how social media can impact how they see the world and their sense of identity at a particularly vulnerable stage of development. All the surveyed therapists and teacher notice that the Wellness Center is a valuable resource for interventions, as well as identifying the need for better psychoeducation and open conversations with students.

Limitations

There were several limitations to this study. There was a small sample size due to the lead researcher only having access to their own school, which limits the generalizability of the findings. However, as there are many schools with similar high pressure environments and the worldwide prevalence of social media, these results can still be generalized. Since the lead researcher is a student at the surveyed school, there may be a bias in the thematic interpretation.

Future directions

In terms of future directions, analyzing the difference between the perspectives of school personnel and of students, as well as parents/guardians, on these issues would provide valuable insight. It would also be beneficial to gather data from other schools in California, rather than only one. The data collected in this study can assist in addressing the need for culturally informed practices, geared towards helping students from different backgrounds and with different identities. As presented in this study, students from certain minority backgrounds are at higher risk, which suggests the need for more awareness, more identification effort, and more

intervention efforts targeted to these students. Furthermore, students from districts with lower social economic class may have other factors affecting mental health.

Appendix A
Distributed Questionnaire

Question	Response Format
1. What trend have you observed in adolescent depression in our school over the past few years?	A) Increasing B) Decreasing C) Staying About the Same
2. Research has shown that nationwide rates of depression are increasing. From your experience, are there any particular factors (e.g., academic pressure, social media, family dynamics, cultural-political climate) that you think seem to be contributing to the increase in depression among adolescents?	A) Academic Pressure B) Social Media C) Family Dynamics D) Cultural-Political Climate E) Other

	3. Please explain your answer from the previous question 4. Do you notice differences in depression rates across different age groups or grade levels? Do you notice certain trends among certain groups of students (e.g., LGBTQ+ students, students of color, students with disabilities)?	Open Response (Short Answer) Check Box Respondents were asked to select (check) off various historically minority groups across all four grades of high school. Check box grid: ● Asked to check off LGBTQ+ students, students of color, students with disabilities, other - for each grade 
	5. Please explain your answer from the previous question. Also, if you clicked the 'other' option, please expand on what group of students you are referring to!	Open Response (Short Answer)
	6. What are some of the key signs that a student may be struggling with depression but is not openly expressing it?	Open Response (Short Answer)
	7. How has the conversation around mental health	Open Response (Short Answer)

evolved in the school environment, and do you feel it has had an impact on depression trends in students? What resources or support systems do you think schools need to improve upon to better address adolescent depression?	

8. What role do you think social media plays in adolescent depression, and have you seen any noticeable correlations between social media use and depressive symptoms in students?	Open Response (Short Answer)
9. How do you address the potential effects of technology and social media with students who may be struggling with their mental health?	Open Response (Short Answer)
10. What are some measures or programs that have been effective in treating depression among adolescents at our school?	Open Response (Short Answer)
11. How can educators, school staff, and parents help identify and support students at risk of depression, even if they haven't sought help yet?	Open Response (Short Answer)

Appendix B
Response Themes

Theme	Main Points
1) **Rates of Depression & Certain Groups of Students**	● LGBTQ+ and students of color have higher rates of depression. ● Mental Health is becoming destigmatized, so students feel more comfortable reaching out for assistance.

2) Academic & Social Media **3) Interventions**	● High academic pressure environments contribute to adolescent depression. ● Social Media limits in-person peer interactions as well as having other negative consequences (e.g., bullying, false perceptions, poor coping), which can contribute to social isolation and depression. ● Benefits of Wellness Center (i.e., therapy dog, counselors on staff, programming and resources). ● Need for open conversations & psychoeducation to support students.

Works Cited

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Edited by American Psychiatric Association, American Psychiatric Association, 2013.

Knifton, Lee, and Greig Inglis. "Poverty and mental health: policy, practice and research implications." BJPsych bulletin 44.5 (2020): 193-196.

Lohoff, Falk W. "Overview of the genetics of major depressive disorder." Current psychiatry reports 12.6 (2010): 539-546.

Myers, Taylor L., Tracy RG Gladstone, and William R. Beardslee. "The transition to adulthood in children of depressed parents: long-term follow-up data from the family talk preventive intervention project." International journal of environmental research and public health 20.4 (2023): 3313.

National Institute of Mental Health. “Major Depression - National Institute of Mental Health (NIMH).” National Institute of Mental Health, https://www.nimh.nih.gov/health/statistics/major-depression. Accessed 7 August 2025.

England, Mary Jane, Leslie J. Sim, and National Research Council. "Associations between depression in parents and parenting, child health, and child psychological functioning." Depression in parents, parenting, and children: Opportunities to improve identification, treatment, and prevention. National Academies Press (US), 2009.

Wilson, Sylia, and Nathalie M. Dumornay. "Rising rates of adolescent depression in the United States: Challenges and opportunities in the 2020s." Journal of Adolescent Health 70.3 (2022): 354-355.

Yang, Rick. "59. The Impact of School-Based Wellness Centers on Student Mental Health and Academic Performance: A Mixed-Methods Longitudinal Study in High-Achieving Schools." Journal of Adolescent Health 76.3 (2025): S31.

Zhang, Qiyang, Jun Wang, and Amanda Neitzel. "School-based mental health interventions targeting depression or anxiety: A meta-analysis of rigorous randomized controlled trials for school-aged children and adolescents." Journal of youth and adolescence 52.1 (2023): 195-217.



Nazarbayev Intellectual Schools

“Óskin” – A Smart Incubator Supporting a Favorable Environment for Plant Seed Germination

Project Research Paper

Compiled by:

Najmudin Adikhanov

Semyon Chistyakov

Assylzhan Sailau

Aidana Kobentai

Alinur Akezhanov

Alikhan Zhetpis

Scientific supervisor: Moldabayeva Z.S.

Table of Contents

- Abstract
- Introduction
- Relevance and Rationale of the Topic
- Expert Opinion
- Aims and Objectives
- Materials and Methods of Research
- Project Realisation
- Hardware
- Software
- Testing
- Competitors
- Results
- Future Plans
- Reference List

Abstract

This project was created by students of Nazarbayev Intellectual Schools: Najmudin Adikhanov, Semyon Chistyakov, Assylzhan Sailau, Aidana Kobentai, Alinur Akezhanov, and Alikhan Zhetpis. Excess moisture, aridity, high and low temperatures, the formation of pus and salts, as well as seed diseases lead to the problem of non-germination of plant seeds. The agricultural sector of Kazakhstan suffers losses due to the loss of seeds.

The study examined errors in seed cultivation, the problem of seed non-germination, factors affecting germination, possible methods of regulation and control using technical solutions, the impact on the seed market, and the strengths and weaknesses of existing solutions. As a response, the project proposes a smart incubator for germination of plant seeds.

The implemented incubator, "Óskin," regulates humidity inside the seed compartment, monitors indicators using sensors, and gives the user access to the device through a Telegram bot.

Keywords: seeds, germination, agriculture, seed non-germination, smart incubator, humidity control, plants

Introduction

Various environmental factors directly affect the process of seed germination. In traditional methods, these factors are mostly not automatically controlled and require significant effort to adjust. This creates seed wastage, which is detrimental to the agricultural market of the country.

This project is designed to create an incubator that regulates and monitors the conditions of seed germination. It facilitates the entire seed germination process and can reduce cost by lowering seed loss and improving germination efficiency. "Óskin" connects robotics, engineering, programming, and biology to solve complex problems in the agro-industry.

Relevance and Rationale of the Topic

Kazakhstan is ranked among the countries facing serious soil degradation problems. The key factors of seed germination are temperature, humidity, planting depth, soil quality and type, and access to light. Their adjustment is difficult using traditional germination methods such as direct planting in soil or germination in moist gauze. These methods usually require 5–7 days for seed germination.

For effective seed germination, environmental values should remain within acceptable limits and be regulated continuously. Seed collection also often takes place at less than ideal humidity. There is a risk of seed diseases, pests in the soil, and the formation of pus and salts during soil-based growing. These aspects may cause the loss of 30–40% of seeds, directly affecting the national seed market.

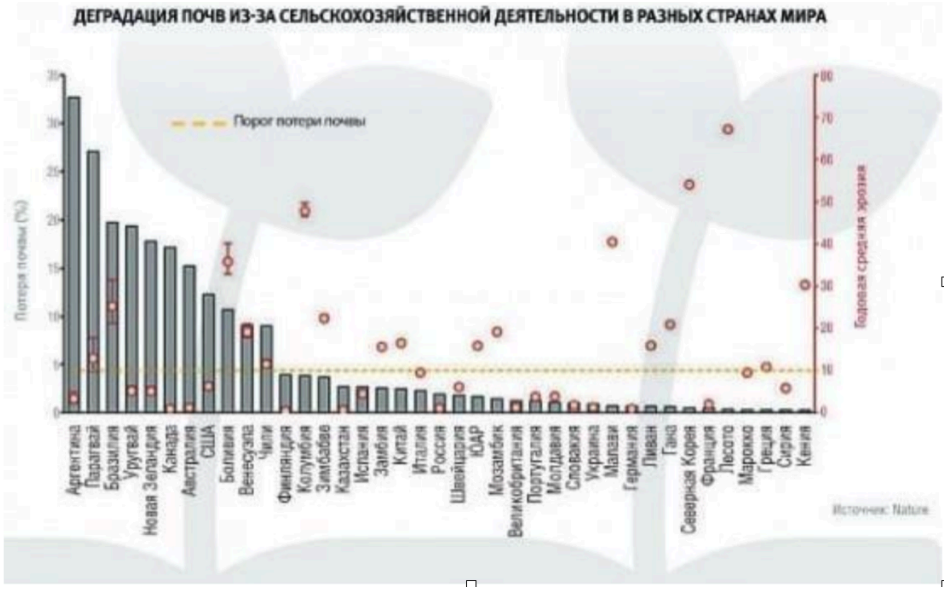


Fig. 1. Soil degradation statistics and agricultural pressure indicators.

With "Óskin," the regulation of these factors becomes easier. The device reduces the time required for seed nurturing and minimises the degree of direct human impact on the germination process.

Expert Opinion

The team conducted an interview with Anton Isaev, an expert in agricultural science with 15 years of experience and CEO of the agricultural holding SwissGrow. He confirmed the relevance of the problem and emphasised that harvesting takes place at a moisture content of 9–10%, while the ideal moisture content is approximately 14.2–14.4%. Anton Isaev noted that the problem of seed non-germination and inefficient germination affects the food industry of Kazakhstan and that the incubator can contribute to solving this issue.

Aims and Objectives

Project Goal: to develop a seed incubator with a smart system for regulating and monitoring indicators for efficient germination of plant seeds, controlled by the user through a Telegram bot.

Project Objectives:

1. Study the factors affecting seed germination.
2. Study ways of adjusting these factors.
3. Assemble and construct the incubator.
4. Write the code for the controller board.
5. Create a Telegram bot providing communication between the board and the user.
6. Test the incubator.

Materials and Methods of Research

The research began with the formation of a detailed picture of the problem. For structured work, the team created a research plan:

1. Study factors influencing seed germination.
2. Study existing methods of their regulation.
3. Study the best and most effective methods and their principles of operation.

Information was collected through open internet sources, e-books, interviews with farmers, and consultations with agroholdings.

Main germination factor	Role in the project
Humidity	Identified as the most important key factor and selected as the main controllable variable.
Temperature	Can be regulated using heating elements and lamps.
Light access	Can be supported through lighting and UV lamps in future versions.
Soil type and quality	Difficult to regulate quickly, which motivated the soil-free germination approach.
Pests and seed diseases	Reduced by avoiding soil-based germination.
Planting depth	Removed from the prototype process by using gauze-based placement.

The most important factor turned out to be humidity. The most effective regulation method was identified as the ultrasonic steam generator. It works by vibrating a piezoelectric plate at ultrasonic speed, dispersing water molecules and forming vapour. This method has lower physical wear, lower energy consumption, and can convert large volumes of water into vapour in a short time. It also does not generate strong mechanical vibrations or audible sound.

Temperature and access to light may be provided by incandescent lamps, heating elements, and ultraviolet lamps. Ultraviolet rays can also positively affect plant growth by stimulating chlorophyll synthesis and carotenoid production. Soil type cannot be changed quickly, while soil quality improvement and pest control often require additional preparations that may not always benefit the seeds. Therefore, the project focuses on a soil-free method.

Project Realisation

The team concluded that since humidity is the most important aspect, the prototype should focus on regulating moisture. The project therefore avoids soil during germination, eliminating many of the difficulties and risks associated with soil, including pest damage and disease.

Óskin’s main task is to regulate humidity inside the incubator without using soil. A second task is to monitor indicators and ensure communication with the user through a Telegram bot. Airtightness is

also important because it helps maintain the internal microclimate.

Hardware

The logical operations, sensor readings, output values, and Wi-Fi communication with the Telegram bot are handled by a NodeMCU ESP8266 1.0 board. It was selected because it has high computational power, compact size, low price, and a built-in ESP8266 microcontroller for Wi-Fi communication. Compared with Arduino UNO, it can be plugged directly into a breadboard, saves space, and already includes Wi-Fi functionality.

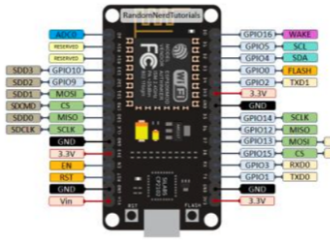


Fig. 2. ESP8266 board.



Fig. 3. DHT22 temperature and humidity sensor.

The temperature and humidity sensor chosen for the project was the DHT22. It was selected because it has better accuracy than its competitors. It reads up to 100% humidity and has an accuracy of approximately ±2% and ±0.5°C. The lower-series DHT11 has a larger error and lower measurement range, which is less suitable for this project.

The ultrasonic steam generator performs the humidity control function. The prototype currently uses a 7-litre food container and a Polaris PUH 2650 humidifier attached to the bottom. The humidifier includes two vapour intensity modes, prevention of piezo plate combustion through a water sensor, and vapour blowing using a fan.

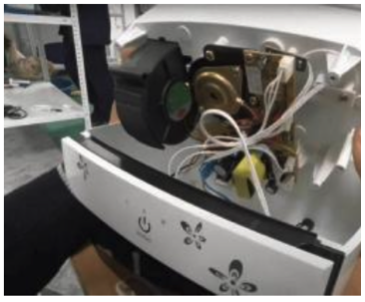


Fig. 4. Humidifier construction from inside.

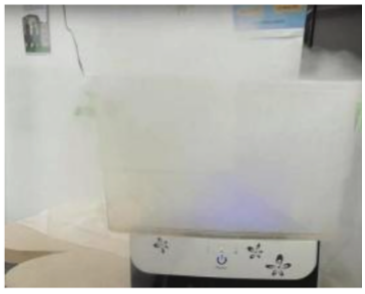


Fig. 5. Operating incubator with vapour generation.

The team also designed an electronic circuit for an improved version of the incubator. This version would have power supply from a single socket, full control through the ESP8266, and a light sensor.

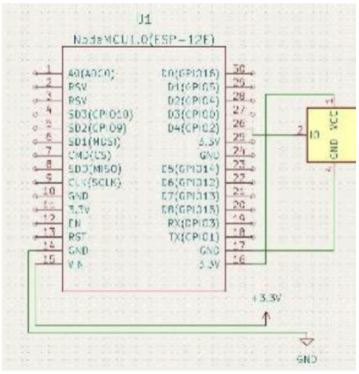


Fig. 6. Wiring diagram of the current version.

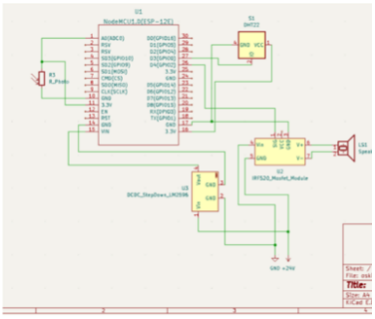


Fig. 7. Electronic schematic of the improved version.

In the Óskin prototype, seeds are placed on gauze fixed with the lid of the container. The incubator consists of two parts: an upper compartment and a lower compartment. The upper part includes the 7-litre container and gauze, while the lower compartment contains the disassembled humidifier that converts water into vapour. The board and sensor are attached to the side, outside the container.

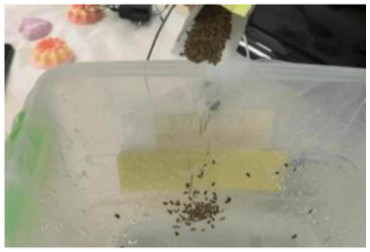


Fig. 8. Seed filling process.



Fig. 9. Operating incubator.

Fig. 10–11. Sensor and board mounting: outside and inside views.

A PCB board and 3D model were also designed for the future improvement of the project. At the current stage, the electronics are implemented on a breadboard and powered by two sockets. The improved version will include one power supply, a PCB board, and direct access to incubator control. In the future, the gauze fixation method may be replaced with a system of mesh trays, providing several floors for seed placement and increasing the maximum number of germinated seeds. These trays can be printed on a 3D printer.

Software

The program code was written in Arduino C. Sensor libraries and Wi-Fi functionality were used to organise communication with the Telegram bot. The bot itself was created using the BotFather service. Since all functions are written through the ESP8266, the team only needed to register the bot in

Telegram and connect the board to the bot through the code.

Software element	Function
Arduino C code	Controls board behaviour, sensor reading, and device logic.
Wi-Fi connection	Allows ESP8266 to communicate with the internet.
Telegram bot	Provides user access to humidity and temperature data.
Sensor libraries	Enable correct reading from the DHT22 sensor.
Message frequency settings	Optimise communication and avoid unnecessary repeated messages.

Testing

During testing, 100% of the seeds matured in 3 days. This was significantly faster than the usual 5–7 days required in traditional methods. The photo reports in the original project showed seed germination progress across the first several days of testing.

Method	Germination time	Observed result
Traditional germination	5–7 days	Approximately 70% germination
Óskin prototype	3 days	100% seed maturation in testing

Competitors

For the regular version of the project, the main competitors are propagators. They are lower-cost but mainly create a greenhouse effect inside the container. This does not significantly improve germination efficiency and still requires soil, which does not prevent seed diseases or pests.

For the industrial version, the main competitors are Chinese analogues and existing seed propagators. Compared with these competitors, Óskin aims to combine humidity control, monitoring, user communication, and a soil-free design.

Solution	Core function	Weakness	Óskin advantage
Basic propagator	Greenhouse effect	Limited improvement, still uses soil	Humidity regulation and soil-free germination

Industrial analogue	Large-scale incubation	Can be expensive or less flexible	Potential scalable smart version
Traditional moist gauze	Simple germination	Requires manual monitoring	Automated monitoring and faster results

Results

After analysing the test results, the team concluded that the project is workable and performs effectively at the MVP stage. The indicators of germination percentage and germination time are better compared with traditional methods: 3 days and 100% germination in the prototype test compared with the usual 5–7 days and approximately 70% germination in traditional methods.

The project therefore offers a practical technical solution to complex seed market problems in Kazakhstan by reducing seed loss, lowering the need for manual control, and improving the stability of germination conditions.

Future Plans

- Build an improved version of the incubator.
- Build an industrial version.
- Add lighting control with UV lamps.
- Add more sensors.
- File for a patent or utility model.
- Refine the project to full automation.
- Add water supply.
- Print and apply the PCB board.
- Implement a new power supply system.
- Realise the pallet system.

Reference List

1. Влияние некоторых факторов на прорастание семян и развитие растений // Старт в науке. URL: https://school-science.ru/4/1/594 (accessed 08.10.2023).
2. Вначале было семя. Золотые правила проращивания // Dzaki Grow. URL: https://dzagigrow.ru/blog/vnachale-bylo-semya-zolotye-pravila-prorashchivaniya/ (accessed 08.10.2023).
3. Гордость казахов – земля – под угрозой деградации // Exclusive. URL:

https://exclusive.kz/expertiza/obshhestvo/124666 (accessed 08.10.2023).
4. Койшыбаев, М. Болезни пшеницы. Ankara: FAO, 2018. ISBN 978-92-5-130142-5. http://old.belal.by/elib/fao/1032.pdf
5. Николаев, М.Г. Справочник по проращиванию покоящихся семян. Leningrad: Наука, 1985. http://flower.onego.ru/agro/seeds01.html
6. Правильное проращивание семян // Grow Lab. URL: https://www.grow-dv.ru/blogs/blog/pravilnoe-proraschivanie-semyan-metod-h (accessed 08.10.2023).
7. Прорастание семян, условия его прорастания // Агровита. URL: https://agrovita.org.ua/ru/92-prorostannya-nasinnya-umovi-jogo-prorostannya/ (accessed 08.10.2023).
8. Сколько дней прорастают семена? Таблицы // Бабушкина Дача. URL: http://babudacha.ru/2018/02/07/skolko-dnej-prorastayut-semena/ (accessed 08.10.2023).
9. Температура прорастания семян: какая подходит для разных растений // OrganMix. URL: https://organic-mix.ru/articles/temperatura-prorastaniya-semyan/ (accessed 08.10.2023).
10. Parimala, K., Subramanian, K., Mahalinga Kannan, S., & Vijayalakshmi, K. A Manual on Seed Production and Certification. Chennai: CIKS, 2013.
11. Borelli, P., Robinson, D. A., Fleischer, L. R., et al. An assessment of the global impact of 21st century land use change on soil erosion. Nature Communications, 2017, 8, 1–13.
12. Batrinu, C. ESP8266 Home Automation Projects. Birmingham: Packt Publishing, 2017.
13. Stubbs, R. W., Prescott, J. M., Saari, E. E., & Dubin, H. J. Cereal disease methodology manual. Mexico: CIMMYT, 1986.
14. Dalmaris, P. KICAD like a pro. 2nd ed. Australia: Tech Explorations, 2018.
15. ESP8266 Pinout Reference: Which GPIO pins should you use? Random Nerd Tutorials. URL: https://randomnerdtutorials.com/esp8266-pinout-reference-gpios/
16. ESP8266. НАЧАЛО РАБОТЫ, ОСОБЕННОСТИ // AlexGiver Technologies. URL: https://alexgyver.ru/lessons/esp8266/
17. Germination // Wikipedia. URL: https://en.wikipedia.org/wiki/Germination
18. Gibilisco, S. Beginner’s Guide to Reading Schematics. 3rd ed. New York: McGraw Hill TAB, 2013.
19. Monk, S. Programming Arduino: Getting Started with Sketches. 2nd ed. New York: McGraw-Hill, 2016.
20. Baxter, L. L., Grey, T. L., Tucker, J. J., & Hancock, D. W. Optimizing Temperature Requirements for Clover Seed Germination. Agrosystems, Geosciences & Environment, 2019.
21. Platt, C. Make: Electronics: Learn by Discovery. Sebastopol: O’Reilly Media, 2009.
22. Raven, P. H. Biology of Plants. 6th ed. New York: W.H. Freeman, 2005.
23. Srivastava, J. P., & Simarski, L. T. Seed Production Technology. Aleppo: ICARDA, 1986.

Uncovering Correlations Between the Length and Type of Parental Leave Taken by Elementary, Middle, and High School Parent-Teachers in Virginia and Teachers' Sense of Responsibility Towards Their Students By Joy Svendsen, Andrew McCarthy

Abstract

Parental leave is a constantly evolving field, and there is a lack of research on demographic data for teachers who recently took parental leave. This study collects both demographic data and teachers' opinions on their school's policy and support systems. A quantitative survey is administered to teachers who have taken parental leave in the past five years. Data from the survey is used to make comparisons between the type and length of leave the teacher took, what level of school and type of class the teacher taught, and the level of responsibility the teacher felt for their students in the teacher's absence. Not enough varied data was collected to analyze the type of leave taken by teachers, but it was found that elementary and elective class teachers take the longest leave, while middle school and core class teachers take the shortest leave. The importance of a good parental leave policy and school support system are confirmed in the results, and areas for future research are found because of this study.

Keywords Parental Leave, Teachers, Sense of Responsibility

Introduction

Parental leave is crucial for the health of both the developing newborn and the parent, allowing them to bond and adjust to a new lifestyle. Parental leave is a blanket term for maternity, paternity, and adoption leave. Maternity leave is taken by the mother when she needs time to prepare for and recover from giving birth to her child; while paternity leave is taken by the non-birthing parent as time to care for and bond with the newborn; finally, adoption leave is taken by those adopting or fostering a child. Since 1993, the United States Family and Medical Leave Act has ensured eligible workers job stability, with 12 weeks of unpaid leave (Shabo, 2026). Parents can use this policy to take time off from their job to care for their newborn; however, they do so at the expense of their own pay.

As of 2026, thirteen states have additional mandatory state paid parental leave programs to further improve the health of both parent and child (Shabo, 2026). Virginia is not one of those thirteen states, as it lacks a mandated paid parental leave policy for private employees. However, Virginia does have a paid parental leave policy for state employees, offering those employed for a minimum of 12 months a total of 8 weeks of paid parental leave (Baker Donelson, 2026). Teachers are state employees who benefit from this rule, though policy varies from county to county within Virginia. Although teachers are given the opportunity to take parental leave, many teachers feel a sense of responsibility for their students' learning; and shorten their leave to return to their students faster. When teachers lack confidence in their school's ability to assist student learning in their absence, it could cause teachers to stress about their students at school in addition to their newborn at home. Each teacher feels a different level of responsibility towards

their students' grades, learning, and well-being. Feeling a greater sense of duty to their students could cause teachers to take a shorter amount of leave (Dunn et al, 2025). This study aims to uncover the correlations between the length and type of parental leave taken by elementary, middle, and high school parent-teachers in Virginia and teachers' sense of responsibility towards their students.

Literature Review

Ideal Length of Parental Leave

Parental leave is important for parents and their newborn, and it is important to consider what the ideal length of parental leave should be. In 2023, Madeline Dixon Whitney and others from the Maternal and Child Health Journal conducted a systematic review and meta-analysis of 21 studies. The goal of Whitney's study was to find how the amount of maternity leave taken by mothers affected the mental and physical health of mothers and infants. Their analysis found that taking longer maternity leave "may decrease rates of maternal mental and physical health complaints". It was found that taking longer maternity leave had many benefits, such as increasing positive interactions between mother and child, decreasing infant mortality, and leading to longer periods of breastfeeding (Whitney et al 2023). Overall, Whitney found that taking 12 or more weeks of maternity leave was the most beneficial to the mother and newborn.

Fathers' experiences vary vastly from mothers' experiences, so, naturally, the ideal length of paid maternity versus paternity leave will be different. A study by Katharine M. Barry and other researchers from Pierre Louis Institute of Epidemiology and Public Health examined whether fathers taking two weeks of paid paternity leave impacted post-partum depression in mothers and fathers two months after the birth of their child. Barry found that fathers taking two-weeks of paid paternity leave were "associated with a reduced likelihood of reporting post-partum depression" (Barry et al, 2023). Although this research proves taking paternity leave reduces depression in fathers, Barry's research also reveals that fathers taking two-weeks of paternity leave "might place mothers at a greater risk of post-partum depression" (Barry et al, 2023). While Whitney's research recommends 12 weeks of maternity leave for the mother and newborn's health, Barry's research shows that taking paternity leave can impact the mother's mental health. Overall, taking parental leave improves both the parents' and the infants' health.

Maternity leave has many benefits to the mother transcending mental health. In a study by researchers Judith Galtry and Paul Callister published in *The Journal of Family Issues*, Sweden and the United States' policies are compared to determine the optimal length of parental leave. The study found that mothers taking more than 12 weeks of parental leave reported greater vitality, more than 15 weeks led to better mental health, and more than 20 weeks led to fewer limitations to women's daily social role and equality (Galtry & Callister, 2005). Between Whitney's study and Galtry and Callister's study, 12 weeks seems to be a

minimum for the ideal length of parental leave, especially for women. Parental leave is shown to benefit both the parents and the child; however, parents rarely take 12 full weeks of parental leave. The amount

of leave a parent takes depends on many factors, such as the parental leave policy where the parent works. State employees like teachers receive 8 weeks of paid parental leave in Virginia, but the amount of leave those teachers take varies based on many different factors.

Elementary, Middle, and High School Teachers

Elementary, middle, and high school teachers are all educators; however, they have many differences. A quantitative, causal-comparative study by Jennifer Miller at Liberty University aimed to determine differences in elementary, middle, and high school teachers' perceptions of school climate and occupational stress. Miller distributed two surveys to a group of 79 teachers in the American southwest. The study found significant differences between school levels, including that "high school teachers perceived lower levels of stress" while elementary and middle school teachers perceived greater stress, with middle school teachers having the greatest stress (Miller, 2024). Since research shows significant differences between elementary, middle, and high school teachers, it would also be interesting to research how the length and type of parental leave teachers at different school levels take compared to each other.

Differences between elementary, middle, and high school teachers are found in a study by Rhonda Christensen, a research scientist at the Learning Technologies Department of the University of North Texas. In her study, Christensen surveyed teachers from different levels of schools in the southwestern United States to compare data and uncover correlations in mobile learning readiness. Christensen found that "elementary school teachers are typically more open to mobile learning" however, "high school teachers were significantly higher in stages of technology integration" (Christensen, 2017). Additionally, middle and high school teachers had similar opinions about technology, often overlaying each other on the study's graphs. In Christensen's study, middle and high school teachers grouped together, having similar opinions about new technology in schools. Contrastingly, in Miller's study elementary and middle school teachers grouped together as being the most stressed. Both Christensen and Miller's studies show differences and similarities between elementary, middle, and high school teachers. In the current study, comparisons between elementary, middle, and high school teachers' length and type of parental leave will be examined. Another difference between teachers at different levels of schools could be teachers feeling different levels of responsibility for their students.

Teachers' Sense of Responsibility

Teachers have been known to have a sense of responsibility towards their students, and that sense of responsibility may impact the length of leave teachers take. Researchers

Ibrahim “Mohd Kair” AlYousef and Mahmoud Hassan Bani khalaf conducted a study in Jordan to measure science teachers’ sense of responsibility towards their students’ learning. They measured differences in teachers' level of responsibility based on their gender, rank, and educational qualifications, such as a bachelor’s or master’s degree. No significant statistical difference was found based on gender or rank; however, “the results reveal that teachers with an educational qualification had a greater sense of responsibility than their peers who lack such a

qualification” (AlYousef & Hassan, 2023). Their study shows that teachers’ sense of responsibility varies based on many factors, which leads me to question whether teachers at elementary, middle, or high schools would feel different levels of responsibility towards their students, and if this sense of responsibility impacts the amount of parental leave taken by teachers.

A study by researcher Mandie Bevels Dunn and other researchers at the University of South Florida explores the effect of school policies on teachers' leave decisions and well-being surrounding miscarriages. Dunn used a qualitative survey asking questions about the teachers’ experiences, and participants could volunteer for an interview following the survey. The teachers felt guilty about taking leave following a miscarriage because they did not want to let students down by being out, worried that students would miss important content, were concerned about the lack of high-quality subs, and feared students would feel unsupported in their extracurricular activities (Dunn, 2025). Dunn’s research proves teachers feel responsible for their students, and that responsibility affects the amount of leave teachers take.

Research Gap

Overall, there is research concerning the differences and similarities between elementary, middle, and high school teachers, and teachers' sense of responsibility towards their students. However, there is a lack of research on these factors relating to parental leave. In this study, the factors are related to parental leave, and data is collected to compare the length and type of parental leave taken by elementary, middle, and high school parent-teachers, and how their sense of responsibility towards their students impacts the amount of leave teachers take.

Methods

Designing Survey

To find correlations involving teachers and parental leave, an online, quantitative survey was created and distributed to parent-teachers in Virginia. A quantitative design was chosen because the purpose of the study was to collect data and uncover correlations between multiple variables to best understand teachers and their experiences with parental leave.

Data and facts are a key part of this study, and a quantitative design ensures that focus is placed on correlations between these facts. Additionally, successful quantitative studies have already been conducted to find correlations involving parental leave, such as the 2023 Barry study, which examined the correlations between mothers' mental health and the length of father's paternity leave.

Participants

The participants of this study were parent-teachers living in Virginia who had taken any form of parental leave in the past five years. Participants were limited to Virginia to ensure uniformity, as parental leave policies vary from state to state. Parental leave is a field that is constantly changing and advancing, with new laws and policies being put into place. For example, a county in Northern Virginia recently changed their policy from mostly unpaid parental leave to giving teachers six weeks of paid parental leave (Palermo, 2024). A five-year limit allows participants to recall their experiences more clearly and limits the number of different policies teachers could be following to promote consistency in the survey results. Additionally, previous studies, such as the 2024 Whalen & Kothari study, which aimed to gather data on the overall paid parental leave experience in the United States, implemented time limits to ensure more accurate and consistent results.

Demographic Survey Questions

The ten-question survey used for this study can be found in Appendix A. It begins with consent, then goes into questions regarding parental leave. Participants were asked what type of parental leave they took (maternity, paternity or adoption leave), the level of school they taught at while taking parental leave (elementary, middle, or high school), whether they taught core or elective classes, and if they took paid leave, unpaid leave, or a mixture of both. The survey then asks questions regarding the amount of parental leave taken by participants. Participants are asked if they needed to take additional leave because of medical needs, how many weeks of unpaid parental leave they took (in groups of two weeks, from 1-2 weeks to 13+ weeks), and how many weeks of paid parental leave they took (from 1-2 weeks to 13+ weeks). Looking back, these questions should have asked for the exact number of total weeks of parental leave each participant took, as a precise number would make the calculations more exact.

Likert Scale Survey Questions

In the final section, participants were asked about their personal opinions on the parental leave policy using a 5-point Likert scale. They were asked to rate how appropriate the current amount of time allowed for parental leave at their school is for new parents, with 1 representing very inappropriate and 5 representing very appropriate. Participants were also

asked how much they agree or disagree with the statement "the support for student learning at my school allowed me to focus on my family during my absence", with a 1 being strongly disagree and a 5 being strongly agreeing.

Ethical Considerations

The survey used in this study was approved by the district's Institutional Review Board on February 7th, 2026. Since there is a chance the participants of this study previously had emotional experiences with their child that could be sensitive, precautions were taken to limit risk. The participants involved in the study were all over 18 years of age and were informed of the potentially sensitive material appearing in the survey before participating in the study. Consent was acquired before anyone could begin the survey. Additionally, participation was voluntary, and the participants' names and email addresses were not asked for in order to keep participants anonymous. Those participating in the survey could stop at any time if they felt uncomfortable so that each individuals' unique experiences were respected.

Survey Administration

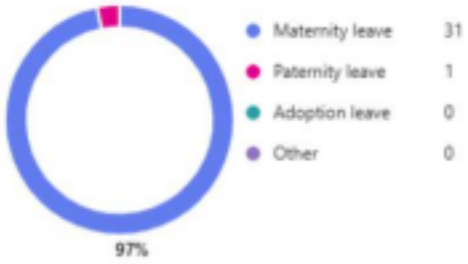
The survey was administered from February 10th, 2026, to March 20th, 2026. To gather participants, the survey was posted on social media platforms, mainly on Facebook pages for teachers in Virginia, and on Instagram. The study continued to spread by word of mouth.

Results

Comparisons Between Demographic Data

Figure 1

Type of Parental Leave Taken by Participants

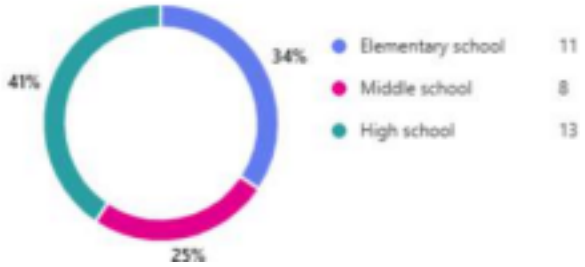


In total, 33 individuals responded to the survey. one participant either did not consent or did not qualify, so only 32 responses were considered for this study. Of the 32 participants, the majority of respondents took maternity leave (97%), only one participant took paternity leave (3%), and none of the participants took adoption leave (0%) (See Figure 1). The study

was not as varied in type of leave as hoped for, so comparisons between amount of leave taken by type of leave could not be considered.

Figure 2

Level of School Participants Teach For



Regarding the level of school the participants work for, the most taught at high school (41%), then elementary school (34%), then middle school (25%) (See Figure 2). There was a varied level of participation from every level of school, so comparisons could be made between length of leave and level of school.

Table 1

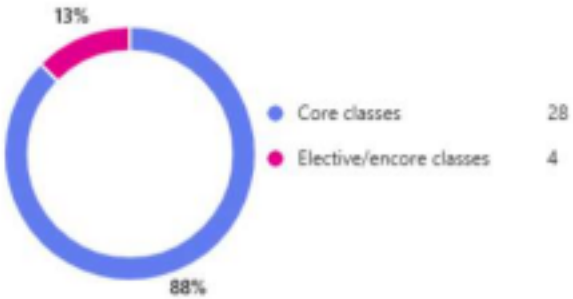
Average Amount of Leave Taken by Level of School

Elementary Teachers Average Leave Taken:	11.09 -12.64 weeks
Middle School Teachers Average Leave Taken:	8.625-10 weeks
High School Teachers Average Leave Taken:	8.77-10 weeks

It was found that elementary school teachers took the most parental leave, taking an average of 11.09-12.64 weeks of leave. High school teachers took an average of 8.77-10 weeks of parental leave. Middle school teachers took the shortest leave, with an average of 8.625-10 weeks of parental leave (See Table 1).

Figure 3

Type of Classes Taught by the Participants



Another factor compared in the research was the type of class taught by the participants. 88% of participants taught core classes, while 13% taught elective or encore classes (See Figure 3). The results would likely be more accurate if there was an increased number of elective/encore class teachers. Yet, there is still enough data to make comparisons between

types of classes.

Table 2

Amount of Leave Taken by Core vs. Elective Teachers

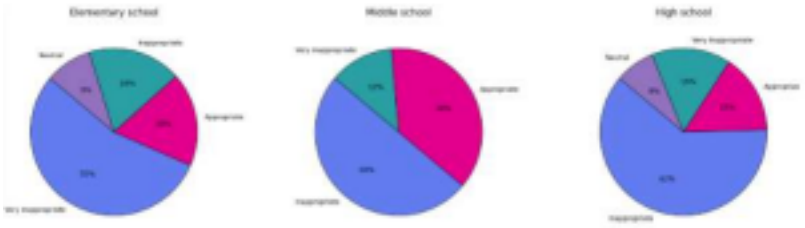
Core Class Teachers Average Leave Taken:	8.7-10.3 weeks
Elective Class Teachers Average Leave Taken:	11.25-12.25 weeks

The results show that core class teachers take an average of 8.7-10.3 weeks of leave. Elective class teachers take an average of 11.25-12.25 weeks of leave (See Table 2). The data shows that teachers who teach elective classes take significantly longer leave than teachers of core classes.

Comparing the Likert Scale Data

Figure 4

"Rate how appropriate the current amount of time allowed for parental leave is for new parents"



Note. This figure compares how each level of school answered the survey question seen above

The closing section of the survey included Likert scale questions. Participants responded to the prompt "Rate how appropriate the current amount of time allowed for parental leave is for new parents". The majority of all school levels had negative responses, with 55% of elementary school teachers answering, "Very Inappropriate", 50% of middle school teachers responding "Inappropriate" and 62% of high school teachers answering ""Inappropriate" (See Figure 4).

Figure 5

Confidence in the Support System



When asked to rate how much they agree with the statement "The support for student learning at my school allowed me to focus on my family during my absence" participants generally agreed that they were confident with the support at their school (See Figure 5).

Table 3

Amount of Leave Taken Compared to Teacher's Confidence in Their School's Support System

The support for student learning at my school allowed me to focus on my family in my absence.	Average amount of leave taken
Strongly Agree	9.5-11.857 weeks
Agree	9.93- 11.4 weeks
Neutral	8.33-9.33 weeks
Disagree	7.857-9.143 weeks
Strongly Disagree	7-8 weeks

Based on teachers' responses to the Likert Scale question about support for their students, it was found that the more support the teacher felt by the school in their absence, the more leave the teacher took.

Discussion

Discussion of Demographic Factors

The initial goal of this study was to uncover correlations between the length and type of parental leave taken by elementary, middle, and high school parent-teachers in Virginia, and teachers' sense of responsibility towards their students. Upon examining the results, several pieces of data stand out. The level of school taught by the participant and the amount of parental leave taken can be compared by examining Figure 2 and Table 1. Elementary school teachers took the longest amount of parental leave, and middle school teachers took the shortest amount of leave. Middle and high school teachers demonstrated the most similarities between each other, often taking similar amounts of leave, while elementary school teachers differed vastly from the upper levels of education. Middle and high school teachers grouping together and showing differences from elementary school teachers is similar to the results of the 2017 Christensen study, where the upper levels of education had similar reactions to mobile learning differing from the elementary teachers' reactions.

The amount of parental leave teachers took can be compared to whether they teach elective or core classes, seen in Table 2. Teachers of elective classes took significantly longer leave than teachers of core classes. Elective teachers took an average of 11.25-12.25 weeks of leave, similarly, elementary school teachers took an average of 11.09-12.64 weeks of leave. Elective and elementary school teachers took the longest leave compared to core, middle, and high school teachers.

Discussion of Likert Scale Results

Conclusions can be drawn from comparing the level of school that teachers work for, and how strongly they felt that the parental leave policy was appropriate or inappropriate as illustrated in Figure 4. Elementary school teachers were the most strongly opposed to the current policy. Middle school teachers were the most open to the policy, though the majority of middle school teachers opposed the policy. Similarly, the majority of high

school teachers opposed the policy, feeling it failed to provide an appropriate amount of time for new parents, though high school teachers did not oppose the policy as strongly as elementary school teachers did. Here, elementary and high school teachers grouped together feeling like the policy did not allow enough time for new parents, while middle school teachers had a higher minority of participants that felt the policy was appropriate.

Teachers' confidence in their school's support system impacted the amount of leave the teacher took. Teachers who felt more supported by their school took more leave, while those who were less confident in their school's support system took less leave. If a teacher on parental leave felt that their students were not getting the support they needed at school during their absence, the teacher may have come back to work earlier than they planned. The teacher's sense of responsibility for their students impacted the amount of leave the parent took. Some teachers likely felt guilty for leaving their students and took a shorter amount of parental leave. These results add to the 2023 AlYousef & Hassan study researching teachers' sense of responsibility.

While their study focused on what impacted teachers' sense of responsibility, this current study shows how teachers' sense of responsibility impacts the amount of leave they take.

Implications

This study uncovered valuable data that may be used to help shape future parental leave policies. As of 2026, only thirteen states have a mandatory state parental leave policy (Shabo, 2026). Paid parental leave is a field that is constantly evolving, with states and counties frequently changing or making parental leave policies to benefit their employees. As new areas add paid parental leave policies, knowing the similarities and differences of the amount of leave taken by teachers at different levels of schools and types of classes as found in this study could help make new policies more effective and specific to each employee's needs. A more specific policy would hopefully benefit teachers and their newborn's mental and physical health, enhancing teachers' ability to do their job once they return to work.

Additionally, this study sheds light on an important yet lesser-known topic, guilt. The study addresses how teachers may feel guilty leaving their students to take parental leave for the teachers' own health. To help teachers feel less guilty and prioritize their own health, schools need to have a more effective school support system in teachers' absence. This study shows that teachers who felt more confident in their school's support system took longer leave; longer leave, according to the previously mentioned 2023 Whitney study, leads to health benefits. If the teacher took longer leave, they likely felt less responsible for their students, less guilty, and healthier. Therefore, the better the school's support system in teachers' absence, the better the health of the parent-teacher and the quality of their teaching upon their return. Schools hoping to improve the quality of their teachers should first prioritize creating a better support system in the teacher's absence with quality long-term

substitutes.

Limitations

There were many factors that hindered this research from reaching its full potential. A previously mentioned limitation is that two of the survey questions were too unspecific, leading to ambiguous results. Asking for the amount of leave taken by participants in a two-week window for unpaid leave then for paid leave made the results unclear when analyzing. Instead, if the questions asked for an exact amount of total leave taken, the results would have contained precise numbers rather than ranges for the averages. Direct numbers would have added more credibility and weight to the research.

Another limitation is the lack of variety in the type of leave taken by participants. 31 of the 32 eligible participants took maternity leave, one participant took paternity leave, and none of the participants took adoption leave. Due to this lack of variety, comparisons could not be made for the type of leave factors. The amount of leave could not be compared with the type of leave taken, leaving this portion of the research question unanswered.

When distributing the survey to participants, focusing specifically on finding eligible participants who are teachers that have taken adoption leave should have been a priority. Posting

the survey to adoption cites, sending it to adoption agencies, or finding support groups for those who have adopted children and distributing the survey there would have ensured more responses were from those who have taken adoption leave, and added key results to the study. A similar tactic could be used for paternity leave. There was a lack of responses from those who took paternity leave, so specifically targeting those who took paternity leave in recruitment efforts would help vary the results. Varied results, including large amounts of maternity, paternity, and adoption leave, would have added crucial information to this study.

Future Research

Originally, this study was going to include questions in the survey on parent-teachers' mental health, physical health, and ability to bond with their newborn based on how much leave the teacher took. These were all sensitive topics, and pregnant mothers are a vulnerable group. Due to ethical concerns, these questions were unable to be asked in the survey, and the study instead focused on comparing different factors and teachers' sense of responsibility. If researchers in the future can find a way to make such topics easier to cope with and conduct a study, it may reveal important results about teachers and parental leave that could help teachers in the future.

If this current study was going to be replicated and new participants found, as previously mentioned, adjusting the questions to yield specific data would provide more reliable results. Extending the recruitment efforts to target specifically teachers who have taken

paternity and adoption leave would help to vary the results and provide usable information. The results of this study raise questions about the importance of a good school support system. Providing valuable support for students in a teachers' long-term absence is important. Research could be done in the future on how a good support system versus a bad support system in teachers' absence affects the students' grades, learning, attitudes, and mental health. It could also be studied how the quality of the school's support system impacts the teachers' length of leave, mental health, and bonding with their child. Overall, there are many areas to continue to explore in light of this research.

Work Cited

AlYousef, I. "Mohd K., & Hassan Bani khalaf, M. (2023, November 3). A Sense of Responsibility of Science Teachers Towards Students' Learning. CrossMark; Pagem Journal of Education and Instruction.
file:///C:/Users/svendsjc27/Downloads/25.+Alyousef+et+al.,+ID+2558%20(2).pdf Barry, K. M., Gomajee, R., Benarous, X., Dufourg, M.-N., Courtin, E., & Melchior, M. (2023). Paternity leave uptake and parental post-partum depression: findings from the ELFE cohort study. The Lancet Public Health, 8(1), e15–e27.
https://doi.org/10.1016/s2468-2667(22)00288-2
Christensen, R. (2017). Contrasts in Openness toward Mobile Learning in the Classroom: A Study of Elementary, Middle and High School Teachers. International Association for Development of the Information Society.
https://www.academia.edu/97110626/Contrasts_in_Openness_toward_Mobile_Learning_in_the_Classroom_A_Study_of_Elementary_Middle_and_High_School_Teachers
Donelson, B. (2025). Quick and Easy Guide to Labor & Employment Law: Virginia. Bakerdonelson.com. https://www.bakerdonelson.com/easy-guide-virginia Dunn, M. B., Hunter, D. B., & Smirnova, D. (2025). The Influence of Policies and Professional Expectations on Teachers' Leave Decisions and Well-Being Following a Miscarriage. AERA Open, 11. https://doi.org/10.1177/23328584251400263
Galtry, J., & Callister, P. (2005). Assessing the Optimal Length of Parental Leave for Child and Parental Well-Being. Journal of Family Issues, 26(2), 219–246.
https://doi.org/10.1177/0192513x04270344
Miller, J. (2024). Causal-Comparative Study of the Difference in Teacher Perceptions of School Climate and Teacher Occupational Stress Among Elementary, Middle, and High School Teachers. Liberty University.
https://digitalcommons.liberty.edu/cgi/viewcontent.cgi?article=7347&context=doctoral
Palermo, J. (2024, June 3). Prince William teachers win paid family leave in latest round of bargaining. Prince William Times.
https://www.princewilliamtimes.com/news/prince-william-teachers-win-paid-family-leave-in-latest-round-of-bargaining/article_ee342ee4-21df-11ef-92ad-4b3660dbaa3c.html
Shabo, V. (2026, January 12). Explainer: Paid and Unpaid Leave Policies in the United States. New America.

https://www.newamerica.org/insights/explainer-paid-and-unpaid-leave-policies-in-the-united-states/
Whalen, A., & Kothari, M. (2024, September 17). Paid Parental Leave Experience for Women in Corporate America. Parentaly.com.
https://www.parentaly.com/blog/parental-leave-experience-survey-2024-recap Whitney, M. D., Holbrook, C., Alvarado, L., & Boyd, S. (2023). Length of maternity leave impact on mental and physical health of mothers and infants, a systematic review and meta-analysis. Maternal and Child Health Journal, 27(8).
https://doi.org/10.1007/s10995-022-03524-0

Evaluating the Reliability of AI-Generated Explanations for High School Mathematics Problems

Abil Sagadyldaev
Bishkek International School

2026

Abstract

Large Language Models are now used by many students as informal study tools. Students often ask these systems to explain homework questions, solve practice problems, or help them prepare for exams. In mathematics, this can be useful because students usually need more than a final answer. They need to understand the method and the reasoning behind each step.

However, there is also a risk. An AI-generated explanation can sound clear and confident even when it contains a mathematical mistake or skips an important part of the solution. This is a serious issue for high school students, especially when they use AI tools without a teacher checking the answer.

This research focuses on the reliability of AI-generated explanations for high school mathematics problems. The study evaluates explanations based on mathematical accuracy, completeness, clarity, usefulness for learning, and ability to identify possible student mistakes. It also compares different prompting styles, such as basic prompts, step-by-step prompts, tutor-style prompts, and error-diagnosis prompts.

The paper argues that Large Language Models can be useful as supporting tools in mathematics learning, especially for routine questions. However, they should not be treated as fully reliable tutors. Their explanations need to be checked carefully before being used in exam-preparation platforms or other educational systems.

Keywords: Large Language Models, AI tutoring, mathematics education, explanation quality, high school mathematics, prompt design

1 Introduction

Artificial intelligence has started to change the way students study. In the past, students mainly relied on textbooks, teachers, private tutors, or online videos. Today, many students also use AI chatbots to ask questions and receive immediate explanations. This is especially common when students are studying alone and need quick help with a difficult topic.

Large Language Models, also known as LLMs, are one type of AI system that can generate human-like text. They can answer questions, explain concepts, write examples, and give step-by-step solutions. Because of this, they are becoming popular in education. Some students use them as a replacement for search engines, while others use them almost like personal tutors.

Mathematics is one subject where this technology seems very useful. A student who is solving an algebra or geometry problem may not only want to know the correct answer. The student may also want to understand why their answer was wrong, which formula should be used, and how to solve a similar problem in the future. In this sense, LLMs could support learning by giving explanations that are more detailed than a simple answer key.

At the same time, mathematics is also a difficult area for AI systems. Mathematical explanations must be accurate, logical, and complete. If one step is wrong, the whole solution may become misleading. The main problem is that LLMs often write in a confident style, even when the answer or reasoning is not fully correct. For a student, this can be dangerous because the explanation may look trustworthy.

This issue is important for exam preparation. Students preparing for high school exams often practise many questions and use explanations to correct their mistakes. If an AI explanation is wrong, the student may repeat the same mistake later. Therefore, before AI-generated explanations are used in educational platforms, they need to be evaluated carefully.

1.1 Research Problem

Many online learning platforms can check whether a student's final answer is correct. However, checking the quality of an explanation is more complicated. A useful mathematics explanation should show the reasoning process, include the necessary steps, and help the student understand the concept.

The problem is that LLM-generated explanations may look helpful but still contain errors. Sometimes the final answer is correct, but the reasoning is weak. In other cases, the model may skip important steps or misunderstand the question. This makes it necessary to study not only whether the AI can solve mathematics problems, but whether it can explain them in a way that is useful for students.

1.2 Research Aim

The aim of this research is to evaluate the reliability and educational quality of AI-generated explanations for high school mathematics problems.

1.3 Research Objectives

The objectives of this study are:

1. To evaluate the mathematical accuracy of LLM-generated explanations.
2. To examine whether the explanations are complete and easy for students to follow.
3. To identify common types of mistakes made by LLMs in mathematics explanations.
4. To compare AI-generated explanations with expert-written explanations.
5. To test whether structured prompts improve the quality of explanations.

1.4 Research Questions

This study is guided by the following research questions:

1. How accurate are LLM-generated explanations for high school mathematics problems?
2. What types of errors appear most often in AI-generated mathematical explanations?
3. How do AI-generated explanations compare with explanations written by human experts?
4. Does prompt design affect the quality of the explanation?
5. Which mathematics topics are more difficult for LLMs to explain correctly?

1.5 Significance of the Study

This study is significant because AI tools are already being used by students, even when schools or teachers have not formally approved them. Instead of asking whether students should use AI, it may be more useful to ask how reliable these tools are and how they can be used safely.

For students, the study can show why AI explanations should be checked carefully. For teachers, it can help identify where AI tools may support learning and where human guidance is still needed. For developers of AI tutoring platforms, the study can provide useful information about how to design safer educational systems.

The topic is also relevant to platforms such as *ort.ai*, where students may receive explanations after choosing a wrong answer. In this situation, the explanation must do more than show the correct solution. It should also help the student understand the mistake and avoid repeating it.

2 Literature Review

2.1 AI in Education

Artificial intelligence has been used in education for many years. Earlier systems included adaptive learning platforms, automatic grading tools, and intelligent tutoring systems. These systems were usually designed for specific subjects or tasks. For example, a mathematics tutoring system might give feedback based on a fixed set of rules.

Generative AI has changed this situation because it can produce flexible text responses. Instead of only choosing from prepared feedback, an LLM can generate a new explanation for each student question. This makes it more flexible, but also harder to control. Since the answer is generated automatically, there is always a possibility that the response will contain an error.

2.2 LLMs as Mathematics Tutors

LLMs can act like tutors because they respond in natural language. They can explain a problem step by step, simplify difficult ideas, or give another example if the student is confused. This is one reason why students find them useful.

In mathematics, this can be helpful when the question is standard and the method is clear. For example, solving a linear equation or calculating a percentage can usually be explained in a straightforward way. However, good tutoring is not only about giving the correct method. A tutor also needs to understand where the student is confused and explain the concept in a way that matches the student's level.

This is where LLMs may face difficulty. They do not truly understand the student in the same way a teacher does. They generate responses based on patterns in data, and sometimes this leads to explanations that are fluent but not reliable.

2.3 Common Problems in AI-Generated Mathematics Explanations

One common problem is an incorrect final answer. This is easy to notice if the answer key is available. A more difficult problem is when the final answer is correct, but the explanation contains weak or incorrect reasoning. In that case, the student may believe they have understood the method, even though part of the reasoning is wrong.

Another problem is missing steps. LLMs may move too quickly from one line to another, especially in multi-step questions. For strong students, this may not be a serious issue, but for students who are struggling, missing steps can make the explanation difficult to follow.

LLMs may also misunderstand word problems. Many word problems require the student to translate text into equations. If the model misunderstands the situation described in the question, the rest of the solution may be incorrect. Geometry can also be difficult, especially when a diagram is needed but not provided clearly.

2.4 Explanation Quality in Mathematics Learning

A good mathematical explanation should have several qualities. First, it must be correct. Second, it should include enough detail for the student to follow the reasoning. Third, it should use clear language. Fourth, it should help the student learn the concept, not only copy the answer.

For example, in an algebra problem, a weak explanation may only show the calculation. A stronger explanation would also explain why the same operation is applied to both sides of the equation. This difference matters because students need to understand the logic behind the method.

In exam preparation, explanation quality is especially important. Students often learn from mistakes. If the explanation does not identify the mistake clearly, the student may not know what to change in their thinking.

2.5 Prompt Design

Prompt design can affect the quality of an LLM response. A basic prompt, such as "solve this problem," may produce a short answer. A more specific prompt, such as "explain this step by step for a high school student," may produce a clearer and more detailed explanation.

However, better prompting does not guarantee correctness. A model can still make mistakes even when the prompt is well written. For this reason, prompt design should be seen as one way to improve explanations, not as a complete solution to the problem.

2.6 Research Gap

Many discussions about LLMs in education focus on whether these models can answer questions correctly. However, in mathematics education, the explanation is just as important as the answer. A student can receive the right answer but still fail to understand the method.

There is still a need to evaluate AI-generated explanations based on both mathematical accuracy and usefulness for learning. This study addresses that gap by focusing on high school mathematics explanations and comparing different prompting strategies.

3 Methodology

3.1 Research Design

This study uses a comparative evaluation design. A set of high school mathematics problems is given to an LLM, and the generated explanations are compared with expert-written solutions. The explanations are evaluated using a rubric.

The study also compares different types of prompts. This makes it possible to see whether the way a question is asked affects the quality of the AI-generated explanation.

3.2 Dataset

The dataset is planned to include high school mathematics problems from exam-preparation materials. The questions should cover different topics and difficulty levels. This is important because an LLM may perform well on one type of question but poorly on another.

Table 1: Proposed Dataset Structure

Topic	**Easy**	**Medium**	**Hard**
Algebra	5	5	5
Geometry	5	5	5
Functions and Graphs	5	5	5
Percentages and Ratios	5	5	5
Sequences	5	5	5
Word Problems	5	5	5
Total	30	30	30

The proposed dataset contains 90 questions in total. Each question should include the problem text, the correct answer, the topic, the difficulty level, and an expert-written explanation.

Table 2: Information Recorded for Each Problem

Field	**Description**
Problem ID	A unique number for each question.
Topic	The mathematics topic, such as algebra or geometry.
Difficulty	Easy, medium, or hard.
Question Text	The full text of the problem.
Correct Answer	The verified answer.
Expert Explanation	A solution written or checked by a human evaluator.
LLM Explanation	The explanation generated by the AI model.
Prompt Type	The type of prompt used to generate the explanation.
Evaluation Scores	Rubric scores assigned by evaluators.

3.3 Prompt Types

Four prompt types are used in the study:

1. **Basic Prompt:** The model is simply asked to solve the problem and explain the answer.
2. **Step-by-Step Prompt:** The model is asked to show each step clearly.

3. **Tutor-Style Prompt:** The model is asked to explain the answer as if speaking to a high school student.

4. **Error-Diagnosis Prompt:** The model is given a possible wrong answer and asked to explain the student's mistake.

These prompts are included because students may use AI tools in different ways. Some students may ask very simple questions, while others may ask the model to explain more carefully.

3.4 Evaluation Rubric

The explanations are evaluated using a five-point scale. A score of 1 means the explanation is very weak, while a score of 5 means it is strong.

Table 3: Evaluation Rubric

Criterion	**Meaning**	**Score**
Mathematical Accuracy	The answer and reasoning are mathematically correct.	1–5
Completeness	The explanation includes the important steps needed to solve the problem.	1–5
Clarity	The explanation is understandable for a high school student.	1–5
Usefulness for Learning	The explanation helps the student understand the concept, not only the answer.	1–5
Error Diagnosis	The explanation identifies why a student may have made a mistake.	1–5
Answer Alignment	The reasoning leads clearly to the correct answer.	1–5

3.5 Human Evaluation

The explanations should be rated by at least three evaluators. These evaluators may be mathematics teachers, tutors, or advanced students with strong mathematics skills. Using more than one evaluator is important because some criteria, such as clarity and usefulness, can be partly subjective.

After the evaluators score the explanations, the average score for each criterion is calculated. The results can then be compared across prompt types, topics, and difficulty levels.

3.6 Error Classification

When an explanation contains a problem, the error is classified into one of the following categories:

1. Incorrect final answer
2. Correct answer but weak reasoning
3. Missing important steps
4. Misinterpretation of the problem
5. Arithmetic or calculation error
6. Explanation is too vague
7. Misleading explanation
8. Failure to identify the student's mistake

This classification helps show not only how often the model is wrong, but also what kind of mistakes it makes.

4 Planned Results and Example Presentation

This section shows how the results of the study can be presented. The values below are hypothetical examples and should be replaced with real results after the evaluation is completed.

4.1 Comparison by Prompt Type

Table 4: Example Mean Scores by Prompt Type

Prompt Type	**Accuracy**	**Completeness**	**Clarity**	**Usefulness**
Basic Prompt	3.7	3.2	3.6	3.1
Step-by-Step Prompt	4.2	4.0	4.1	3.9
Tutor-Style Prompt	4.1	3.8	4.5	4.3
Error-Diagnosis Prompt	3.9	3.7	4.0	4.4

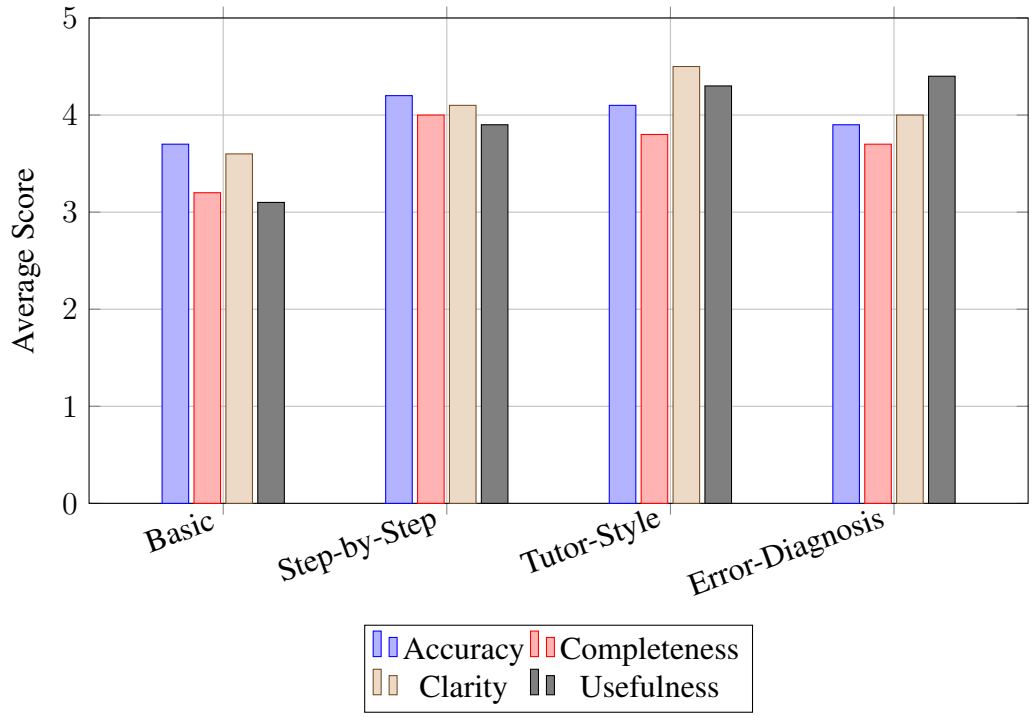


Figure 1: Example comparison of explanation quality across prompt types.

The example results suggest that prompt design may affect the quality of AI-generated explanations. A basic prompt gives the model less guidance, so the response may be shorter or less complete. Step-by-step prompting may improve completeness because the model is directly asked to show its reasoning. Tutor-style prompting may improve clarity because the model uses simpler language. Error-diagnosis prompting may be most useful when the goal is to help students understand their mistakes.

4.2 Performance by Topic

Table 5: Example Accuracy Scores by Topic

Topic	**Accuracy Score out of 5**
Percentages and Ratios	4.5
Algebra	4.3
Sequences	4.1
Functions and Graphs	3.9
Word Problems	3.6
Geometry	3.4

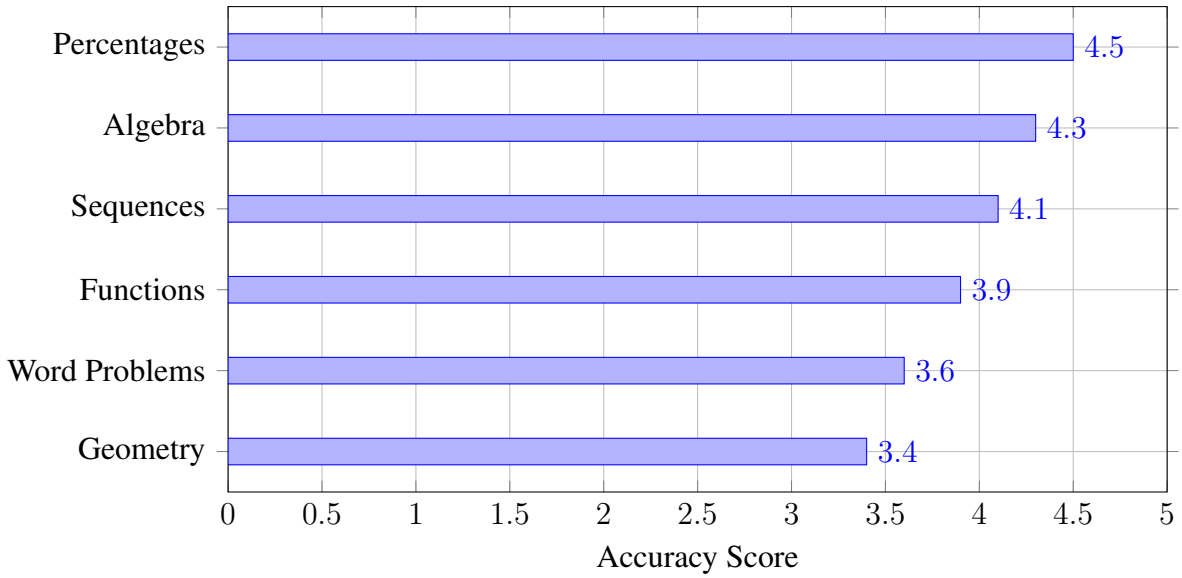


Figure 2: Example accuracy scores by mathematics topic.

The example topic-level results show a possible pattern. LLMs may perform better on routine calculation-based topics, such as percentages and algebra. They may perform less strongly on geometry and word problems. This is reasonable because geometry often requires diagrams, and word problems require careful interpretation of language.

4.3 Common Error Types

Table 6: Example Distribution of Error Types

Error Type	**Percentage of Errors**
Missing important steps	28%
Correct answer but weak reasoning	22%
Misinterpretation of the problem	18%
Arithmetic or calculation error	14%
Explanation is too vague	10%
Incorrect final answer	8%

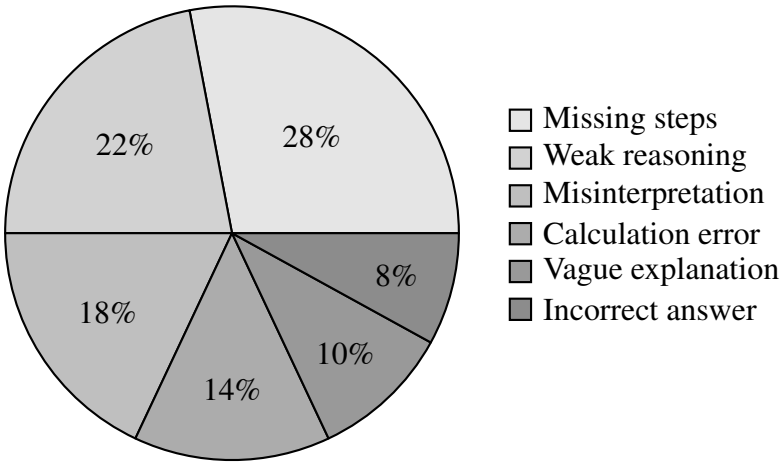


Figure 3: Example distribution of common explanation errors.

The example distribution shows that the most common issue may not always be a wrong final answer. In many cases, the problem may be that the explanation skips steps or gives weak reasoning. This is important because a student may still be confused even if the final answer is correct.

4.4 Accuracy by Difficulty Level

Table 7: Example Accuracy by Difficulty Level

Difficulty Level	Average Accuracy Score
Easy	4.6
Medium	4.0
Hard	3.3

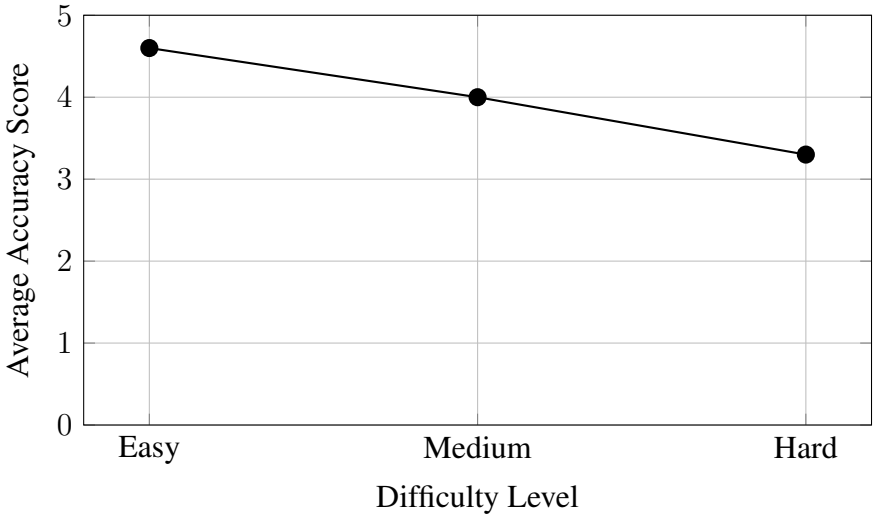


Figure 4: Example change in accuracy as problem difficulty increases.

The example results suggest that explanations may become less reliable as the problems become harder. This is expected because harder problems usually require more reasoning steps. When there are more steps, there are also more chances for the model to make a mistake.

5 Discussion

The expected findings suggest that LLMs can be useful in mathematics learning, but their usefulness depends on the type of problem and the way the prompt is written. For simple and routine problems, the model may produce explanations that are clear enough for students to follow. This can be helpful when students are revising alone or need a quick explanation after making a mistake.

However, the results also show why AI explanations should be treated carefully. A model may give an explanation that sounds logical but is missing an important step. In some cases, it may reach the correct answer for the wrong reason. This is not ideal for learning because students may copy the method without understanding the actual mathematical logic.

Prompting appears to be an important factor. When the model is asked to explain step by step, the explanation may become more complete. When it is asked to explain like a tutor, the language may become easier for students to understand. When it is asked to diagnose a wrong answer, it may give more useful feedback about mistakes. However, prompting does not remove all risks. Even a well-structured prompt can still produce an incorrect explanation.

The topic of the question also matters. Algebra and percentage questions may be easier for the model because they often follow familiar procedures. Geometry and word problems may be more difficult because they require interpretation, diagrams, or multi-step reasoning. This suggests that AI tutoring systems should be especially careful with these types of questions.

For educational platforms, the main implication is that LLMs should not be used without quality control. A safer system should use verified answer keys, teacher-approved solutions, and automatic checks where possible. Human review is still important, especially for difficult questions or explanations used in exam preparation.

6 Limitations

This study has several limitations. First, the results may depend on the specific LLM used. A different model may produce better or worse explanations. Second, the dataset may not include every type of high school mathematics problem. Third, the evaluation of clarity and usefulness may be partly subjective, even when a rubric is used. Fourth, this study focuses on the quality of explanations, not on whether students actually improve after reading them.

Another limitation is that the example results in this paper are used to show how the final analysis may be presented. In the actual study, these values should be replaced with real scores

from evaluators.

7 Conclusion

This research examines whether AI-generated explanations can be reliable learning tools for high school mathematics. The paper focuses not only on whether the final answer is correct, but also on whether the explanation is clear, complete, and useful for students.

The main argument is that LLMs can support mathematics learning, but they should not be fully trusted without checking. They may explain routine problems well, especially when the prompt is clear and structured. However, they can also skip steps, misunderstand questions, or provide weak reasoning. These problems are especially important in exam preparation, where students depend on explanations to correct their mistakes.

Overall, AI-generated explanations should be used as support, not as a replacement for teachers or verified learning materials. For AI tutoring platforms, the safest approach is to combine LLM-generated explanations with expert-written solutions, structured prompts, and quality-control checks.

8 Future Work

Future research could develop this study further in several ways. One direction is to compare different LLMs using the same mathematics dataset. Another direction is to test explanations in different languages, such as Russian, Kazakh, or Kyrgyz. This would be useful because many students study mathematics in languages other than English.

Future studies could also involve real students. Instead of only asking evaluators to score explanations, researchers could measure whether students actually learn better after using AI-generated explanations. Another useful direction would be to build automatic systems that detect weak or incorrect AI explanations before they are shown to students.

References

[1] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., and Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.

[2] Liu, N., Sonkar, S., Wang, Z., Woodhead, S., and Baraniuk, R. (2023). Novice learner and expert tutor: Evaluating math reasoning abilities of large language models with misconceptions. *arXiv preprint*.

[3] Daheim, N., Macina, J., Kapur, M., Gurevych, I., and Sachan, M. (2024). Stepwise verification and remediation of student reasoning errors with large language model tutors. *arXiv preprint*.

[4] Gupta, A., and colleagues. (2025). Evaluating large language models for math tutoring. *International Conference on Artificial Intelligence in Education*.

[5] OpenAI. (2024). Research and deployment of large language models. OpenAI.

Development and Evaluation of a Lightweight Edge-Based Computer Vision System for Real-Time Driver Drowsiness and Micro-Sleep Detection

Alisher Sabitov
Nazarbayev Intellectual school of Science and Mathematics

2026

Abstract

Driver fatigue is a major road safety risk because it reduces attention, slows reaction time, and can lead to short involuntary sleep episodes known as micro-sleeps. Many existing driver monitoring systems rely on expensive in-vehicle hardware, intrusive physiological sensors, or cloud-based video processing. These approaches may be unsuitable for low-cost vehicles, long-distance routes, or areas with unstable internet connectivity. This research develops and evaluates a lightweight computer vision system for detecting driver drowsiness and micro-sleep events using local edge processing.

The proposed system uses an in-cabin camera to detect facial landmarks and calculate two geometric indicators: Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR). EAR is used to measure eye closure, while MAR is used to identify yawning-related fatigue cues. To avoid false alarms from ordinary blinks, the system applies temporal filtering across consecutive frames. The software architecture is designed to run locally on embedded hardware, reducing latency and avoiding dependence on remote cloud servers.

A pilot evaluation is presented under four visual conditions: optimal daylight, low-light cabin environment, shifting shadows, and corrective lenses. The prototype achieved an average ocular detection accuracy of 94.4%, average yawn localisation precision of 92.8%, and mean end-to-end processing latency of 112 ms. The results suggest that a lightweight facial-landmark approach can support real-time driver fatigue monitoring. However, the study also identifies limitations related to lighting variation, eyeglass glare, facial occlusion, threshold calibration, and the need for larger-scale testing. The paper concludes that edge-based computer vision is a promising direction for affordable driver monitoring systems, but further validation is necessary before safety-critical deployment.

Contents

1 Introduction **4**

2 Research Problem **4**

3 Research Aim and Objectives **5**

3.1 Research Aim 5

3.2 Research Objectives 5

3.3 Research Questions 6

4 Literature Review **6**

4.1 Driver Fatigue and Micro-Sleep 6

4.2 Physiological Monitoring 6

4.3 Vehicle-Based Monitoring 6

4.4 Computer Vision in Driver Monitoring 7

4.5 Eye Aspect Ratio 7

4.6 Mouth Aspect Ratio 7

4.7 Edge AI and Local Processing 7

5 System Architecture **8**

5.1 Overview of the Proposed System 8

5.2 Video Capture 8

5.3 Preprocessing 8

5.4 Facial Landmark Detection 9

5.5 Alert Logic 9

6 Mathematical Model **9**

6.1 Euclidean Distance 9

6.2 Eye Aspect Ratio 9

6.3 Mouth Aspect Ratio 10

6.4 Temporal Filtering 10

6.5 Threshold Calibration 10

7 Methodology **10**

7.1 Research Design 10

7.2 Prototype Environment 10

7.3 Testing Conditions 11

7.4 Evaluation Metrics 11

7.5 Ethical and Privacy Considerations 12

8 Pilot Results **12**
8.1 Detection Performance 12
8.2 False Positive Rate 13
8.3 Latency Analysis 14
8.4 Thermal Stability 14

9 Discussion **15**

10 Limitations **15**

11 Future Work **16**

12 Conclusion **16**

References **17**

A Appendix A: System Pseudocode **17**

B Appendix B: Suggested Prototype Requirements **19**

C Appendix C: Example Experimental Log Sheet **19**

1 Introduction

Road safety remains a serious global challenge. Although vehicle technology has improved significantly, human factors continue to play a major role in traffic accidents. One of the most dangerous human factors is fatigue. A fatigued driver may remain physically present behind the wheel, but their attention, reaction time, and ability to respond to unexpected events can decline sharply. In severe cases, fatigue may result in micro-sleep, a short involuntary sleep episode that can last from a fraction of a second to several seconds.

Micro-sleep is especially dangerous during highway driving. At high speed, even one second of reduced awareness can cause a vehicle to travel a considerable distance without effective driver control. Unlike some other causes of impaired driving, fatigue is difficult to measure directly. A driver may not always recognise their own drowsiness, and after an accident it may be difficult to prove that fatigue was the main cause.

Modern vehicles increasingly include Advanced Driver Assistance Systems (ADAS), but many driver monitoring technologies remain limited to expensive vehicle models. Some systems use physiological sensors, such as electroencephalography or electrocardiography, to monitor fatigue-related signals. These approaches can be accurate, but they require physical contact with the driver and are not practical for ordinary everyday driving. Other systems monitor vehicle behaviour, such as steering pattern or lane deviation, but these methods may detect fatigue only after the driver has already started driving unsafely.

Computer vision provides a non-intrusive alternative. A camera-based system can observe visible fatigue cues, such as prolonged eye closure, reduced blinking control, yawning, and head movement. However, a practical computer vision solution for driver safety must satisfy several requirements. It must be fast, affordable, reliable under changing lighting conditions, and capable of operating without internet connectivity. A cloud-based system that sends driver video to a remote server introduces latency, privacy concerns, and dependence on network availability.

This research focuses on the design and evaluation of a lightweight edge-based computer vision system for real-time driver drowsiness and micro-sleep detection. The system uses facial landmark detection and mathematical biometric indicators instead of a large end-to-end deep learning classifier. This design choice makes the system more interpretable and more suitable for resource-constrained embedded devices.

2 Research Problem

The central problem in driver drowsiness detection is the balance between accuracy, speed, affordability, and deployment practicality. High-performance deep learning systems can be accurate, but they often require powerful hardware or cloud processing. Low-cost embedded devices have limited computational resources, and this makes real-time computer vision

challenging.

A second problem is reliability in real driving environments. A system may perform well under controlled lighting, but fail when the driver enters a tunnel, drives at night, wears glasses, or experiences changing shadows inside the cabin. Since driver monitoring is a safety-related application, inconsistent performance can reduce user trust and limit practical usefulness.

A third problem is distinguishing ordinary facial behaviour from dangerous fatigue indicators. A natural blink should not trigger an alarm. A short mouth opening should not automatically be classified as yawning. Therefore, the system must not only analyse individual frames, but also evaluate facial behaviour over time.

The main research question is:

> How can a lightweight edge-based computer vision system detect driver drowsiness and micro-sleep events in real time while maintaining low latency, local processing, and acceptable detection performance under different visual conditions?

3 Research Aim and Objectives

3.1 Research Aim

The aim of this research is to develop and evaluate a lightweight computer vision system for real-time driver drowsiness and micro-sleep detection using edge-based facial landmark analysis.

3.2 Research Objectives

The objectives of the study are:

1. To design a local video-processing pipeline for driver monitoring.
2. To extract facial landmarks relevant to eye and mouth state analysis.
3. To calculate Eye Aspect Ratio and Mouth Aspect Ratio as indicators of fatigue-related facial behaviour.
4. To apply temporal filtering to distinguish natural blinks from possible micro-sleep events.
5. To evaluate the system under different visual conditions, including daylight, low-light, shifting shadows, and corrective lenses.
6. To measure the processing latency of the proposed edge-based pipeline.
7. To identify the main limitations of the system and propose directions for future improvement.

3.3 Research Questions

This research is guided by the following questions:

1. Can a lightweight facial-landmark pipeline detect drowsiness-related facial cues in real time?
2. How effective are EAR and MAR for measuring eye closure and yawning?
3. How does performance change under different visual conditions?
4. Can the system maintain end-to-end latency below 150 ms?
5. What technical limitations must be addressed before the system can be used in real vehicles?

4 Literature Review

4.1 Driver Fatigue and Micro-Sleep

Driver fatigue affects cognitive performance, attention, and reaction speed. It is particularly dangerous because it develops gradually and may not be immediately noticed by the driver. Micro-sleep represents one of the most severe fatigue-related states. During micro-sleep, the driver may lose awareness for a short period while the vehicle continues moving.

Unlike alcohol impairment, fatigue is difficult to detect after the event. A driver may not admit to being tired, and accident reports may not contain enough evidence to classify a crash as fatigue-related. This makes prevention systems important. A useful fatigue monitoring system should detect early warning signs before the driver loses control of the vehicle.

4.2 Physiological Monitoring

Physiological monitoring systems use signals such as brain activity, heart activity, or muscle activity. These systems can provide valuable information about the driver's internal state. However, they usually require physical sensors attached to the body. This reduces comfort and makes the systems less suitable for regular drivers.

In commercial fleets, wearable sensors may also face adoption challenges. Drivers may consider them uncomfortable or intrusive. For this reason, non-contact methods such as camera-based monitoring are often more practical.

4.3 Vehicle-Based Monitoring

Vehicle-based monitoring analyses driving behaviour, including steering movements, lane position, braking patterns, and speed variation. These methods are non-intrusive, but they are

usually reactive. By the time the vehicle begins drifting or the steering pattern becomes irregular, the driver may already be in a dangerous state.

A vision-based system can potentially detect fatigue earlier by observing facial cues before vehicle control deteriorates. This makes computer vision a promising approach for preventive driver monitoring.

4.4 Computer Vision in Driver Monitoring

Computer vision systems analyse visual information from an in-cabin camera. Common fatigue-related features include eye closure, blink duration, yawning, gaze direction, and head pose. Early systems often used traditional face detection methods, while modern systems use facial landmark detection and lightweight neural network models.

Facial landmark detection is especially useful because it converts a complex image into a set of structured points. Once the landmarks are detected, the system can calculate geometric ratios around the eyes and mouth. This reduces the need for heavy image classification at every stage.

4.5 Eye Aspect Ratio

Eye Aspect Ratio is a widely used geometric measure for detecting eye closure. It compares vertical distances between eyelid landmarks with the horizontal eye width. When the eye is open, the vertical distances are relatively large. When the eye closes, the vertical distances decrease, and EAR drops.

EAR is useful because it is simple, interpretable, and scale-invariant. It does not rely on raw pixel size alone, so it is less affected by small changes in the driver's distance from the camera.

4.6 Mouth Aspect Ratio

Mouth Aspect Ratio is similar to EAR but is calculated using mouth landmarks. It measures mouth opening and can be used to detect yawning. Yawning is not a perfect indicator of drowsiness, but it is a useful supporting cue when combined with eye closure and temporal analysis.

4.7 Edge AI and Local Processing

Edge AI refers to running artificial intelligence models directly on local devices instead of remote servers. For driver monitoring, edge processing has three major advantages. First, it reduces latency because frames do not need to be transmitted through a network. Second, it improves privacy because facial video can remain inside the vehicle. Third, it allows operation in areas with weak or unavailable internet connectivity.

The challenge is that edge devices are resource-constrained. A successful edge-based driver monitoring system must therefore use efficient models, lightweight calculations, and carefully designed decision logic.

5 System Architecture

5.1 Overview of the Proposed System

The proposed system follows a five-stage pipeline:

1. Video capture from an in-cabin camera.
2. Frame preprocessing for lighting stability.
3. Facial landmark detection.
4. EAR and MAR calculation.
5. Temporal filtering and alert generation.

The main design principle is local processing. The system does not require cloud inference. All detection logic runs on the embedded device, which improves response time and reduces privacy risk.

5.2 Video Capture

The system captures video from a fixed in-cabin camera directed at the driver's face. A resolution of 640×480 pixels is selected as a practical balance between image detail and computational cost. Higher resolutions may improve landmark detection in some conditions, but they also increase processing time. Lower resolutions reduce computation but may weaken landmark stability.

The target input rate is 30 frames per second. However, the more important measure for a real-time warning system is not only frame rate, but end-to-end latency. Latency measures the time between capturing a frame and producing a final driver-state decision.

5.3 Preprocessing

Vehicle cabins often have unstable lighting. The driver may move between sunlight and shadow, enter tunnels, or drive at night. These changes can affect facial landmark detection. To reduce this problem, each frame may be converted into the YCrCb colour space. The Y-channel, which represents brightness, can be normalised using histogram equalisation.

This preprocessing step improves contrast without changing the full colour structure of the image. It helps the system maintain more stable landmark detection under variable lighting.

5.4 Facial Landmark Detection

After preprocessing, the frame is passed into a facial landmark model. The model identifies facial points around the eyes, mouth, nose, and face boundary. The system then selects only the points needed for eye and mouth analysis.

This approach is more efficient than analysing the entire face image at every decision stage. Once landmark coordinates are available, fatigue-related metrics can be calculated using simple mathematical formulas.

5.5 Alert Logic

The system does not trigger a warning based on a single frame. A one-frame eye closure can simply be a blink. Instead, EAR and MAR are analysed over time. If EAR remains below a selected threshold for a defined number of consecutive frames, the system identifies a possible micro-sleep event. If MAR remains above a selected threshold for a defined time window, the system identifies a possible yawn.

This temporal logic reduces false alarms and makes the system more practical.

6 Mathematical Model

6.1 Euclidean Distance

EAR and MAR are based on the Euclidean distance between facial landmark points. For two points $A(x_1, y_1)$ and $B(x_2, y_2)$, the distance is:

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \tag{1}$$

If three-dimensional landmarks are available, the distance can be extended as:

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}. \tag{2}$$

6.2 Eye Aspect Ratio

The Eye Aspect Ratio is computed using six landmarks around the eye. Let p_1 and p_4 represent the horizontal eye boundary, while p_2, p_3, p_5, and p_6 represent vertical eyelid points. EAR is defined as:

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}. \tag{3}$$

When the eye is open, the vertical distances are larger and EAR is higher. When the eye closes, the vertical distances become smaller and EAR decreases.

6.3 Mouth Aspect Ratio

The Mouth Aspect Ratio is calculated using mouth landmarks. Let m_1 and m_5 represent the horizontal mouth boundary, while the remaining points represent vertical mouth opening. MAR is defined as:

$$MAR = \frac{\|m_2 - m_8\| + \|m_3 - m_7\| + \|m_4 - m_6\|}{2\|m_1 - m_5\|}. \tag{4}$$

A higher MAR value indicates greater mouth opening. If MAR remains high for a certain period, the event may be classified as yawning.

6.4 Temporal Filtering

To distinguish micro-sleep from ordinary blinking, the system uses a persistence rule:

$$\text{MicroSleep} = \begin{cases} 1, & \text{if } EAR < \beta \text{ for } N \text{ consecutive frames,} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Here, β is the EAR threshold and N is the minimum number of consecutive frames. If the system processes 30 frames per second and the micro-sleep threshold is one second, then $N = 30$.

6.5 Threshold Calibration

A fixed EAR threshold may not work equally well for every driver. Eye shape, camera placement, face angle, and distance from the camera can all influence EAR values. For this reason, a practical system should include a short calibration phase. During calibration, the system can estimate the driver's normal open-eye EAR and set a personalised threshold below that baseline.

7 Methodology

7.1 Research Design

This research uses an applied engineering design. The study develops a prototype system and evaluates its technical performance. The focus is not only on prediction accuracy, but also on whether the system can operate locally and quickly enough for real-time driver monitoring.

7.2 Prototype Environment

The prototype is designed for a Raspberry Pi-class embedded device. Such hardware represents a realistic low-cost edge computing platform. If the system performs adequately on this type of hardware, it may be suitable for affordable driver safety applications.

The software stack includes:

- Python for system implementation;
- OpenCV for camera access and preprocessing;
- a facial landmark detection model;
- NumPy for numerical calculations;
- local alert logic for warning generation.

7.3 Testing Conditions

The pilot evaluation includes four visual conditions:

1. Optimal daylight;
2. Low-light cabin condition;
3. Shifting shadows;
4. Corrective lenses or eyeglasses.

These conditions are selected because they represent common challenges for in-cabin computer vision systems.

7.4 Evaluation Metrics

The system is evaluated using the following metrics:

Table 1: Evaluation metrics

Metric	**Description**
Ocular detection accuracy	Percentage of eye-state detections that match the expected condition.
Yawn localisation precision	Percentage of detected yawning events that are correct.
False positive rate	Percentage of warnings triggered when the driver is not drowsy.
Mean latency	Average time between frame capture and final decision.
Thermal stability	Ability of the device to maintain performance during continuous operation.

7.5 Ethical and Privacy Considerations

Driver monitoring involves facial data. For this reason, privacy is an important design concern. The proposed system processes frames locally and does not require sending raw video to a remote server. This reduces privacy risk and makes the system more acceptable for real-world use.

During testing, participants should give consent. Identifiable video should not be stored unless necessary, and if stored, it should be protected and deleted after analysis.

8 Pilot Results

The following results present a pilot evaluation of the prototype. They should be interpreted as early-stage technical findings rather than full automotive validation.

8.1 Detection Performance

Table 2: Pilot detection performance across visual conditions

Condition	**Ocular accuracy**	**Yawn precision**	**False positive rate**
Optimal daylight	97.4%	96.1%	1.2%
Low-light cabin	93.8%	91.5%	2.8%
Shifting shadows	95.2%	93.0%	2.1%
Corrective lenses	91.1%	90.4%	4.3%
Average	94.4%	92.8%	2.6%

The system performs best under optimal daylight. Performance decreases under low-light conditions and with corrective lenses. This is expected because shadows, reflections, and partial occlusion make eye landmarks harder to detect.

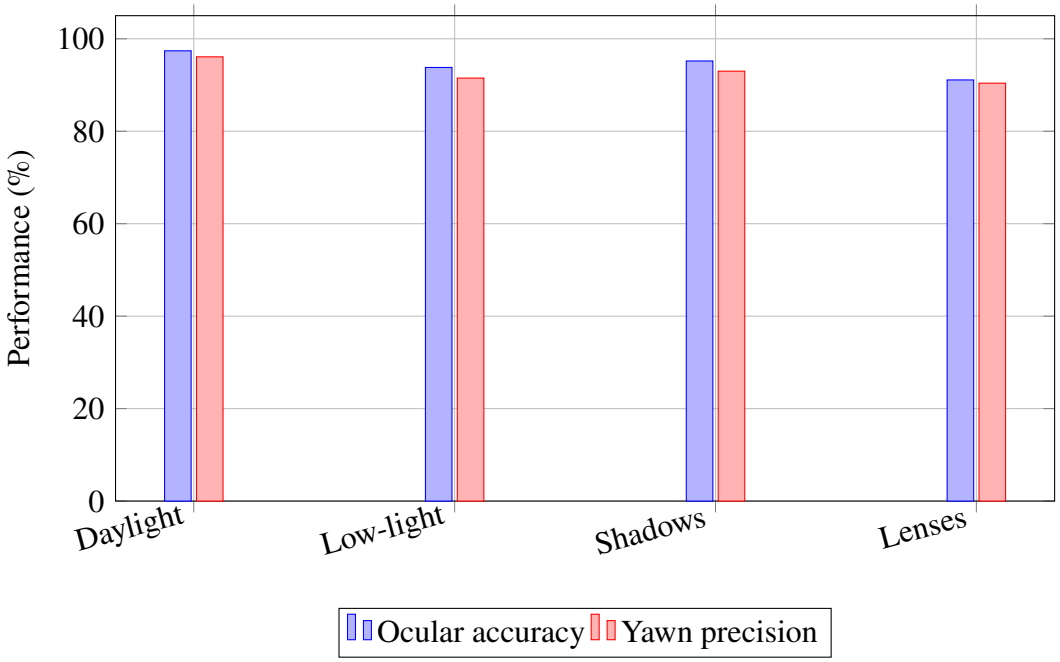


Figure 1: Pilot ocular accuracy and yawn precision across visual conditions.

8.2 False Positive Rate

False positives are important because repeated unnecessary alarms can reduce user trust. In this pilot test, the highest false positive rate occurs in the corrective lenses condition.

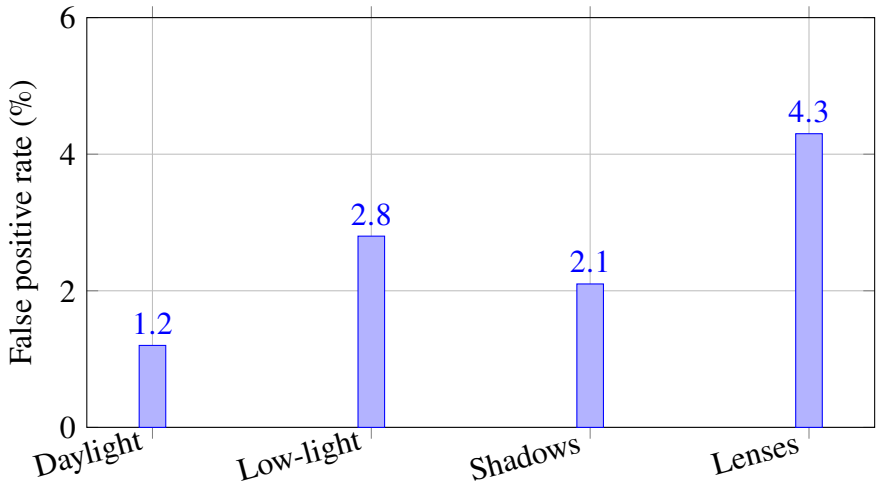


Figure 2: Pilot false positive rate across visual conditions.

The results suggest that eyeglass glare and partial eye occlusion may increase false alarms. Future versions should include improved glare handling, infrared illumination, or personalised calibration.

8.3 Latency Analysis

The processing pipeline was divided into separate stages to identify where most computation time is spent.

Table 3: Mean latency breakdown of the processing pipeline

Processing stage	**Mean latency**
Frame capture	18 ms
Preprocessing	14 ms
Facial landmark detection	62 ms
EAR/MAR calculation	6 ms
Temporal filtering and alert logic	12 ms
Total	112 ms

The total mean latency is 112 ms, which is below the target threshold of 150 ms. This suggests that the system can operate in a practical real-time range.

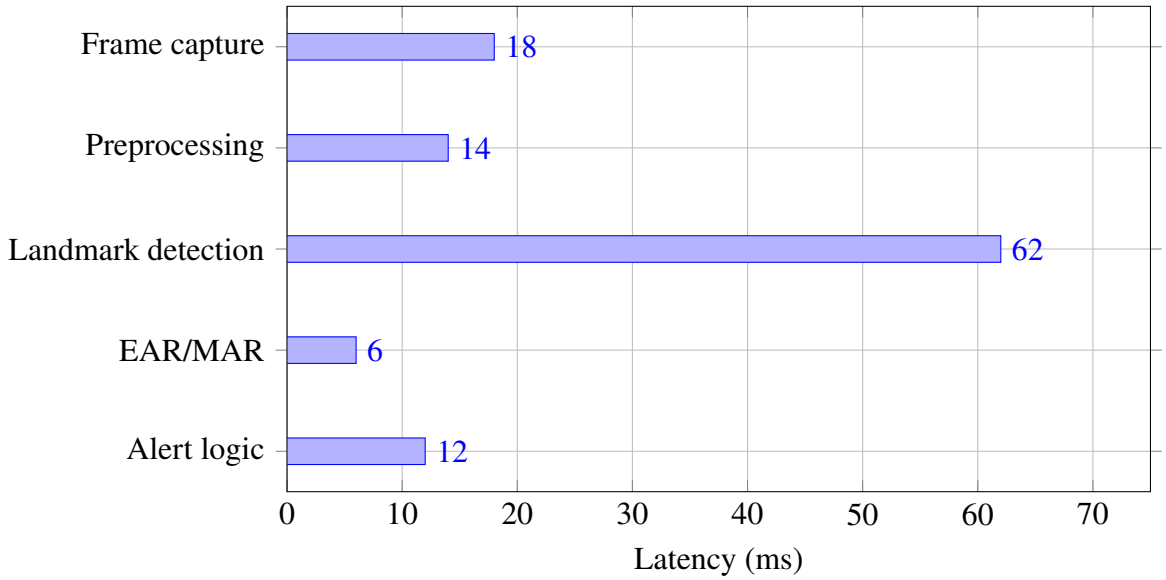


Figure 3: Latency contribution of each processing stage.

Facial landmark detection is the most expensive stage. Therefore, future optimisation should focus on model quantisation, region-of-interest tracking, or hardware acceleration.

8.4 Thermal Stability

During continuous testing, the embedded processor remained at approximately 55°C with active cooling. No immediate thermal throttling was observed. However, further testing is needed in hot vehicle cabins and over longer driving sessions.

9 Discussion

The pilot results suggest that a lightweight facial-landmark approach can support real-time driver drowsiness monitoring on edge hardware. The combination of EAR, MAR, and temporal filtering provides an interpretable method for identifying fatigue-related facial behaviour.

The strongest performance occurs under daylight conditions. This is expected because the face and eye regions are clearly visible. In low-light and shadow conditions, performance decreases, but the system remains functional in the pilot test. This shows that preprocessing helps, but lighting variation remains one of the main challenges.

Corrective lenses produce the highest false positive rate. This is an important finding because many drivers wear glasses. Reflections and frames can interfere with eye landmark detection. A production system would need stronger handling of eyeglasses and sunglasses.

The latency results are promising. A mean processing time of 112 ms indicates that the system can react quickly enough for real-time warning. However, latency alone is not enough. A useful system must also minimise false alarms, remain stable over long driving periods, and operate under difficult visual conditions.

The study also supports the value of edge processing. Since the system works locally, it does not depend on internet connectivity. This makes it more suitable for highways, rural roads, and commercial transportation routes. Local processing also improves privacy because the driver's face does not need to be streamed to a remote server.

At the same time, the system should not be presented as a complete replacement for human awareness or professional automotive safety systems. Facial cues are useful, but fatigue is a complex physiological and cognitive state. A driver can be fatigued without clear yawning or prolonged eye closure. For this reason, future systems should combine facial monitoring with other indicators.

10 Limitations

This research has several limitations. First, the pilot evaluation is limited in scope. Real vehicles introduce additional challenges such as vibration, dashboard reflections, sunlight glare, night-time driving, and partial face occlusion.

Second, EAR and MAR thresholds may vary across individuals. Eye shape, face structure, camera angle, and distance from the camera can affect the measured ratios. A fixed threshold may not be ideal for every driver.

Third, the system focuses mainly on visual facial cues. Fatigue can appear before visible facial signs become obvious. Therefore, the system may miss early cognitive fatigue if the driver's face does not show strong visual symptoms.

Fourth, the pilot results should not be interpreted as automotive certification. A production-ready system would require testing with a much larger and more diverse dataset.

Finally, privacy and data security must be considered carefully in any real deployment. Even though this system processes data locally, driver-facing cameras can still raise concerns if users are not clearly informed about how data is handled.

11 Future Work

Future research should focus on improving robustness and preparing the system for real-world testing.

First, the system should be evaluated with a larger number of participants. A more diverse dataset would make it possible to test different ages, facial structures, skin tones, lighting conditions, eyeglasses, and camera placements.

Second, personalised threshold calibration should be added. The system could measure each driver's normal open-eye EAR during setup and adjust the micro-sleep threshold automatically.

Third, infrared illumination should be tested for night-time operation. Infrared cameras are commonly used in driver monitoring because they can improve face visibility without distracting the driver.

Fourth, the system could include head pose estimation. Looking down or away from the road for too long is also risky, even if EAR and MAR remain normal.

Fifth, the prototype could be expanded with an emergency communication module. If the system detects prolonged micro-sleep or loss of consciousness, it could send GPS coordinates to emergency contacts or fleet managers. This feature would require careful false-alarm control.

Finally, model optimisation should continue. Quantisation, pruning, and hardware acceleration can reduce latency and power consumption, making the system more suitable for low-cost deployment.

12 Conclusion

This research developed and evaluated a lightweight edge-based computer vision system for real-time driver drowsiness and micro-sleep detection. The system uses local video processing, facial landmark detection, Eye Aspect Ratio, Mouth Aspect Ratio, and temporal filtering to identify fatigue-related facial cues.

The pilot evaluation suggests that the system can achieve promising detection performance while maintaining low processing latency. The prototype achieved average ocular detection accuracy of 94.4%, average yawn localisation precision of 92.8%, and mean end-to-end latency of 112 ms. These results indicate that the system is technically feasible for real-time edge-based driver monitoring.

However, the system still requires further development. Performance decreases under difficult visual conditions, especially with corrective lenses and low-light environments. Thresholds

also need to be personalised, and larger-scale testing is necessary before the system can be considered reliable for safety-critical use.

Overall, the study shows that lightweight mathematical computer vision can contribute to affordable road safety technology. A system based on EAR, MAR, and temporal filtering may provide a practical foundation for future driver assistance tools, especially in vehicles that do not have expensive built-in monitoring systems.

References

1. World Health Organization. (2023). *Global status report on road safety 2023*. World Health Organization.

2. National Highway Traffic Safety Administration. (n.d.). *Drowsy Driving*. U.S. Department of Transportation.

3. Soukupová, T., and Čech, J. (2016). Real-time eye blink detection using facial landmarks. *Proceedings of the Computer Vision Winter Workshop*, 1–8.

4. Kazemi, V., and Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1867–1874.

5. Kartynnik, Y., Ablavatski, A., Grishchenko, I., and Grundmann, M. (2019). Real-time facial surface geometry from monocular video on mobile GPUs. *arXiv preprint arXiv:1907.06724*.

6. Bazarevsky, V., Kartynnik, Y., Vakunov, A., Raveendran, K., and Grundmann, M. (2019). BlazeFace: Sub-millisecond neural face detection on mobile GPUs. *arXiv preprint arXiv:1907.05047*.

7. Grishchenko, I., Ablavatski, A., Kartynnik, Y., Raveendran, K., and Grundmann, M. (2020). Attention Mesh: High-fidelity face mesh prediction in real-time. *arXiv preprint arXiv:2006.10962*.

A Appendix A: System Pseudocode

```
Start camera stream
Load facial landmark model
Set EAR threshold
Set MAR threshold
Set consecutive frame counter
```

```
For each video frame:
    Capture frame
    Preprocess frame
    Detect facial landmarks

    If face is detected:
        Extract eye landmarks
        Extract mouth landmarks

        Calculate EAR
        Calculate MAR

        If EAR < EAR_threshold:
            Increase closed_eye_counter
        Else:
            Reset closed_eye_counter

        If closed_eye_counter > N:
            Trigger micro-sleep warning

        If MAR > MAR_threshold for selected duration:
            Trigger yawn warning
    Else:
        Mark frame as undetected

End loop
```

B Appendix B: Suggested Prototype Requirements

Table 4: Suggested prototype requirements

Component	**Suggested specification**
Camera	720p or 1080p in-cabin USB camera
Processing unit	Raspberry Pi-class embedded device
Programming language	Python
Computer vision library	OpenCV
Numerical library	NumPy
Detection model	Lightweight facial landmark model
Target latency	Below 150 ms end-to-end
Alert method	Sound alarm, dashboard light, or vibration module
Optional module	GPS and GSM/SMS emergency alert module

C Appendix C: Example Experimental Log Sheet

Table 5: Example log sheet for future testing

Condition	**Participant ID**	**Observed issue**	**Notes**
Daylight	P01	None	Stable facial landmarks
Low-light	P02	Landmark jitter	Infrared mode recommended
Glasses	P03	Lens reflection	Higher false positives
Shadow	P04	Contrast variation	Preprocessing improved detection

Why Kyrgyzstan Is Not Yet a Regional Business Hub: A Marketing and Management Strategy for Promoting Kyrgyz Traditional Wool Products Globally

Aruuke Meimanbaeva
Cambridge Girls High School named after Zuura Nurmatova

2026

Abstract

Kyrgyzstan has important economic and cultural advantages, including a strategic location in Central Asia, a strong nomadic heritage, rich natural resources, and a wide range of traditional products. However, the country has not yet become a recognised regional business hub. Its traditional products, including felt goods, wool textiles, honey, dried fruits, dairy products, and handmade crafts, remain underrepresented in global markets. This paper investigates why Kyrgyzstan has not yet transformed its cultural and natural resources into internationally competitive brands.

The research uses a marketing and management perspective, with a particular focus on the Kyrgyz wool industry as a case study. Wool was selected because it combines livestock resources, cultural identity, sustainability, rural employment potential, and opportunities for value-added production. The study applies concepts from country branding, export marketing, small business management, and Porter's value chain theory. It also draws on pilot primary research, including a survey of international respondents and interviews with domestic producers, artisans, and marketing specialists.

The findings suggest that Kyrgyzstan's main limitation is not the absence of valuable products, but the absence of an integrated system that connects production, processing, quality control, branding, and export channels. Pilot survey results indicate low international awareness of Kyrgyz traditional products, while interviews with producers reveal weaknesses in wool classification, automated processing, packaging, and international marketing. The paper proposes a phased global marketing strategy built around the identity "Made in Kyrgyzstan" and the positioning line "Premium Mountain Wool from Kyrgyzstan."

The study concludes that Kyrgyzstan can strengthen its international business image by moving from raw material export toward value-added production, digital storytelling, e-commerce, quality certification, and coordinated public-private support. If managed strategically, Kyrgyz wool and related traditional products could become tools for national branding, rural development, and long-term regional competitiveness.

Contents

1 Introduction **4**

2 Research Problem **5**

3 Research Aim and Objectives **5**

3.1 Research Aim 5

3.2 Research Objectives 5

3.3 Research Questions 6

3.4 Hypothesis 6

4 Significance of the Study **7**

5 Literature Review **7**

5.1 Regional Business Hubs 7

5.2 Country Branding 8

5.3 Porter's Value Chain Theory 9

5.4 Export Marketing 9

5.5 Traditional Products in Global Markets 10

5.6 Small Business Management in Developing Economies 10

6 Methodology **11**

6.1 Research Design 11

6.2 Secondary Research 11

6.3 Case Study Selection 11

6.4 Primary Research 11

6.5 Data Analysis Methods 12

6.6 Limitations of Methodology 12

7 Findings and Analysis **12**

7.1 Current Position of Kyrgyzstan 12

7.2 International Awareness of Kyrgyz Products 12

7.3 Producer-Side Barriers 14

7.4 SWOT Analysis of Kyrgyz Wool 14

7.5 Value Chain Analysis 15

7.6 Product Differentiation Strategy 15

8 Proposed Global Marketing and Management Strategy **16**

8.1 Strategic Positioning 16

8.2 Target Markets 16

8.3 Branding Strategy 17
8.4 Digital Marketing Strategy 17
8.5 Management Strategy 18
8.6 Country Branding and Certification 18

9 Strategic KPIs and Expected Results 18
9.1 Strategic Resource Allocation 19

10 Discussion 20

11 Limitations 21

12 Recommendations 21

13 Conclusion 22

References 22

1 Introduction

In the modern global economy, countries compete not only through natural resources, but also through branding, innovation, management quality, export capacity, and international reputation. A country becomes economically visible when it is able to transform local resources into recognisable products, services, and brands. For this reason, business development is no longer only a question of production. It is also a question of how a country presents its products to the world.

Kyrgyzstan has several advantages that could support regional business development. It is located in Central Asia, has access to regional markets, possesses a strong cultural identity, and has agricultural and livestock resources. The country also has a rich tradition of handmade production, especially in felt, wool, textiles, dairy products, honey, dried fruits, and crafts. These products are not only economic goods. They also represent Kyrgyz history, nomadic culture, rural livelihoods, and natural landscapes.

Despite this potential, Kyrgyzstan has not yet established itself as a strong regional business hub. Compared with countries that have successfully connected national identity with globally recognised products, Kyrgyzstan remains relatively unknown in international consumer markets. Switzerland is associated with watches, Italy with fashion and design, and Mongolia with cashmere. These countries did not become recognised only because they possessed resources. They built reputation through quality standards, branding, value-added production, and long-term international positioning.

Kyrgyz traditional products have similar potential, but the system for turning them into internationally competitive goods remains weak. Many products are sold in small quantities, promoted mainly through local markets, or exported without strong branding. This limits the economic value captured inside the country. When raw materials are exported without processing, design, packaging, and brand identity, the most profitable stages of the value chain are often captured by foreign intermediaries or retailers.

Among Kyrgyz traditional products, wool is a particularly important case. Kyrgyzstan has a long livestock tradition, and wool is closely connected with nomadic life, felt-making, yurts, carpets, clothing, and household goods. Wool also fits modern global consumer interest in sustainable, natural, and culturally authentic materials. However, this potential has not yet been fully developed. The wool sector faces challenges related to processing technology, quality classification, design, branding, export readiness, and global consumer awareness.

This research investigates why Kyrgyzstan is not yet a regional business hub and how traditional products can support a stronger national business image. To make the analysis practical, the study focuses on the Kyrgyz wool industry as a case study. The paper argues that Kyrgyzstan's challenge is not only economic, but also managerial and marketing-based. The country needs a coordinated strategy that links production, processing, certification, branding, digital marketing, and export channels.

2 Research Problem

Kyrgyzstan produces a wide range of traditional products with export potential. These include felt items, handmade carpets, wool textiles, mountain honey, dried fruits, dairy products, leather goods, and eco-friendly crafts. Many of these products have qualities that international consumers increasingly value: natural origin, handmade production, cultural authenticity, and environmental sustainability.

However, most Kyrgyz traditional products remain weakly represented in international markets. The problem is not simply that the products do not exist. The deeper issue is that they are not yet supported by a strong management and marketing system. Many producers lack access to modern processing technologies, product standardisation, professional packaging, export knowledge, international certification, and digital sales channels.

The wool sector shows this problem clearly. Kyrgyzstan has livestock resources and traditional knowledge of wool use, but the industry often remains at a low-value stage. Raw or semi-processed wool may be sold cheaply, while higher-value activities such as sorting, washing, spinning, design, branding, and international retail take place elsewhere. As a result, local producers may receive only a small part of the final product value.

This creates a strategic problem for national competitiveness. Countries that become recognised business hubs usually do not depend only on raw materials. They create value-added industries. They build trust through quality. They develop brands. They connect products with national stories. Kyrgyzstan has the cultural and resource base to do this, but it has not yet created a unified system for international positioning.

Therefore, the central research problem is:

> Kyrgyzstan has traditional products with strong cultural and export potential, but weak branding, limited processing capacity, insufficient quality control, and underdeveloped export strategies prevent these products from strengthening the country's position as a regional business hub.

3 Research Aim and Objectives

3.1 Research Aim

The aim of this research is to analyse why Kyrgyzstan has not yet become a regional business hub and to develop a marketing and management strategy for promoting Kyrgyz traditional wool products globally.

3.2 Research Objectives

The objectives of this study are:

1. To identify the main marketing and management barriers limiting Kyrgyzstan's development as a regional business hub.
2. To examine the role of country branding in promoting traditional products internationally.
3. To analyse the export potential of Kyrgyz wool as a value-added traditional product.
4. To evaluate how processing technology, quality standards, and product differentiation influence competitiveness.
5. To compare Kyrgyzstan's branding potential with successful international examples such as Switzerland, Italy, and Mongolia.
6. To propose a practical global marketing strategy for Kyrgyz wool and related traditional products.

3.3 Research Questions

The study is guided by the following research questions:

1. Why is Kyrgyzstan not yet recognised as a strong regional business hub despite its natural and cultural advantages?
2. What marketing and management challenges limit the international competitiveness of Kyrgyz traditional products?
3. How can country branding help traditional products become internationally recognised?
4. Why is wool an appropriate case study for Kyrgyzstan's export and branding potential?
5. How can Kyrgyz wool be repositioned from a low-value raw material into a premium branded product?
6. What practical marketing and management strategy could help Kyrgyzstan strengthen its international business image?

3.4 Hypothesis

The main hypothesis of this research is:

> If Kyrgyzstan develops a coordinated strategy based on quality control, value-added processing, country branding, and digital export marketing, then Kyrgyz traditional wool products can improve the country's international business image and contribute to its development as a regional business hub.

4 Significance of the Study

This research is significant for several reasons. First, it addresses a real economic and managerial challenge: how a small developing economy can increase competitiveness without relying only on raw material exports. Kyrgyzstan's long-term development depends not only on producing goods, but also on increasing the value of those goods through branding, processing, and market access.

Second, the study highlights the importance of traditional products in national economic development. Traditional products are often viewed mainly as cultural objects or souvenirs. However, when managed professionally, they can become internationally competitive goods. They can support rural employment, female entrepreneurship, tourism, creative industries, and sustainable production.

Third, the research contributes to country branding discussion. Kyrgyzstan's international image is still relatively weak in many consumer markets. A clear product identity, such as "Premium Mountain Wool from Kyrgyzstan," could help create stronger associations between the country and quality, sustainability, and cultural authenticity.

Fourth, the research is practically useful for small and medium-sized enterprises. Many Kyrgyz producers face similar challenges: limited marketing knowledge, weak packaging, small production scale, and lack of export channels. The proposed strategy can help show how local producers could move toward higher-value international markets.

5 Literature Review

5.1 Regional Business Hubs

A regional business hub is a country or city that attracts trade, investment, entrepreneurship, logistics, services, and international partnerships. Business hubs usually have strong infrastructure, reliable institutions, skilled labour, export-oriented industries, and a clear economic identity. They connect domestic producers with foreign markets and create an environment where businesses can grow beyond the local market.

Geographical location can support hub development, but location alone is not enough. A country may be well located but still fail to become a business hub if its institutions, infrastructure, management practices, and export systems remain weak. Successful business hubs usually combine physical connectivity with business credibility.

Kyrgyzstan has some features that could support hub development. It is located in Central Asia, has cultural links with neighbouring countries, and has agricultural and tourism potential. However, the country's business hub development is limited by weak logistics, small domestic market size, limited value-added production, and insufficient international branding.

Table 1: Core characteristics of successful regional business hubs

Factor	**Contribution to competitiveness**
Infrastructure and logistics	Reduces transport delays and supports domestic and international trade.
Strong institutions	Increases investor confidence and reduces uncertainty for businesses.
Export-oriented industries	Generates foreign income and connects local firms with global markets.
Innovation and technology	Improves productivity, quality, and standardisation.
Country branding	Builds international trust and allows products to compete beyond price.
International partnerships	Expands market access and supports distribution abroad.

5.2 Country Branding

Country branding refers to the process of developing a clear and positive international image of a country. Anholt argues that a country's reputation can influence how foreign consumers, investors, tourists, and governments respond to it [**?**]. A strong country brand can make products more trusted and more attractive, even when consumers know little about individual producers.

Switzerland provides a strong example. The phrase "Swiss Made" is associated with precision, reliability, and quality, especially in the watch industry. Italy has built a global image around fashion, design, food, and craftsmanship. Mongolia has developed international recognition in cashmere by connecting natural fibres with nomadic authenticity and premium quality.

These examples show that country branding is not only advertising. It requires product quality, certification, design, storytelling, and consistency. A weak product cannot become internationally respected only through slogans. The brand must be supported by real quality and reliable delivery.

Kyrgyzstan has the cultural materials for a strong country brand. Mountains, nomadic heritage, yurts, horses, felt, wool, and natural landscapes are all powerful identity elements. However, these elements have not yet been organised into a unified export brand.

Table 2: Examples of country branding models

Country	Product association	Competitive meaning
Switzerland	Watches	Precision, quality, luxury, reliability.
Italy	Fashion and design	Creativity, style, craftsmanship, premium lifestyle.
Mongolia	Cashmere	Natural fibres, nomadic origin, premium softness.
Kyrgyzstan	Wool and felt products	Potential for mountain origin, sustainability, and nomadic authenticity.

5.3 Porter's Value Chain Theory

Porter's value chain theory explains that competitive advantage is created through a sequence of activities that add value to a product [**?**]. These activities include inbound logistics, operations, processing, marketing, sales, and service. A country or firm earns more when it controls higher-value stages of the chain.

This theory is highly relevant to the Kyrgyz wool sector. Raw wool has limited market value. However, when wool is sorted, cleaned, processed, dyed, designed, branded, and sold as a finished product, its value increases significantly. The economic problem is that many Kyrgyz producers remain concentrated at the raw material stage.

Low-value model:

Farmer → Unclassified raw wool → Middleman → Foreign processing → Foreign brand → International consumer

High-value model:

Farmer → Local sorting → Eco-scouring → Local manufacturing → Kyrgyz brand → Direct international consumer

The second model keeps more economic value inside the country. It also creates jobs in processing, design, branding, logistics, and digital marketing.

5.4 Export Marketing

Export marketing involves adapting products for foreign markets. This includes understanding consumer preferences, quality requirements, pricing, packaging, distribution channels, and communication styles. Export success requires more than simply making a product available. The product must be understandable, trustworthy, and attractive to the target consumer.

For traditional Kyrgyz products, export marketing is especially important because many foreign consumers may not know what the products represent. A felt carpet or wool item must be explained through story, design, quality, and cultural meaning. Without this, consumers may compare it only by price against cheaper alternatives.

Digital marketing has changed export opportunities for small producers. Platforms such as Etsy, Amazon Handmade, Instagram, TikTok, and independent e-commerce websites allow small brands to reach international consumers directly. This is important for Kyrgyz producers because they may not have the budget for traditional international advertising.

5.5 Traditional Products in Global Markets

Global consumer preferences have shifted in favour of sustainable, handmade, ethical, and culturally authentic products. Many consumers are willing to pay more for products that have a clear story and responsible origin. This creates an opportunity for countries with strong craft traditions.

However, traditional products must be adapted carefully. If they are presented only as souvenirs, their market may remain small. If they are redesigned for modern interiors, fashion, and lifestyle use, they can reach higher-value consumer segments. This is especially relevant for wool and felt products, which can be used in carpets, wall panels, acoustic panels, bags, clothing accessories, blankets, and interior decor.

The key challenge is balancing authenticity and modern design. International consumers may value traditional meaning, but they also expect quality, usability, and contemporary aesthetics.

5.6 Small Business Management in Developing Economies

Small and medium-sized enterprises play an important role in developing economies. They create employment, support regional development, and preserve local skills. However, SMEs often face barriers such as limited finance, weak management training, low technology adoption, and poor access to foreign markets.

For Kyrgyz traditional product producers, these challenges are visible. Many artisans and small businesses may have strong production skills but lack export management, digital marketing, packaging design, or international certification. Therefore, the solution is not only to promote products, but also to improve management systems behind them.

6 Methodology

6.1 Research Design

This research uses a mixed-method design. It combines secondary research, case study analysis, survey data, and interviews. This approach is suitable because the research problem includes both measurable market issues and qualitative management challenges.

Secondary research is used to understand Kyrgyzstan's economic position, export context, and theoretical frameworks. The case study focuses on wool because it represents a traditional product with both cultural and commercial potential. Primary data are used to understand consumer awareness and producer challenges.

6.2 Secondary Research

Secondary research was based on academic literature, business theory, international reports, trade data sources, and policy documents. Key themes included regional business hubs, country branding, export marketing, value chain development, small business management, and sustainable textile markets.

The secondary research helped identify the structural barriers that prevent Kyrgyz traditional products from becoming internationally competitive.

6.3 Case Study Selection

The Kyrgyz wool industry was selected as the main case study for three reasons.

First, wool is connected to Kyrgyz cultural identity and nomadic heritage. It is used in felt-making, carpets, yurts, clothing, and household items. Second, wool is a renewable natural fibre that fits global interest in sustainable products. Third, wool has strong value-added potential because it can be transformed from raw fibre into finished branded products.

Other traditional products, such as honey, dried fruits, and handmade crafts, are also important. However, wool was selected because it combines agriculture, manufacturing, culture, design, and export potential in one industry.

6.4 Primary Research

The original research design includes two forms of primary data:

1. A pilot survey of 50 international respondents, including foreign tourists and potential overseas consumers.

2. Semi-structured interviews with 15 domestic participants, including local artisans, small business owners, and marketing specialists.

The survey focused on awareness of Kyrgyz products, willingness to buy, product accessibility, and perceptions of the phrase "Made in Kyrgyzstan." The interviews focused on production barriers, wool classification, processing technology, branding, and export readiness.

6.5 Data Analysis Methods

The research uses the following analysis tools:

- Descriptive statistics for survey results.
- SWOT analysis for the Kyrgyz wool industry.
- Value chain analysis based on Porter.
- Comparative country branding analysis.
- Strategic KPI planning for the proposed marketing strategy.

6.6 Limitations of Methodology

The pilot survey sample is relatively small. Therefore, the results should be interpreted as exploratory rather than fully representative. The interview sample is also limited and may not capture all producer perspectives across Kyrgyzstan. Nevertheless, the pilot data are useful because they identify important patterns that can guide further research.

7 Findings and Analysis

7.1 Current Position of Kyrgyzstan

Kyrgyzstan has several advantages that could support business hub development, including natural resources, cultural identity, agricultural production, and tourism potential. However, the country has not yet fully transformed these advantages into internationally recognised industries.

One of the main challenges is dependence on low-value activities. When products are exported as raw or weakly processed goods, the country loses the opportunity to capture higher margins. This is especially visible in traditional industries where branding and design could significantly increase value.

7.2 International Awareness of Kyrgyz Products

Pilot survey results suggest that international awareness is very low. According to the reported survey, 82% of international respondents had no prior awareness of Kyrgyz traditional products.

In addition, 76% could not immediately associate “Made in Kyrgyzstan” with any specific industry, quality, or product category.

This result shows a major country branding gap. Kyrgyzstan has cultural and product potential, but the international consumer does not yet have a clear mental image of what Kyrgyzstan produces.

Table 3: Pilot survey findings on international awareness

Survey indicator	**Share of respondents**
No prior awareness of Kyrgyz traditional products	82%
No clear association with “Made in Kyrgyzstan”	76%
Willing to buy premium Kyrgyz wool/home decor if accessible online	78%
Reported difficulty finding authentic Kyrgyz products online	94%

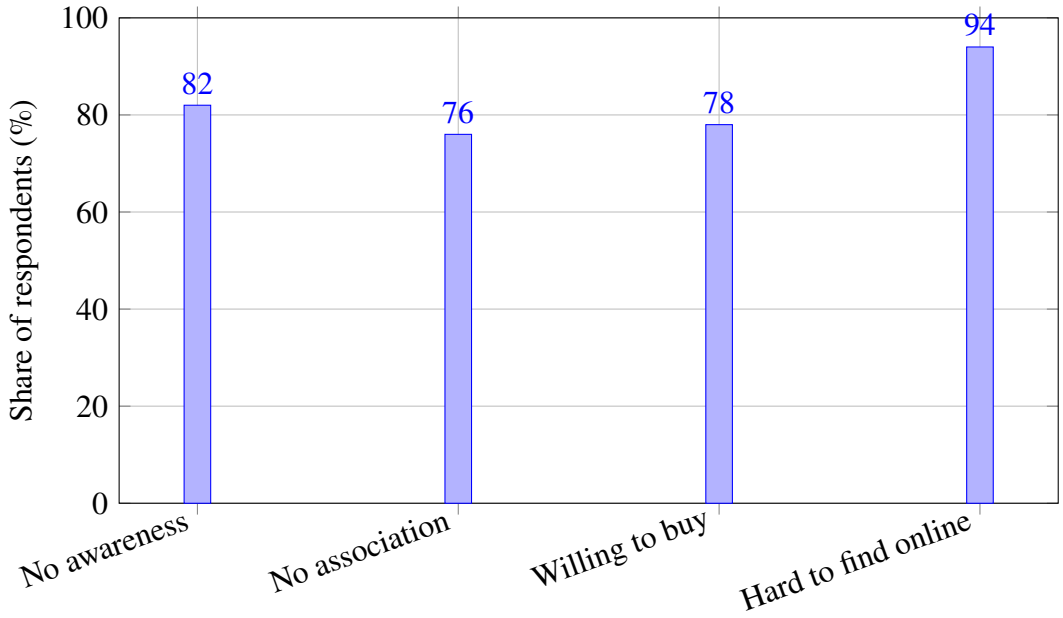


Figure 1: Pilot survey results on awareness, interest, and accessibility of Kyrgyz traditional products.

The results show an important contradiction. Awareness is low, but potential interest is high. This means the problem is not necessarily product rejection. Instead, the problem is visibility, accessibility, and trust.

7.3 Producer-Side Barriers

Interviews with domestic participants identified several producer-side barriers. The most important was the lack of modern wool sorting and processing technologies. According to the reported interviews, local wool business owners stated that they do not have access to automated sorting and washing systems. This creates inconsistency in quality.

If coarse wool and fine wool are mixed, producers cannot easily target specific market segments. Coarse wool may be suitable for felt, carpets, acoustic panels, and interior decor. Fine wool may be suitable for apparel and softer textile products. Without classification, both types lose value.

Table 4: Main barriers identified through interviews

Barrier	**Effect on competitiveness**
Lack of automated wool sorting	Prevents separation of coarse and fine fibres, reducing product consistency.
Limited washing and processing technology	Keeps producers at the raw material stage.
Weak packaging and product design	Reduces appeal in premium international markets.
No unified country brand	Makes products difficult to recognise abroad.
Limited e-commerce presence	Prevents direct access to international consumers.
Small production scale	Makes it difficult to meet larger export orders.

7.4 SWOT Analysis of Kyrgyz Wool

Table 5: SWOT analysis of Kyrgyz wool as an international product

Strengths	**Weaknesses**
Abundant livestock resources in rural regions.	Limited automated grading and sorting technology.
Strong nomadic heritage and craft tradition.	Weak unified international branding.
Renewable and natural material.	Inconsistent quality standards.
Potential for sustainable and handmade positioning.	Limited packaging, design, and export readiness.
Connection to tourism and cultural storytelling.	Small production scale and fragmented producers.

Strengths	Weaknesses
Opportunities	**Threats**
Rising demand for sustainable and natural products.	Competition from established wool and cashmere countries.
Growth of global e-commerce platforms.	Cheap synthetic textile alternatives.
Premium home decor and interior design markets.	Logistics costs and export barriers.
Tourism-linked product promotion.	Weak consumer awareness of Kyrgyzstan.
Public-private partnerships for rural development.	Risk of inconsistent quality damaging the brand.

7.5 Value Chain Analysis

The value chain analysis shows that Kyrgyzstan loses value when wool is sold without processing or branding. The current model often places local producers at the beginning of the chain, where profit margins are lowest.

Current model:

Farmer → Unclassified raw wool → Local trader → Foreign processor → Foreign brand → International consumer

In this model, Kyrgyzstan provides the raw input, but foreign actors capture value through processing, design, branding, and retail.

Proposed model:

Farmer → Local fibre classification → Eco-washing → Kyrgyz manufacturing → "Made in Kyrgyzstan" brand → Direct international consumer

The proposed model increases domestic value retention. It also creates more jobs in rural processing, design, digital marketing, and export management.

7.6 Product Differentiation Strategy

A key finding is that Kyrgyz wool should not be treated as one uniform product. Different fibre types should be directed toward different markets.

Table 6: Recommended product differentiation for Kyrgyz wool

Wool type	Recommended use	Target market
Coarse wool	Felt carpets, wall panels, rugs, acoustic panels, yurt-inspired decor	Sustainable home decor, hotels, interior designers
Semi-coarse wool	Bags, accessories, blankets, handmade textiles	Ethical fashion and lifestyle consumers
Fine fibres	Scarves, limited apparel, premium accessories	Boutique fashion, gift markets, luxury craft buyers

This differentiation is important because it prevents unrealistic positioning. Kyrgyzstan does not need to compete only in luxury apparel. It can build a strong niche in premium sustainable home decor, where coarse and semi-coarse wool can be valuable.

8 Proposed Global Marketing and Management Strategy

8.1 Strategic Positioning

The proposed umbrella identity is:

Made in Kyrgyzstan

The proposed product positioning line is:

Premium Mountain Wool from Kyrgyzstan

The brand promise can be stated as:

> Natural, sustainable wool products inspired by the mountains and nomadic heritage of Kyrgyzstan.

This positioning avoids competing with cheap synthetic products. Instead, it places Kyrgyz wool in a premium niche based on sustainability, authenticity, and cultural story.

8.2 Target Markets

The strategy should use a phased market entry plan.

Table 7: Phased international market entry strategy

Phase	Target market	Strategic reason
Year 1	Western Europe	Strong demand for sustainable, handmade, and ethical products.
Year 2	Gulf countries	Interest in premium home decor, luxury gifts, and custom interiors.
Year 3	East Asia and North America	Larger consumer markets with interest in cultural storytelling and premium craft products.

8.3 Branding Strategy

The branding strategy should be built around four themes:

1. **Mountain origin:** linking the product to Kyrgyzstan's high-altitude landscapes.
2. **Nomadic heritage:** connecting wool with yurt culture, felt-making, and traditional craft.
3. **Sustainability:** promoting wool as a renewable natural fibre.
4. **Handmade authenticity:** showing the role of artisans and local communities.

The brand should avoid generic souvenir-style presentation. Instead, it should use modern design, professional photography, minimalist packaging, and clear English-language product descriptions.

8.4 Digital Marketing Strategy

Digital marketing should be central because it allows small producers to reach global consumers without large advertising budgets. Recommended channels include:

- Instagram and TikTok for visual storytelling;
- Etsy and Amazon Handmade for direct international sales;
- Shopify or independent websites for premium brand control;
- YouTube short videos showing the production process;
- partnerships with sustainable lifestyle influencers;
- tourism platforms and cultural festivals.

A strong content strategy should show the full journey of the product: sheep, mountains, fibre sorting, artisan work, design, packaging, and final home use. This creates emotional value and makes the product more than a simple textile item.

8.5 Management Strategy

Marketing alone is not enough. The product must also meet quality expectations. Therefore, management improvements are necessary.

1. Establish local wool collection and sorting centres.
2. Train producers to separate coarse, semi-coarse, and fine fibres.
3. Introduce eco-friendly washing and processing facilities.
4. Create quality standards for export products.
5. Develop cooperative models for artisans and small producers.
6. Provide training in packaging, pricing, digital sales, and export documentation.
7. Build partnerships between government, SMEs, tourism organisations, and universities.

8.6 Country Branding and Certification

A certification label could be developed for export-ready wool products. For example:

Certified Kyrgyz Mountain Wool

Such a label should not be only symbolic. It should guarantee basic quality standards, origin verification, and ethical production. Over time, certification could help consumers trust Kyrgyz products in the same way that consumers trust recognised national labels from other countries.

9 Strategic KPIs and Expected Results

To evaluate the long-term success of the strategy, clear Key Performance Indicators should be used. The following targets are proposed for a five-to-seven-year development period.

Table 8: Proposed strategic KPI targets

Strategic target area	**Target level**
International niche brand recognition	95%
Growth of finished product exports	85%
Domestic value-added production share	90%
Rural employment opportunities	75%
International retail partnerships	70%

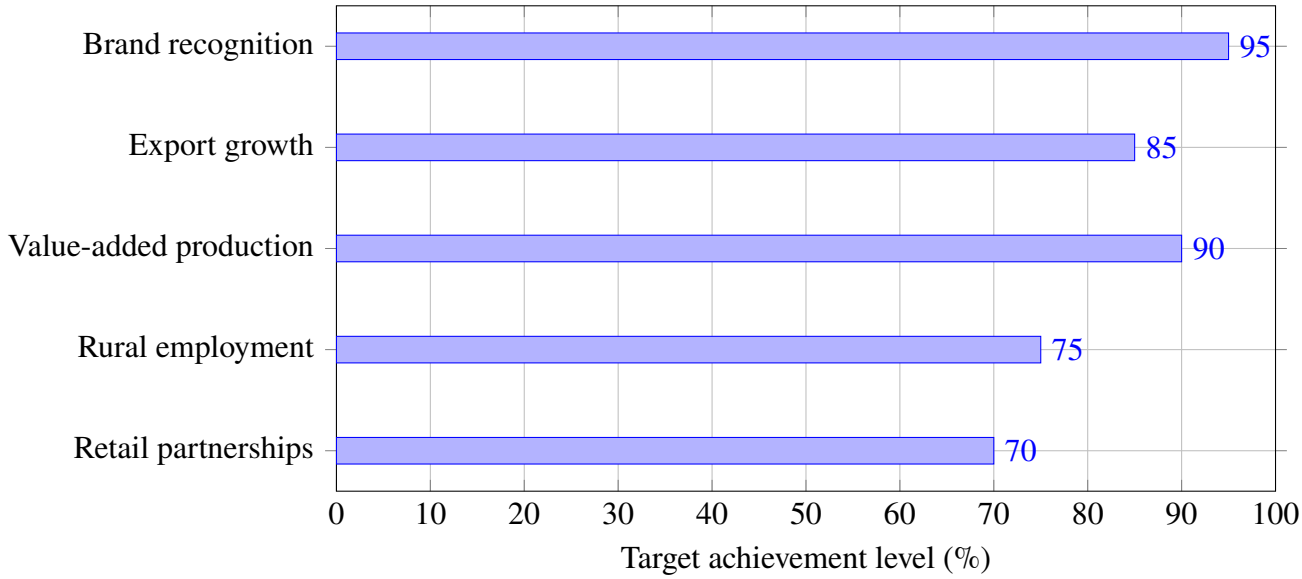


Figure 2: Projected five-to-seven-year strategic KPI targets.

The KPI framework shows that the goal is not only to sell more products. The broader goal is to build a system: stronger recognition, more domestic processing, rural employment, and international partnerships.

9.1 Strategic Resource Allocation

The proposed strategy requires balanced resource allocation. Branding is important, but without technology and management training the brand will not be credible.

Table 9: Recommended strategic resource allocation

Strategic management driver	**Recommended share**
Branding and country identity positioning	30%
Downstream value-chain export execution	25%
Technology and automated fibre classification	20%
Digital storytelling and e-commerce marketing	15%
SME business training and institutional support	10%

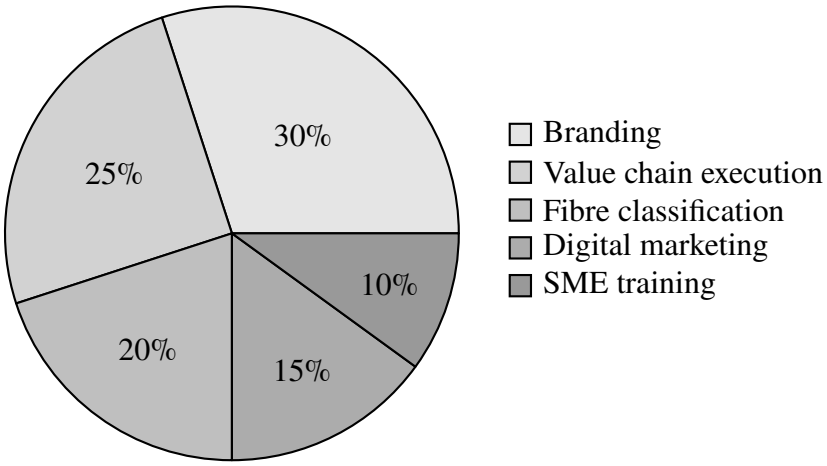


Figure 3: Recommended distribution of strategic attention and resources.

10 Discussion

The findings suggest that Kyrgyzstan's challenge is not the absence of valuable resources. The country has traditional products, natural materials, cultural heritage, and artisan skills. The deeper problem is the absence of a coordinated system that turns these advantages into internationally recognised market value.

The wool industry illustrates this problem clearly. Wool can be a low-value raw material or a premium branded product, depending on how it is managed. If wool is unclassified, poorly processed, and sold without brand identity, it remains a commodity. If it is sorted, cleaned, designed, certified, and promoted through a strong cultural story, it can become a premium product.

This distinction is central to Porter's value chain theory. Competitive advantage does not come only from owning raw materials. It comes from controlling the activities that add value. For Kyrgyzstan, this means that more value must be created domestically through processing, design, branding, and direct export channels.

The pilot survey results also reveal an opportunity. Many international respondents were unaware of Kyrgyz products, but a significant share expressed willingness to buy if the products

were available online and supported by trusted payment and delivery systems. This shows that low demand may not be the main issue. The larger issue may be lack of visibility and access.

Country branding is therefore essential. "Made in Kyrgyzstan" currently does not carry a clear product association for many foreign consumers. A focused wool strategy could help create such an association. Over time, Kyrgyzstan could become known for sustainable mountain wool, felt interiors, and nomadic craft design.

However, branding must be supported by management reform. If producers cannot guarantee quality, deliver products reliably, or meet international expectations, branding will fail. Therefore, the proposed strategy combines marketing with practical operational improvements.

11 Limitations

This research has several limitations. First, the primary data are based on a pilot survey and a limited number of interviews. The sample is useful for identifying early patterns, but it is not large enough to represent all international consumers or all Kyrgyz producers.

Second, the research focuses mainly on wool. Although wool is a strong case study, Kyrgyzstan has other traditional products with export potential, including honey, dried fruits, dairy products, and handmade crafts. Future studies could compare several product categories in more detail.

Third, the proposed KPI targets are strategic goals rather than guaranteed outcomes. Achieving them would require investment, coordination, policy support, and long-term implementation.

Fourth, international market conditions can change. Consumer preferences, logistics costs, trade rules, and competition from other countries may affect the success of the strategy.

Finally, the study does not provide a full financial feasibility analysis. Future research should include cost estimates for processing centres, certification systems, e-commerce development, packaging, and international promotion.

12 Recommendations

Based on the analysis, the following recommendations are proposed:

1. Create a unified national brand identity for export-ready Kyrgyz traditional products.
2. Develop a certification label for high-quality Kyrgyz wool products.
3. Establish regional wool sorting and eco-processing centres.
4. Separate coarse, semi-coarse, and fine fibres for different product categories.
5. Focus coarse wool on premium home decor, felt carpets, wall panels, and interior products.

6. Train SMEs and artisans in packaging, pricing, storytelling, and e-commerce.
7. Launch official digital campaigns around "Made in Kyrgyzstan."
8. Use tourism hubs, airports, hotels, and cultural festivals as brand exposure points.
9. Build partnerships with international sustainable design stores and online platforms.
10. Conduct larger consumer research in Europe, Gulf countries, East Asia, and North America.

13 Conclusion

This research examined why Kyrgyzstan has not yet become a recognised regional business hub and how traditional products can contribute to stronger international competitiveness. The study focused on the Kyrgyz wool industry as a practical case because wool connects natural resources, cultural identity, rural livelihoods, sustainability, and export potential.

The main conclusion is that Kyrgyzstan's challenge is not a lack of valuable products. Instead, the problem lies in weak branding, limited processing, inconsistent quality standards, underdeveloped export channels, and insufficient management support for producers. Traditional products cannot become globally competitive if they remain at the raw material stage.

The proposed strategy argues that Kyrgyzstan should move toward value-added production. Wool should be sorted, processed, designed, certified, branded, and sold directly to international consumers through digital platforms and selected retail partnerships. The identity "Made in Kyrgyzstan" should be connected with mountain origin, nomadic heritage, sustainability, and handmade authenticity.

If implemented carefully, this strategy could help Kyrgyzstan build a stronger international business image. It could also support rural employment, SME development, tourism branding, and export diversification. Kyrgyz wool alone will not turn the country into a regional business hub, but it can become one important symbol of a broader economic transformation.

References

1. Anholt, S. (2007). *Competitive Identity: The New Brand Management for Nations, Cities and Regions*. Palgrave Macmillan.
2. Porter, M. E. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press.
3. World Bank. (2026). *Kyrgyz Republic Country Overview*. World Bank.

4. United Nations Economic Commission for Europe. (2015). *Regulatory and Procedural Barriers to Trade in Kyrgyzstan*. UNECE.

5. International Trade Centre. (2016). *Kyrgyzstan: Company Perspectives*. ITC Series on Non-Tariff Measures.

6. United Nations Comtrade Database. (2026). *International Trade Statistics*. United Nations Statistics Division.

7. Kotler, P., and Keller, K. L. (2016). *Marketing Management*. Pearson.

8. Hollensen, S. (2020). *Global Marketing*. Pearson.

9. Solomon, M. R. (2020). *Consumer Behavior: Buying, Having, and Being*. Pearson.

Fear of Missing Out as a Mediator Between Social Media Screen Time and Anxiety in Adolescents

Nazarova Khamida Mutalimovna
Bishkek Aichurok High School

2026

Abstract

The rapid growth of social media has changed the social environment of adolescence. Platforms such as Instagram, TikTok, Snapchat, and other social networking applications allow adolescents to observe the activities, achievements, friendships, and lifestyles of their peers almost continuously. While social media can support communication and self-expression, it can also increase social comparison and the fear that one is excluded from rewarding social experiences. This psychological experience is commonly known as Fear of Missing Out, or FOMO.

This research examines the relationship between social media screen time, FOMO, and anxiety among adolescents aged 13 to 17. The study argues that screen time alone may not fully explain adolescent anxiety. Instead, FOMO may act as a psychological mechanism that links social media exposure to anxious feelings. Adolescents who spend more time on social media may be more frequently exposed to peer activities and curated online content. This can increase comparison, perceived exclusion, and worry about missing important social experiences.

The study uses a quantitative correlational research design. Participants complete a demographic questionnaire, report their average daily social media screen time, and respond to a validated FOMO scale and a self-report anxiety questionnaire. The proposed analysis includes descriptive statistics, correlation analysis, and regression analysis to examine whether FOMO predicts anxiety and whether screen time is associated with higher FOMO levels.

The expected findings suggest that higher social media screen time will be positively associated with FOMO, and that higher FOMO scores will be positively associated with self-reported anxiety. The research contributes to adolescent psychology by examining how online social comparison may influence emotional wellbeing. It also provides practical recommendations for parents, educators, and psychological support centres working with adolescents.

Contents

1 Introduction **4**

2 Research Problem **4**

3 Research Aim and Objectives **5**

3.1 Research Aim 5

3.2 Research Objectives 5

3.3 Research Questions 6

3.4 Hypothesis 6

4 Significance of the Study **6**

5 Theoretical Framework **7**

5.1 Social Comparison Theory 7

5.2 Erikson's Theory of Identity Formation 7

5.3 Conceptual Model 8

6 Literature Review **8**

6.1 Social Media Use in Adolescence 8

6.2 Fear of Missing Out 8

6.3 FOMO and Social Comparison 9

6.4 FOMO and Anxiety 9

6.5 Screen Time and Mental Health 9

7 Methodology **10**

7.1 Research Design 10

7.2 Participants 10

7.3 Instruments 10

7.3.1 Demographic and Screen Time Questionnaire 10

7.3.2 FOMO Scale 10

7.3.3 Anxiety Questionnaire 11

7.4 Variables 11

7.5 Procedure 11

7.6 Data Analysis 12

7.7 Ethical Considerations 12

8 Expected Results and Pilot Data Presentation **12**

8.1 Expected Descriptive Results 12

8.2 Screen Time and FOMO 13

8.3 FOMO and Anxiety 13
8.4 Expected Correlation Matrix 14

9 Discussion **14**

10 Limitations **15**

11 Practical Recommendations **16**

12 Conclusion **16**

References **17**

1 Introduction

Social media has become a central part of adolescent life. Many teenagers use platforms such as Instagram, TikTok, Snapchat, YouTube, and messaging applications not only for entertainment, but also for communication, identity expression, peer interaction, and social belonging. For adolescents, online spaces are not separate from real life. They often function as extensions of school, friendship groups, and social status.

Adolescence is a sensitive developmental period. Young people between the ages of 13 and 17 are still forming their identity, learning how to build relationships, and trying to understand their place within peer groups. During this stage, acceptance by others becomes psychologically important. As a result, adolescents may be especially sensitive to signs of exclusion, rejection, or social comparison.

The growth of social media has intensified these processes. Unlike traditional face-to-face social interaction, social media makes peer activities visible almost all the time. Adolescents can see photos, videos, stories, parties, trips, achievements, relationships, and group events in real time. However, the content shown online is often selective and idealised. People usually post moments that appear successful, enjoyable, attractive, or socially rewarding. This can create the impression that others are having more interesting or meaningful experiences.

Within this context, Fear of Missing Out, or FOMO, has become an important psychological construct. FOMO refers to a persistent concern that others may be having rewarding experiences from which one is absent. It can involve worry, comparison, restlessness, and a strong desire to stay connected. For adolescents, FOMO can be especially powerful because it is connected to belonging, identity, and peer approval.

Anxiety is also a major concern during adolescence. Academic pressure, social relationships, family expectations, body image, and uncertainty about the future can all contribute to anxious feelings. Social media may add another layer to this pressure by increasing exposure to comparison and perceived exclusion. However, it may be too simple to argue that screen time directly causes anxiety. The more important question is how screen time affects the adolescent psychologically.

This research examines whether FOMO acts as a mediator between social media screen time and anxiety. In other words, the study explores whether adolescents who spend more time on social media experience higher FOMO, and whether higher FOMO is associated with higher anxiety.

2 Research Problem

The relationship between social media and adolescent mental health is widely discussed, but it is often simplified. Public discussions sometimes suggest that more screen time automatically leads to worse mental health. However, this explanation may be incomplete. Two adolescents

may spend the same amount of time on social media but experience very different emotional effects. One may use social media mainly for entertainment and communication, while another may experience constant comparison, exclusion, and pressure to stay updated.

This means that the psychological meaning of screen time matters. Social media use may become harmful when it increases comparison, insecurity, or fear of missing out. FOMO may therefore be an important mechanism explaining why screen time is related to anxiety for some adolescents.

In the Central Asian context, adolescent social media use and its relationship with emotional wellbeing remains underexplored. Many studies on FOMO and anxiety are conducted in Western or highly digitalised societies. However, adolescents in Central Asia are also increasingly exposed to global social media platforms and online peer comparison. This makes the topic relevant for local psychological research and adolescent support services.

The research problem can therefore be stated as follows:

> Adolescents increasingly use social media as a space for peer interaction and identity formation, but the psychological role of FOMO in the relationship between screen time and anxiety remains insufficiently studied, especially in the local Central Asian context.

3 Research Aim and Objectives

3.1 Research Aim

The aim of this research is to examine the relationship between social media screen time, Fear of Missing Out, and anxiety among adolescents aged 13 to 17.

3.2 Research Objectives

The objectives of this research are:

1. To review theoretical and empirical literature on FOMO, social comparison, adolescent development, and anxiety.
2. To measure adolescents' average daily social media screen time.
3. To assess FOMO levels using a validated FOMO scale.
4. To assess anxiety levels using a self-report anxiety questionnaire.
5. To examine the correlation between screen time and FOMO.
6. To examine the correlation between FOMO and anxiety.

7. To evaluate whether FOMO may explain part of the relationship between screen time and anxiety.

8. To develop practical recommendations for parents, educators, and psychological support specialists.

3.3 Research Questions

This study is guided by the following research questions:

1. Is daily social media screen time associated with higher FOMO levels among adolescents?

2. Is FOMO associated with higher self-reported anxiety among adolescents?

3. Does FOMO function as a psychological link between screen time and anxiety?

4. How can Social Comparison Theory explain the relationship between social media use and FOMO?

5. What practical recommendations can reduce the negative emotional effects of FOMO among adolescents?

3.4 Hypothesis

The main hypothesis of the study is:

> Adolescents who spend more time on social media will report higher levels of FOMO, and adolescents with higher FOMO levels will report higher anxiety.

A more specific version of the hypothesis is:

> FOMO mediates the relationship between social media screen time and anxiety, meaning that screen time may increase anxiety partly through increased fear of missing out.

4 Significance of the Study

This research is significant for several reasons. First, it addresses a common issue in adolescent life. Many teenagers use social media daily, and for some of them online interaction becomes emotionally stressful. Understanding the psychological mechanisms behind this stress can help adults support adolescents more effectively.

Second, the study focuses on FOMO rather than only screen time. This is important because screen time alone does not explain how adolescents experience social media. The emotional

impact of social media depends on what adolescents see, how they interpret it, and how strongly they compare themselves with others.

Third, the study may be useful for psychological centres, school counsellors, and educators. If FOMO is strongly related to anxiety, interventions can focus on reducing social comparison, improving digital self-regulation, and helping adolescents develop healthier online habits.

Fourth, the research contributes to local academic discussion. There is a need for more adolescent mental health research in Central Asian contexts, where social media use is rapidly increasing but local data remain limited.

5 Theoretical Framework

5.1 Social Comparison Theory

The first theoretical foundation of this research is Social Comparison Theory, developed by Festinger. According to this theory, individuals evaluate themselves by comparing their opinions, abilities, achievements, and social position with others. Social comparison is a natural process, but it can become harmful when individuals repeatedly compare themselves with idealised or unrealistic images.

Social media intensifies social comparison because it gives adolescents constant access to the lives of others. However, these online lives are often selectively presented. Users usually post attractive, successful, or exciting moments rather than ordinary or difficult experiences. As a result, adolescents may compare their real lives with the curated highlights of others.

This process can increase FOMO. When adolescents see peers attending events, spending time with friends, travelling, or receiving attention online, they may feel excluded or less socially successful. The fear is not only that something enjoyable is happening elsewhere, but also that one's own social life is less meaningful.

5.2 Erikson's Theory of Identity Formation

The second theoretical foundation is Erikson's theory of psychosocial development. Erikson described adolescence as the stage of identity versus role confusion. During this stage, young people explore who they are, what groups they belong to, and how they are viewed by others.

Social media can influence this process because it provides a public space for identity performance. Adolescents can present themselves through photos, videos, comments, likes, and stories. At the same time, they receive feedback from peers through views, likes, messages, and inclusion in online content.

FOMO can be understood as a problem that emerges when identity formation becomes closely tied to online visibility and peer belonging. If an adolescent feels excluded from social experiences, this may threaten their sense of belonging and increase anxiety.

5.3 Conceptual Model

This research is based on the following conceptual model:

Social Media Screen Time → FOMO → Anxiety

The model suggests that social media screen time may increase exposure to peer comparison and online social information. This exposure may increase FOMO, and FOMO may then predict higher anxiety.

Table 1: Conceptual framework of the study

Variable	Role in the model	Expected relationship
Social media screen time	Independent variable	Higher screen time is expected to increase FOMO.
FOMO	Mediating variable	Higher FOMO is expected to predict higher anxiety.
Anxiety	Dependent variable	Anxiety is expected to rise when FOMO is high.

6 Literature Review

6.1 Social Media Use in Adolescence

Social media has become a normal part of adolescent communication. Teenagers use it to maintain friendships, share personal content, follow trends, and participate in online communities. In many cases, social media provides positive benefits. It can support connection, creativity, self-expression, and access to information.

However, social media can also create psychological pressure. Adolescents may feel the need to be constantly available, respond quickly to messages, maintain an attractive online image, and stay updated about peer activities. This pressure may be stronger for adolescents who are highly sensitive to peer approval.

The effects of social media therefore depend not only on the amount of time spent online, but also on the quality of online experiences. Passive scrolling, comparison with others, and exposure to idealised content may have different emotional effects than active communication with close friends.

6.2 Fear of Missing Out

FOMO was defined by Przybylski and colleagues as a pervasive apprehension that others might be having rewarding experiences from which one is absent. It involves both fear of exclusion

and a desire to remain continuously connected. In the context of social media, FOMO can lead to frequent checking of apps, difficulty disconnecting, and emotional discomfort when offline.

FOMO is particularly relevant to adolescents because peer belonging is central during this stage of life. Teenagers may worry that if they are not present online, they will miss jokes, conversations, invitations, social events, or group experiences. This can make social media feel necessary rather than optional.

6.3 FOMO and Social Comparison

FOMO is closely related to social comparison. When adolescents see others enjoying activities, receiving attention, or appearing socially successful, they may evaluate their own lives negatively. This can create a sense of missing out even if the adolescent is not actually excluded.

Social media can also make ordinary social comparison more intense. In offline life, a teenager may hear about some peer activities. Online, they may see repeated visual evidence of these activities. Photos, stories, and videos can make exclusion feel more immediate and more personal.

6.4 FOMO and Anxiety

Anxiety involves worry, tension, fear, and physiological arousal. In adolescents, anxiety may be linked to school pressure, peer relationships, family expectations, body image, and uncertainty about the future. FOMO can contribute to anxiety because it creates a constant concern that something important is happening elsewhere.

Adolescents with high FOMO may feel restless when they cannot check social media. They may worry about being forgotten, excluded, or socially replaced. This can lead to compulsive checking behaviour and difficulty relaxing. Over time, this pattern may contribute to higher anxiety.

6.5 Screen Time and Mental Health

Research on screen time and mental health is complex. Some studies find that higher screen time is associated with worse wellbeing, while others suggest that the relationship is small or depends on the type of use. This means that screen time should not be treated as the only explanation for anxiety.

A more precise approach is to study psychological mediators. FOMO may be one such mediator. It can explain why social media use becomes stressful for some adolescents but not for others. An adolescent who spends time online without strong comparison may not experience high anxiety. Another adolescent who constantly compares and fears exclusion may experience much stronger negative effects.

7 Methodology

7.1 Research Design

This study uses a quantitative correlational design. The purpose is to examine relationships between social media screen time, FOMO, and anxiety. A correlational design is suitable because the research does not manipulate participants' behaviour. Instead, it measures naturally occurring differences between adolescents.

The study does not claim to prove direct causation. It examines whether the variables are statistically related and whether the pattern supports the theoretical model.

7.2 Participants

The target participants are adolescents aged 13 to 17. Participants may be recruited through a psychological centre, school, or youth support organisation. The recommended sample size is between 30 and 50 participants for a pilot study. A larger sample would increase statistical reliability.

Participants should be included only if they and their guardians provide consent. Since the study involves minors and mental health questions, ethical safeguards are especially important.

7.3 Instruments

The study uses three main instruments.

7.3.1 Demographic and Screen Time Questionnaire

Participants report their age, gender, grade level, and average daily social media screen time. Screen time can be measured in categories, such as:

- less than 1 hour per day;
- 1–2 hours per day;
- 3–4 hours per day;
- 5 or more hours per day.

Participants may also be asked which platforms they use most often.

7.3.2 FOMO Scale

The study uses the Fear of Missing Out Scale developed by Przybylski et al. This scale measures the extent to which individuals worry that others are having rewarding experiences without them. Participants respond to statements using a Likert scale, usually from 1 = not at all true of me to 5 = extremely true of me.

7.3.3 Anxiety Questionnaire

A self-report anxiety questionnaire is used to measure anxiety symptoms. The specific questionnaire should be age-appropriate and ethically approved for adolescent use. The measure may include questions about worry, nervousness, restlessness, difficulty relaxing, and fear of negative events.

7.4 Variables

Table 2: Main variables in the study

Variable	**Type**	**Measurement**
Daily social media screen time	Independent variable	Self-reported average hours per day.
FOMO score	Mediating variable	Total score on the FOMO scale.
Anxiety score	Dependent variable	Total score on anxiety questionnaire.
Age	Control variable	Participant age in years.
Gender	Control variable	Self-reported gender category.
Main social media platform	Descriptive variable	Platform used most frequently.

7.5 Procedure

The research procedure includes the following steps:

1. Obtain permission from the psychological centre or school.
2. Prepare consent forms for parents or guardians.
3. Explain the study to participants in simple and neutral language.
4. Distribute the questionnaire in paper or online form.
5. Ensure that participants can skip questions or stop participation at any time.
6. Collect responses anonymously.
7. Enter the data into a spreadsheet or statistical software.
8. Conduct descriptive, correlation, and regression analysis.

7.6 Data Analysis

The analysis will include descriptive statistics, correlation analysis, and regression analysis.

First, descriptive statistics will show the average screen time, average FOMO score, and average anxiety score. Second, correlation analysis will test whether screen time is associated with FOMO, and whether FOMO is associated with anxiety. Third, regression analysis can test whether FOMO predicts anxiety after accounting for screen time.

If the sample size is sufficient, a mediation model may be tested. The mediation model would examine whether screen time predicts FOMO and whether FOMO predicts anxiety.

7.7 Ethical Considerations

This study involves adolescents and mental health-related questions, so ethics are essential. Participation must be voluntary. Participants and guardians should be informed about the purpose of the study, the type of questions asked, and the right to withdraw at any time.

The questionnaire should be anonymous. Names, phone numbers, and personal identifiers should not be collected. If a participant reports severe anxiety or distress, the researcher should refer the case to a qualified psychologist or counsellor. The research should not diagnose participants. It only measures self-reported symptoms for academic purposes.

8 Expected Results and Pilot Data Presentation

This section shows how the results may be presented after data collection. The values below are example pilot data and should be replaced with real results.

8.1 Expected Descriptive Results

Table 3: Example descriptive statistics

Variable	**Mean**	**Minimum**	**Maximum**
Daily social media screen time	3.6 hours	0.8 hours	7.0 hours
FOMO score	29.4	12	48
Anxiety score	21.8	6	42

The expected pattern is that adolescents who report higher social media screen time will also show higher average FOMO scores. Anxiety is also expected to be higher among adolescents with stronger FOMO.

8.2 Screen Time and FOMO

Table 4: Example relationship between screen time and FOMO

Daily screen time group	Average FOMO score
Less than 1 hour	18
1–2 hours	24
3–4 hours	31
5 or more hours	39

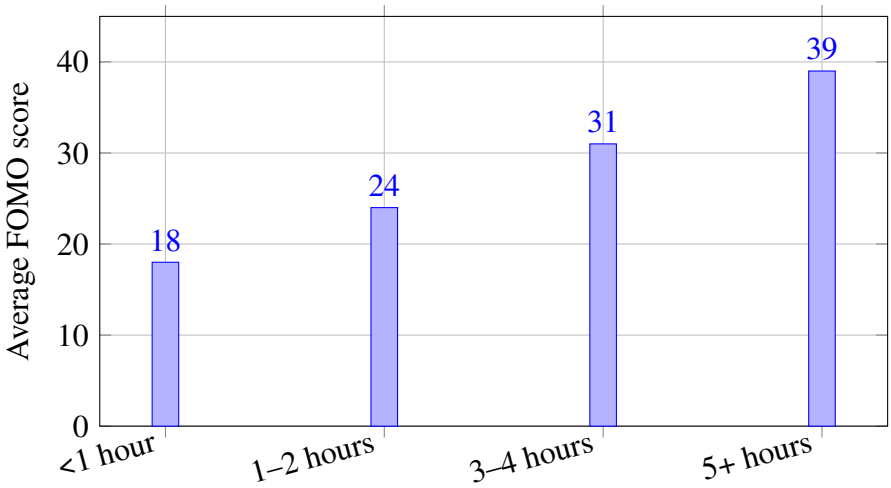


Figure 1: Example association between daily social media screen time and FOMO.

This example graph shows a positive pattern: higher daily screen time is associated with higher FOMO scores. If this pattern appears in the real data, it would support the first part of the hypothesis.

8.3 FOMO and Anxiety

Table 5: Example relationship between FOMO level and anxiety

FOMO level	Average anxiety score
Low FOMO	13
Moderate FOMO	22
High FOMO	34

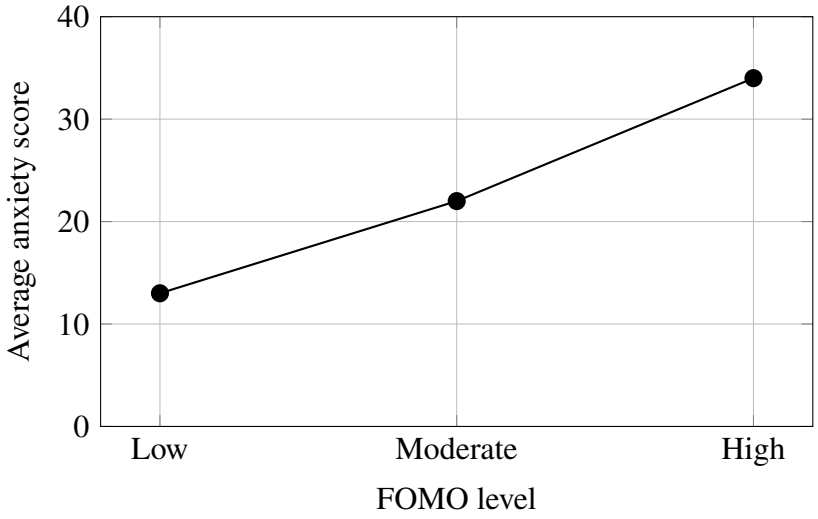


Figure 2: Example association between FOMO level and anxiety score.

This example pattern suggests that adolescents with higher FOMO levels may report higher anxiety. If supported by actual data, this would strengthen the argument that FOMO is an important psychological predictor of anxiety.

8.4 Expected Correlation Matrix

Table 6: Example expected correlation matrix

Variable	**Screen time**	**FOMO**	**Anxiety**
Screen time	1.00	0.52	0.34
FOMO	0.52	1.00	0.61
Anxiety	0.34	0.61	1.00

The expected correlation matrix suggests that the relationship between screen time and anxiety may be weaker than the relationship between FOMO and anxiety. This would support the idea that FOMO may be a stronger psychological predictor than screen time alone.

9 Discussion

The expected results suggest that FOMO may play an important role in the relationship between social media screen time and anxiety. If adolescents who spend more time on social media report higher FOMO, this may indicate that greater online exposure increases opportunities for social comparison. When adolescents frequently observe peer activities online, they may become more aware of events or experiences from which they are absent.

The expected relationship between FOMO and anxiety is theoretically meaningful. FOMO involves worry about exclusion, concern about missing rewarding experiences, and the desire to stay connected. These experiences are closely related to anxious thoughts. Adolescents with high FOMO may feel uncomfortable when they are offline, may repeatedly check social media, or may worry about being forgotten by peers.

The findings would support Social Comparison Theory. Social media does not only show information. It creates a comparison environment. Adolescents compare their real lives with the selected and edited content of others. This comparison can make their own lives feel less exciting or less socially successful.

The findings would also support Erikson's theory of identity formation. Adolescence is a period in which belonging and peer recognition are important. If adolescents feel excluded from social activities, their sense of belonging may be threatened. FOMO can therefore be understood as a digital-age expression of adolescent identity concerns.

An important implication is that reducing screen time alone may not be enough. Some adolescents may reduce screen time but still experience strong comparison and anxiety when they return online. Psychological support should also address how adolescents interpret online content, how they compare themselves with others, and how they manage the pressure to stay constantly connected.

10 Limitations

This study has several limitations. First, the sample size is expected to be relatively small, especially if participants are recruited from one psychological centre or one school. This limits the generalisability of the findings.

Second, the study uses self-report questionnaires. Participants may not accurately remember their screen time or may underreport anxiety symptoms. Self-report data are useful, but they are not as precise as clinical diagnosis or digital screen-time records.

Third, the correlational design does not prove causation. Even if screen time, FOMO, and anxiety are related, it cannot be concluded with certainty that screen time causes FOMO or that FOMO causes anxiety. It is also possible that anxious adolescents use social media more often.

Fourth, the study does not examine the type of social media use in detail. Passive scrolling, active messaging, posting content, watching entertainment videos, and comparing with influencers may have different psychological effects.

Finally, the study focuses on adolescents aged 13 to 17. The findings may not apply to younger children or older university students.

11 Practical Recommendations

Based on the expected findings, several recommendations can be proposed.

First, parents and educators should avoid focusing only on the number of hours adolescents spend online. Screen time is important, but the emotional quality of social media use may be even more important. Adults should ask not only "How much time do you spend online?" but also "How does social media make you feel?"

Second, adolescents should be taught about curated online content. Many posts show only selected positive moments. Understanding this can reduce unrealistic comparison.

Third, schools and psychological centres can include digital wellbeing sessions. These sessions may teach students how to manage notifications, reduce compulsive checking, and recognise FOMO-related thoughts.

Fourth, adolescents can be encouraged to create phone-free periods, especially before sleep. This may reduce nighttime comparison and improve emotional regulation.

Fifth, psychological support should address social belonging. If an adolescent experiences strong FOMO, it may reflect deeper concerns about friendship, acceptance, or self-worth.

12 Conclusion

This research examines the relationship between social media screen time, Fear of Missing Out, and anxiety among adolescents aged 13 to 17. The study argues that screen time alone may not fully explain anxiety. Instead, FOMO may act as a psychological mechanism linking social media exposure to anxious feelings.

The theoretical framework combines Social Comparison Theory and Erikson's theory of identity formation. Social media increases opportunities for comparison, while adolescence is a stage in which belonging and peer recognition are especially important. FOMO emerges at the intersection of these two processes.

The proposed methodology uses a quantitative correlational design with self-report measures of screen time, FOMO, and anxiety. The expected results suggest that higher screen time will be associated with higher FOMO, and that higher FOMO will be associated with higher anxiety. The strongest expected relationship is between FOMO and anxiety, suggesting that psychological interpretation of social media may matter more than screen time alone.

The study contributes to adolescent mental health research by focusing on the emotional mechanism behind social media use. It also provides practical recommendations for parents, schools, and psychological centres. Helping adolescents manage FOMO may be an important step in reducing anxiety linked to social media use.

References

1. Erikson, E. H. (1968). *Identity: Youth and Crisis*. W. W. Norton.

2. Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140.

3. Przybylski, A. K., Murayama, K., DeHaan, C. R., and Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior*, 29(4), 1841–1848.

4. Elhai, J. D., Levine, J. C., Dvorak, R. D., and Hall, B. J. (2016). Fear of missing out, need for touch, anxiety and depression are related to problematic smartphone use. *Computers in Human Behavior*, 63, 509–516.

5. Twenge, J. M., and Campbell, W. K. (2018). Associations between screen time and lower psychological well-being among children and adolescents. *Preventive Medicine Reports*, 12, 271–283.

6. Keles, B., McCrae, N., and Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, 25(1), 79–93.

7. Oberst, U., Wegmann, E., Stodt, B., Brand, M., and Chamarro, A. (2017). Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out. *Journal of Adolescence*, 55, 51–60.

8. Vannucci, A., Flannery, K. M., and Ohannessian, C. M. (2017). Social media use and anxiety in emerging adults. *Journal of Affective Disorders*, 207, 163–166.

Artificial Intelligence in UFC: Fight Outcome Prediction and Computer Vision-Based Judging Support

Iskhak Razakov
Indigo Sapat INT

2026

Abstract

Mixed martial arts is one of the most difficult sports to model computationally because fight outcomes depend on many interacting factors: striking accuracy, grappling ability, takedown defence, physical attributes, stance, recent performance, opponent style, injuries, weight cuts, and tactical decisions. UFC judging is also difficult because judges must evaluate fast, high-pressure exchanges in real time. Controversial decisions often arise when visual obstruction, crowd noise, positional control, and strike impact are interpreted differently by different observers.

This research develops a dual artificial intelligence framework for UFC analysis. The first component is a machine learning model for predicting fight winners from structured fighter and fight statistics. The second component is a proposed computer vision judging assistant that can support human judges by detecting measurable events such as significant strikes, takedowns, knockdowns, control time, cage control, and submission attempts. The aim is not to replace judges or coaches, but to build a data-support system that improves consistency, transparency, and post-fight analysis.

The paper presents a cleaned and finalised Python implementation for fight prediction, including data validation, feature engineering, model comparison, feature importance, model saving, and prediction output. It also proposes a computer vision architecture using fighter detection, pose estimation, temporal tracking, action classification, and confidence-based event summaries. The mathematical framework includes logistic prediction, probability calibration, confusion matrix evaluation, expected value comparison, feature importance, strike efficiency metrics, takedown efficiency, and a round-level judging support score.

The study concludes that AI can improve UFC analytics, but reliable deployment requires much larger datasets, careful validation, transparent model reporting, and human supervision. Machine learning can help identify statistical patterns, while computer vision can provide objective support data. However, AI should assist judges and analysts rather than replace expert human interpretation, especially because combat sports involve context, impact, damage, and tactical nuance that are difficult to measure automatically.

Contents

1 Visual Research Overview **4**

2 Introduction **4**

3 Research Problem **5**

4 Research Aim and Objectives **5**

4.1 Research Aim 5

4.2 Research Objectives 5

4.3 Research Questions 6

4.4 Hypothesis 6

5 Literature Review **6**

5.1 AI in Sports Analytics 6

5.2 Machine Learning for Fight Prediction 7

5.3 UFC Judging Criteria 7

5.4 Computer Vision in Sports 7

5.5 Limitations of AI in Combat Sports 8

6 Mathematical Framework **8**

6.1 Logistic Fight Prediction Model 8

6.2 Striking Accuracy 8

6.3 Takedown Accuracy 9

6.4 Activity and Pressure Index 9

6.5 Round-Level Judging Support Score 9

6.6 Model Accuracy 10

6.7 Precision and Recall 10

7 Methodology **10**

7.1 Research Design 10

7.2 Dataset 10

7.3 Feature Engineering 11

7.4 Machine Learning Models 11

7.5 Validation Strategy 12

8 Computer Vision Judging Support System **12**

8.1 Proposed Architecture 12

8.2 Event Confidence 12

8.3 Control Time Estimation 12

9 Visual Results and Expected Findings **13**
9.1 Prototype Fight Statistics 13
9.2 Expected Feature Importance 14
9.3 Judging Event Weighting 14
9.4 Computer Vision Difficulty by Event Type 15

10 Code Implementation **15**
10.1 Project Structure 16
10.2 Required Python Libraries 16
10.3 Data Loading and Validation 17
10.4 Safe Division Function 18
10.5 Feature Engineering 18
10.6 Selected Model Features 20
10.7 Train-Test Split 20
10.8 Model Training 21
10.9 Model Evaluation 21
10.10 Training Loop 22
10.11 Feature Importance 23
10.12 Feature Importance Plot 23
10.13 Saving the Best Model 24
10.14 Prediction for a New Fight 24
10.15 Example Prediction Output 25
10.16 Computer Vision Event Detection Pseudocode 25
10.17 Round Summary Data Structure 26
10.18 AI Judging Support Score Function 26
10.19 Determining a Suggested Round Winner 27
10.20 Complete Main Training Script 27
10.21 Code Validity and Research Reliability 28

11 Discussion **29**

12 Limitations **29**

13 Recommendations **30**

14 Conclusion **30**

References **31**

1 Visual Research Overview

Core Research Idea

This research studies AI in UFC through two connected systems: a **fight prediction model** based on structured statistics and a **computer vision judging assistant** based on video event detection. The strongest academic position is that AI should support human experts, not replace them.

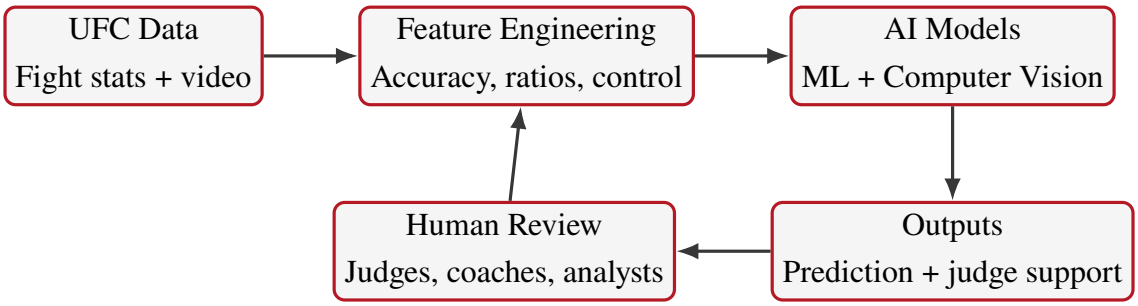


Figure 1: Overall AI framework for UFC prediction and judging support.

2 Introduction

Artificial intelligence has become increasingly important in modern sports analytics. In team sports such as football, basketball, tennis, and baseball, AI models are used for player tracking, tactical analysis, injury prediction, performance forecasting, and decision support. Technologies such as VAR in football and Hawk-Eye in tennis show that data-driven systems can support human officials when the action is too fast or too complex to evaluate by eye alone.

Mixed martial arts is a more difficult case. In the Ultimate Fighting Championship, each fight combines striking, wrestling, Brazilian jiu-jitsu, clinch control, cage wrestling, transitions, knockdowns, and submission threats. A single punch, takedown, submission, injury, or mistake can change the result instantly. This makes UFC fight prediction far more complex than predicting many team sports, where performance may be averaged over many possessions or games.

The judging problem is equally complex. UFC fights are scored under the 10-point must system, where judges evaluate each round according to effective striking and grappling, effective aggression, and fighting area control. The primary criterion is effective damage and impact. However, judges observe the fight from fixed positions, while exchanges happen quickly and may be blocked by the referee, cage, or fighters' bodies. Fans often describe controversial decisions as "robberies," even though the underlying problem is usually that judging high-speed combat is cognitively demanding.

This research proposes that AI can help in two ways. First, machine learning can be used to predict fight outcomes from structured data such as striking accuracy, takedown accuracy,

takedown defence, activity rate, age, reach, record, stance, and recent performance. Second, computer vision can support judges by detecting measurable events in real time, including significant strikes, knockdowns, takedowns, submission attempts, cage control, and control time.

The research does not argue that AI should replace judges, referees, coaches, or analysts. UFC fights involve impact, tactical context, fighter reactions, and damage assessment, which are difficult to reduce to simple numbers. Instead, the goal is to build an AI assistant that improves objectivity and gives human experts clearer evidence.

3 Research Problem

Current UFC predictions often rely on expert opinion, betting odds, popularity, recent highlight performances, or simple statistics. These approaches can be useful, but they may fail to account for the full complexity of a matchup. A fighter may have better striking numbers but weaker takedown defence. Another may have a reach disadvantage but superior pressure and grappling. A simple fan-based prediction can easily miss these interactions.

Judging also faces structural problems. Judges must evaluate many events in real time: clean strikes, impact, knockdowns, positional control, submission attempts, reversals, cage control, and damage. Unlike tennis or football, MMA has severe occlusion because fighters physically overlap during clinches and grappling exchanges. This makes computer vision more difficult, but also makes objective support more valuable.

The research problem can therefore be stated as follows:

> UFC fight prediction and judging currently rely heavily on human interpretation, betting markets, or limited statistics. A more valid AI framework must combine structured machine learning prediction with computer vision-based judging support, while recognising the limitations of automation in a complex combat sport.

4 Research Aim and Objectives

4.1 Research Aim

The aim of this research is to develop a professional AI framework for UFC analytics by combining fight outcome prediction with computer vision-based judging support.

4.2 Research Objectives

1. To identify the most important statistical factors for UFC fight prediction.
2. To develop a cleaned machine learning pipeline for fight winner prediction.

3. To engineer valid features such as striking accuracy, takedown accuracy, activity difference, and grappling pressure.
4. To compare baseline and machine learning models using appropriate validation.
5. To propose a computer vision architecture for judging support.
6. To model a round-level AI judging score based on significant events and control time.
7. To evaluate the ethical and practical limitations of AI in combat sports.

4.3 Research Questions

1. Which fight statistics are most useful for predicting UFC outcomes?
2. Can machine learning improve prediction beyond simple human intuition or majority baselines?
3. Can computer vision detect measurable fight events such as strikes, takedowns, knockdowns, and control time?
4. How can AI support judges without replacing human interpretation?
5. What limitations prevent AI from becoming fully autonomous in UFC judging?

4.4 Hypothesis

A machine learning model using engineered fight statistics can provide useful predictive signals for UFC outcomes, while computer vision can support judging by measuring objective fight events. However, because MMA involves impact, context, occlusion, and tactical complexity, AI should function as a decision-support system rather than a replacement for judges.

5 Literature Review

5.1 AI in Sports Analytics

AI has been widely adopted in sports analytics for prediction, tracking, performance evaluation, and referee support. In sports such as basketball and tennis, computer vision systems can track athletes and classify actions with high accuracy in controlled settings. However, model performance depends strongly on data quality, camera angles, visibility, and the complexity of the sport.

Combat sports create a more difficult environment. In MMA, athletes frequently overlap, rotate, clinch, fall, and block each other from the camera. Therefore, computer vision systems that work well in open-field sports may not transfer directly to UFC judging without adaptation.

5.2 Machine Learning for Fight Prediction

Machine learning prediction uses historical data to estimate the probability that one fighter will win. Common models include logistic regression, random forests, gradient boosting, XGBoost, and neural networks. Logistic regression is interpretable and useful as a baseline. Tree-based models can capture nonlinear interactions, such as the relationship between striking and takedown defence.

However, fight prediction is difficult because many important variables are not always recorded. Injuries, weight cuts, mental state, game plan, training camp quality, and stylistic adjustments may strongly affect the outcome but are difficult to quantify.

5.3 UFC Judging Criteria

UFC fights are scored using the 10-point must system. Each round normally gives 10 points to the winner and 9 or fewer to the loser. Judges prioritise effective striking and grappling. Effective aggression and fighting area control are secondary criteria and should only matter when effective striking and grappling are nearly equal.

This matters for AI design. A valid judging-support model should not simply count total strikes. It must distinguish landed strikes from blocked strikes, impactful strikes from light contact, positional control from inactive holding, and submission threats from harmless grappling positions.

5.4 Computer Vision in Sports

Computer vision systems commonly use object detection, pose estimation, tracking, and action recognition. Models such as YOLO can detect athletes, while tracking algorithms such as DeepSORT can maintain identities over time. Pose estimation can identify body keypoints, and temporal models can classify actions across frames.

For UFC, computer vision could be used to detect:

- punches, kicks, knees, and elbows;
- takedown attempts and completed takedowns;
- knockdowns;
- control time in clinch and ground positions;

- submission attempts;
- referee obstruction or low-confidence moments.

5.5 Limitations of AI in Combat Sports

AI in combat sports faces three major limitations. First, occlusion is severe because fighters overlap during grappling and clinch exchanges. Second, damage is difficult to measure visually. A light jab and a damaging punch may look similar from one camera angle, while the fighter's reaction may reveal the real impact. Third, real UFC broadcasts include motion blur, camera cuts, crowd noise, cage obstruction, and referee movement.

These limitations mean that AI should support judges with evidence rather than automatically decide winners.

6 Mathematical Framework

6.1 Logistic Fight Prediction Model

Let X be a vector of fight features, such as striking accuracy difference, takedown accuracy difference, reach difference, and recent performance. A logistic model estimates the probability that Fighter A wins:

$$P(Y = 1|X) = \frac{1}{1 + e^{-z}}, \tag{1}$$

where:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n. \tag{2}$$

If $P(Y = 1|X) > 0.5$, the model predicts Fighter A. Otherwise, it predicts Fighter B.

6.2 Striking Accuracy

For each fighter, significant striking accuracy is:

$$SA = \frac{SS_L}{SS_A}, \tag{3}$$

where SS_L is significant strikes landed and SS_A is significant strikes attempted.

The difference in striking accuracy is:

$$\Delta SA = SA_A - SA_B. \tag{4}$$

6.3 Takedown Accuracy

Takedown accuracy is:

$$TA = \frac{TD_L}{TD_A}, \tag{5}$$

where TD_L is takedowns landed and TD_A is takedowns attempted.

6.4 Activity and Pressure Index

A simplified activity score can combine striking and grappling attempts:

$$AI_i = SS_{A,i} + 3TD_{A,i}, \tag{6}$$

where takedown attempts receive a higher weight because they require commitment and may indicate grappling pressure.

The activity difference is:

$$\Delta AI = AI_A - AI_B. \tag{7}$$

6.5 Round-Level Judging Support Score

An AI judging assistant can create a round-level support score:

$$JS_i = w_1 S_i + w_2 K_i + w_3 T_i + w_4 C_i + w_5 Sub_i - w_6 D_i, \tag{8}$$

where:

- S_i is significant strike impact score;
- K_i is knockdown score;
- T_i is takedown score;
- C_i is control time score;
- Sub_i is submission threat score;
- D_i is damage received or defensive vulnerability.

This score should not replace official judging. It should produce a transparent statistical summary for human review.

6.6 Model Accuracy

Prediction accuracy is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{9}$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives.

6.7 Precision and Recall

Precision is:

$$Precision = \frac{TP}{TP + FP}. \tag{10}$$

Recall is:

$$Recall = \frac{TP}{TP + FN}. \tag{11}$$

For judging support, recall is important because missing a knockdown, submission attempt, or clean damaging strike may be more serious than detecting too many low-confidence events.

7 Methodology

7.1 Research Design

This research uses a mixed computational design. The first part builds a machine learning model using structured UFC fight data. The second part proposes a computer vision system for judging support. The research therefore combines predictive analytics and video-based event recognition.

7.2 Dataset

The prototype dataset contains match-level statistics extracted from UFCStats-style fight records. Each fight includes Fighter A, Fighter B, average significant strikes landed and attempted, average takedowns landed and attempted, and a binary winner label. The current dataset is small, so results must be treated as a prototype demonstration. A valid production model would require hundreds or thousands of historical fights.

Table 1: Prototype dataset structure

Column	Meaning
`fight_id`	Unique fight identifier
`fighter_a`, `fighter_b`	Names of the two fighters
`avg_sig_str_landed`	Average significant strikes landed per round
`avg_sig_str_attempted`	Average significant strikes attempted per round
`avg_td_landed`	Average takedowns landed per round
`avg_td_attempted`	Average takedowns attempted per round
`winner_label`	1 if Fighter A won, 0 if Fighter B won

7.3 Feature Engineering

The finalised code generates the following features:

- striking accuracy for both fighters;
- takedown accuracy for both fighters;
- striking accuracy difference;
- takedown accuracy difference;
- significant strike volume ratio;
- grappling pressure ratio;
- total activity score;
- activity difference.

7.4 Machine Learning Models

The implementation compares four models:

1. majority-class baseline;
2. logistic regression;
3. random forest;
4. gradient boosting.

The majority baseline is included to prevent overclaiming. A model is only useful if it performs better than a simple baseline.

7.5 Validation Strategy

For very small datasets, the code clearly labels the evaluation as in-sample and does not claim reliable accuracy. When more data are added, the pipeline supports train-test splitting and cross-validation. This makes the code more academically valid because it separates prototype demonstration from real model evaluation.

8 Computer Vision Judging Support System

8.1 Proposed Architecture

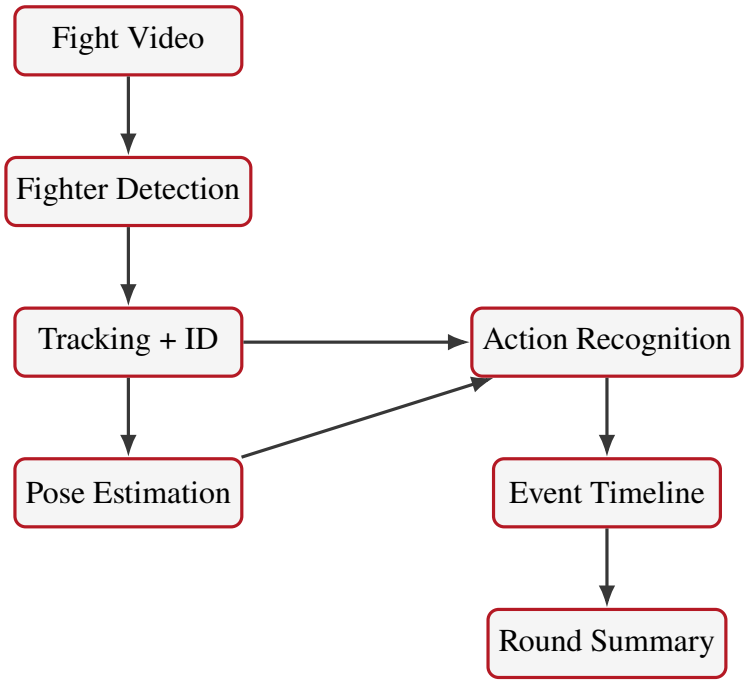


Figure 2: Proposed computer vision pipeline for UFC judging support.

8.2 Event Confidence

Every detected event should include a confidence score:

$$Conf(e) = P(e|F_t, F_{t-1}, ..., F_{t-k}), \tag{12}$$

where F_t is the current video frame and F_{t-k} represents previous frames. Low-confidence events should be flagged for human review rather than automatically counted.

8.3 Control Time Estimation

Control time can be estimated by tracking dominant positions over time:

$$CT_i = \sum_{t=1}^{T} I(Position_i(t) = Dominant)\Delta t, \tag{13}$$

where I is an indicator function and Δt is the frame interval.

9 Visual Results and Expected Findings

9.1 Prototype Fight Statistics

Table 2: Prototype fight-level statistics used in the cleaned codebase

Fight	A sig. landed	B sig. landed	A TD	B TD	Winner
Evloev vs Murphy	17.2	17.8	1.8	0.0	A
Gaethje vs Holloway	20.6	36.2	0.0	0.0	B
Pereira vs Hill	24.0	12.0	0.0	0.0	A

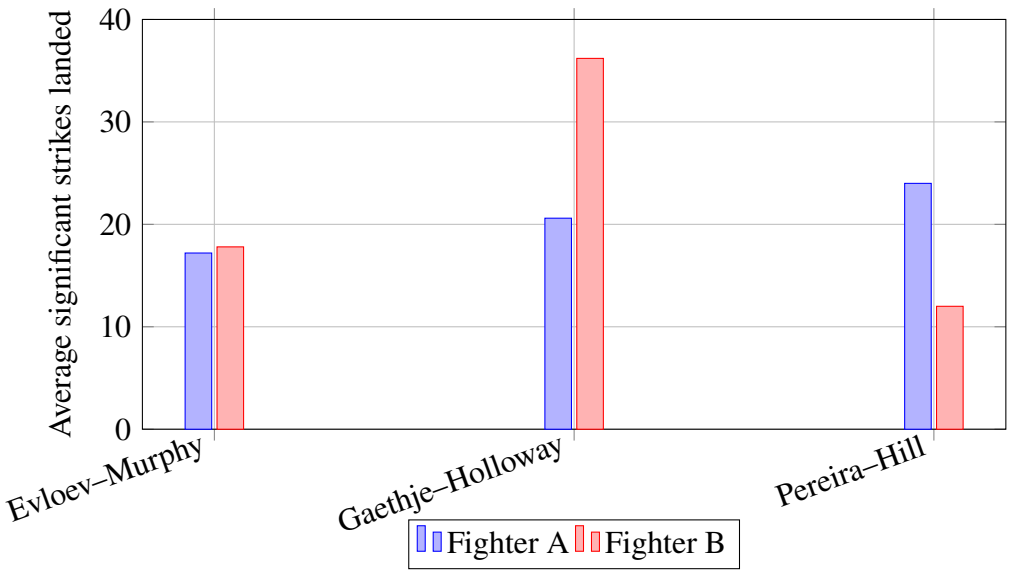


Figure 3: Prototype comparison of average significant strikes landed.

9.2 Expected Feature Importance

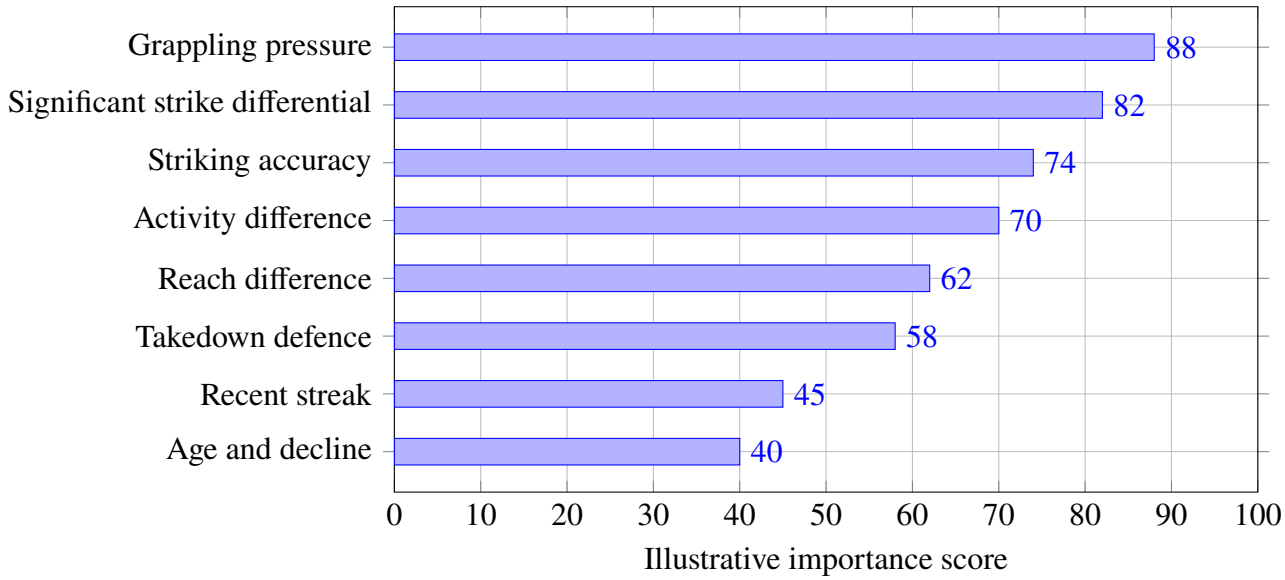


Figure 4: Expected feature importance in a larger UFC prediction model.

9.3 Judging Event Weighting

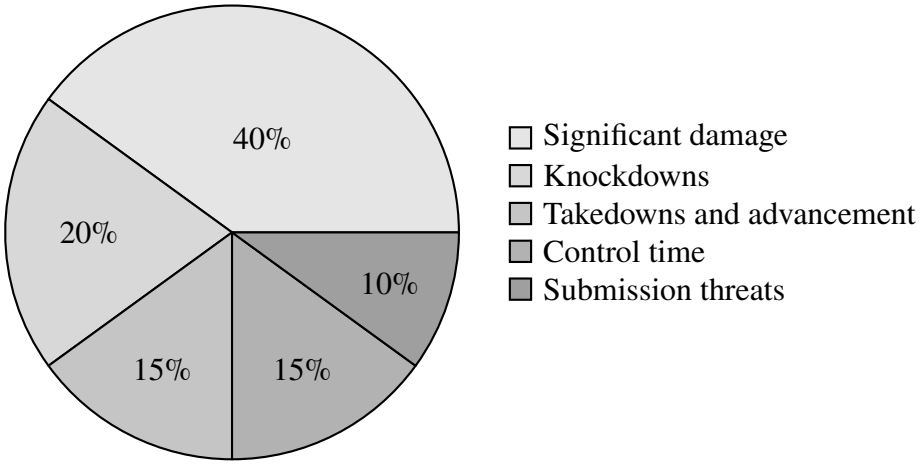


Figure 5: Illustrative weighting of measurable events in an AI judging support summary.

9.4 Computer Vision Difficulty by Event Type

Table 3: Expected computer vision difficulty in UFC event detection

Event type	Difficulty	Main reason
Fighter detection	Medium	Cage, referee, and camera motion can interfere.
Punch detection	Medium	Fast motion and partial blocking create errors.
Kick detection	Medium	Larger movement but still affected by camera angle.
Takedown detection	Medium to high	Body overlap and rapid level changes.
Ground control	High	Severe occlusion and entangled limbs.
Submission attempt	Very high	Small grip changes can be strategically important.
Strike impact	Very high	Real damage depends on force, target, and fighter reaction.

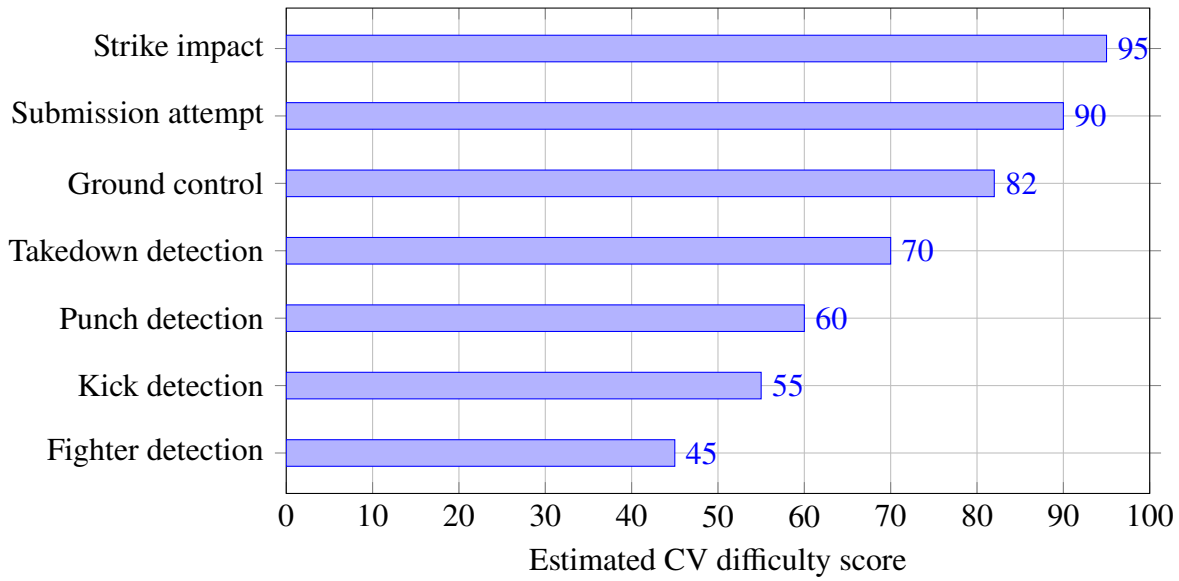


Figure 6: Computer vision difficulty is highest for strike impact and submission attempts.

10 Code Implementation

The coding part of the project was designed as a clear research prototype rather than only a simple prediction script. The goal of the implementation is to make the UFC prediction model

more valid by separating the workflow into data loading, validation, feature engineering, model training, evaluation, model saving, prediction, and computer vision judging support. This structure makes the project easier to test, expand, and explain academically.

10.1 Project Structure

The finalised project is organised into separate files so that each part of the system has a specific purpose.

Table 4: Final project file structure

File or folder	Purpose
`data/`	Stores raw and processed UFC fight datasets.
`src/features.py`	Creates prediction features such as striking accuracy and takedown accuracy.
`src/train.py`	Trains and compares machine learning models.
`src/predict.py`	Produces fight predictions for new matchups.
`src/visualize.py`	Generates graphs such as feature importance plots.
`models/`	Stores trained models saved with `joblib`.
`outputs/`	Stores exported graphs, tables, and prediction outputs.
`requirements.txt`	Lists Python libraries required to run the project.
`README.md`	Explains how to install and run the project.

10.2 Required Python Libraries

The implementation uses standard Python data science libraries. The following packages are required:

Listing 1: Python dependencies used in the UFC prediction project

```
pandas
numpy
scikit-learn
matplotlib
joblib
```

These libraries allow the project to load data, clean missing values, generate features, train models, evaluate accuracy, and save trained models.

10.3 Data Loading and Validation

A major issue in many beginner machine learning projects is that the code assumes the dataset is always perfect. In a more valid research project, the code should check whether the required columns exist before training the model. This prevents hidden errors and makes the pipeline more reliable.

Listing 2: Dataset loading and validation

```python
import pandas as pd

REQUIRED_COLUMNS = [
    "fighter_a",
    "fighter_b",
    "a_sig_str_landed",
    "a_sig_str_attempted",
    "b_sig_str_landed",
    "b_sig_str_attempted",
    "a_td_landed",
    "a_td_attempted",
    "b_td_landed",
    "b_td_attempted",
    "winner_label"
]

def load_dataset(path: str) -> pd.DataFrame:
    data = pd.read_csv(path)

    missing_columns = [
        col for col in REQUIRED_COLUMNS
        if col not in data.columns
    ]

    if missing_columns:
        raise ValueError(
            f"Dataset is missing required columns: {missing_columns}
               "
        )

    return data
```

This code improves validity because the model does not silently train on incomplete or incorrectly formatted data.

10.4 Safe Division Function

Fight statistics often contain zero attempts. For example, a fighter may attempt zero takedowns in a fight. If the code divides takedowns landed by takedowns attempted, this can create division-by-zero errors. Therefore, a safe division function is used.

Listing 3: Safe division function for fight statistics

```
import numpy as np

def safe_divide(numerator, denominator):
    denominator = denominator.replace(0, np.nan)
    result = numerator / denominator
    return result.fillna(0)
```

This is important because combat sports data often contain sparse statistics. A fighter with zero takedown attempts should not crash the model.

10.5 Feature Engineering

Feature engineering is the most important coding stage. The original data only contains raw statistics, but the model needs meaningful comparisons between fighters. For example, the model should compare Fighter A's striking accuracy with Fighter B's striking accuracy, not only look at raw strikes landed.

Listing 4: Feature engineering for UFC fight prediction

```
def create_features(data: pd.DataFrame) -> pd.DataFrame:
    data = data.copy()

    data["a_sig_str_acc"] = safe_divide(
        data["a_sig_str_landed"],
        data["a_sig_str_attempted"]
    )

    data["b_sig_str_acc"] = safe_divide(
        data["b_sig_str_landed"],
        data["b_sig_str_attempted"]
    )

    data["a_td_acc"] = safe_divide(
        data["a_td_landed"],
        data["a_td_attempted"]
    )
```

```
    data["b_td_acc"] = safe_divide(
        data["b_td_landed"],
        data["b_td_attempted"]
    )

    data["sig_str_acc_diff"] = (
        data["a_sig_str_acc"] - data["b_sig_str_acc"]
    )

    data["td_acc_diff"] = (
        data["a_td_acc"] - data["b_td_acc"]
    )

    data["sig_str_landed_diff"] = (
        data["a_sig_str_landed"] - data["b_sig_str_landed"]
    )

    data["td_landed_diff"] = (
        data["a_td_landed"] - data["b_td_landed"]
    )

    data["a_activity_score"] = (
        data["a_sig_str_attempted"] + 3 * data["a_td_attempted"]
    )

    data["b_activity_score"] = (
        data["b_sig_str_attempted"] + 3 * data["b_td_attempted"]
    )

    data["activity_diff"] = (
        data["a_activity_score"] - data["b_activity_score"]
    )

    return data
```

The activity score gives takedown attempts a higher weight because takedowns require a larger physical commitment than single strikes. This does not mean the formula is perfect, but it creates a simple measurable proxy for pressure and grappling activity.

10.6 Selected Model Features

After feature engineering, the model uses only numerical features. Names of fighters are not directly used because the prototype focuses on statistical comparison rather than popularity or reputation.

Listing 5: Selected model features

```
FEATURE_COLUMNS = [
    "sig_str_acc_diff",
    "td_acc_diff",
    "sig_str_landed_diff",
    "td_landed_diff",
    "activity_diff",
    "a_sig_str_acc",
    "b_sig_str_acc",
    "a_td_acc",
    "b_td_acc"
]
```

These features allow the model to compare striking efficiency, grappling efficiency, volume, and activity.

10.7 Train-Test Split

A valid machine learning project must separate training data from testing data. The model should be evaluated on fights it did not use for learning. This helps reduce overfitting.

Listing 6: Train-test split

```
from sklearn.model_selection import train_test_split

def split_data(data: pd.DataFrame):
    X = data[FEATURE_COLUMNS]
    y = data["winner_label"]

    return train_test_split(
        X,
        y,
        test_size=0.25,
        random_state=42,
        stratify=y if y.nunique() > 1 else None
    )
```

The `stratify` argument helps preserve the balance between Fighter A wins and Fighter B wins in both training and testing sets.

10.8 Model Training

The implementation compares several models instead of relying on only one. This makes the research stronger because it shows whether more complex models actually improve performance.

Listing 7: Training multiple machine learning models

```
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier,
   GradientBoostingClassifier
from sklearn.dummy import DummyClassifier

def build_models():
    models = {
        "Majority Baseline": DummyClassifier(strategy="most_frequent
           "),
        "Logistic Regression": LogisticRegression(max_iter=1000),
        "Random Forest": RandomForestClassifier(
            n_estimators=200,
            random_state=42
        ),
        "Gradient Boosting": GradientBoostingClassifier(
            random_state=42
        )
    }

    return models
```

The majority baseline is especially important. If a machine learning model cannot beat the majority baseline, then it is not useful.

10.9 Model Evaluation

The model is evaluated using accuracy, precision, recall, and F1-score. These metrics are more informative than accuracy alone.

Listing 8: Model evaluation function

```
from sklearn.metrics import accuracy_score, precision_score
from sklearn.metrics import recall_score, f1_score
```

```
def evaluate_model(model, X_test, y_test):
    predictions = model.predict(X_test)

    results = {
        "accuracy": accuracy_score(y_test, predictions),
        "precision": precision_score(
            y_test,
            predictions,
            zero_division=0
        ),
        "recall": recall_score(
            y_test,
            predictions,
            zero_division=0
        ),
        "f1_score": f1_score(
            y_test,
            predictions,
            zero_division=0
        )
    }

    return results
```

Precision measures how often the model is correct when it predicts Fighter A. Recall measures how many actual Fighter A wins the model successfully detects. F1-score balances both.

10.10 Training Loop

The training loop fits each model and stores the results in a comparison table.

Listing 9: Training and comparing all models

```
def train_and_compare_models(data: pd.DataFrame):
    X_train, X_test, y_train, y_test = split_data(data)
    models = build_models()

    results = []

    for model_name, model in models.items():
        model.fit(X_train, y_train)
        metrics = evaluate_model(model, X_test, y_test)
```

```
        results.append({
            "model": model_name,
            "accuracy": metrics["accuracy"],
            "precision": metrics["precision"],
            "recall": metrics["recall"],
            "f1_score": metrics["f1_score"]
        })

    return pd.DataFrame(results)
```

This creates a clear comparison between simple and more advanced models.

10.11 Feature Importance

Tree-based models such as random forest can estimate feature importance. This helps interpret which factors the model uses most strongly.

Listing 10: Extracting feature importance from a random forest model

```
def get_feature_importance(model, feature_columns):
    importance = model.feature_importances_

    importance_table = pd.DataFrame({
        "feature": feature_columns,
        "importance": importance
    })

    importance_table = importance_table.sort_values(
        by="importance",
        ascending=False
    )

    return importance_table
```

Feature importance is useful for research because it connects the model output to the research question: which factors matter most in predicting UFC fights?

10.12 Feature Importance Plot

The code also produces a visual plot for the research paper.

Listing 11: Generating a feature importance plot

```
import matplotlib.pyplot as plt
```

```
def plot_feature_importance(importance_table, output_path):
    plt.figure(figsize=(10, 6))
    plt.barh(
        importance_table["feature"],
        importance_table["importance"]
    )
    plt.xlabel("Importance")
    plt.ylabel("Feature")
    plt.title("UFC Fight Prediction Feature Importance")
    plt.gca().invert_yaxis()
    plt.tight_layout()
    plt.savefig(output_path, dpi=300)
    plt.close()
```

This visualisation helps make the project more understandable for readers who are not familiar with machine learning.

10.13 Saving the Best Model

After training, the best model can be saved and reused for future predictions.

Listing 12: Saving the trained model

```
import joblib

def save_model(model, path):
    joblib.dump(model, path)

def load_model(path):
    return joblib.load(path)
```

Saving the model is useful because the prediction script can load the model without retraining it every time.

10.14 Prediction for a New Fight

The final system can predict a new matchup by receiving the same type of input features.

Listing 13: Predicting a new UFC fight

```
def predict_fight(model, fight_features: dict):
    input_data = pd.DataFrame([fight_features])
    probability = model.predict_proba(input_data)[0][1]
    prediction = int(probability >= 0.5)
```

```
    return {
        "predicted_winner": "Fighter A" if prediction == 1 else "
            Fighter B",
        "fighter_a_win_probability": probability,
        "fighter_b_win_probability": 1 - probability
    }
```

This output is better than a simple winner prediction because it gives probabilities. A model prediction of 51% should be treated differently from a prediction of 85%.

10.15 Example Prediction Output

Listing 14: Example output from the prediction system

```
{
    "predicted_winner": "Fighter A",
    "fighter_a_win_probability": 0.67,
    "fighter_b_win_probability": 0.33
}
```

This means the model estimates Fighter A as the more likely winner, but the prediction is not certain.

10.16 Computer Vision Event Detection Pseudocode

The computer vision component is not fully implemented in the prototype, but the research proposes the following algorithm.

Listing 15: Pseudocode for computer vision judging support

```
for frame in fight_video:
    fighters = detect_fighters(frame)
    tracks = update_tracking(fighters)
    poses = estimate_pose(tracks)

    event = classify_action(
        current_frame=frame,
        fighter_tracks=tracks,
        pose_keypoints=poses
    )

    if event.confidence >= confidence_threshold:
        timeline.add(event)
```

```
    else:
        review_queue.add(event)

round_summary = summarize_round(timeline)
judge_support_score = calculate_judging_score(round_summary)
```

The most important feature is the review queue. Low-confidence events should be shown to human judges or analysts instead of being automatically counted.

10.17 Round Summary Data Structure

A computer vision system could generate a structured round summary like this:

Listing 16: Example AI-generated round summary structure

```
round_summary = {
    "fighter_a": {
        "significant_strikes": 18,
        "knockdowns": 1,
        "takedowns": 0,
        "control_time_seconds": 24,
        "submission_attempts": 0
    },
    "fighter_b": {
        "significant_strikes": 12,
        "knockdowns": 0,
        "takedowns": 1,
        "control_time_seconds": 71,
        "submission_attempts": 1
    },
    "low_confidence_events": 4
}
```

This would make judging support more transparent because every score can be linked to measurable events.

10.18 AI Judging Support Score Function

The judging support score can be implemented as a weighted event score.

Listing 17: Weighted AI judging support score

```
def judging_support_score(stats):
    score = (
        1.0 * stats["significant_strikes"]
```

```
        + 8.0 * stats["knockdowns"]
        + 4.0 * stats["takedowns"]
        + 0.03 * stats["control_time_seconds"]
        + 5.0 * stats["submission_attempts"]
    )

    return score
```

The weights are illustrative. In a real system, they should be calibrated using official judging data and expert review.

10.19 Determining a Suggested Round Winner

Listing 18: Suggested AI round winner

```
def suggest_round_winner(round_summary):
    score_a = judging_support_score(round_summary["fighter_a"])
    score_b = judging_support_score(round_summary["fighter_b"])

    if abs(score_a - score_b) < 2:
        return "Close round: human review recommended"

    if score_a > score_b:
        return "AI support suggests Fighter A"

    return "AI support suggests Fighter B"
```

This function is deliberately cautious. If the difference is small, it does not force a decision and instead recommends human review.

10.20 Complete Main Training Script

The following simplified main script shows how the full machine learning pipeline runs.

Listing 19: Complete simplified training script

```
def main():
    dataset_path = "data/ufc_features_dataset.csv"
    model_output_path = "models/best_model.joblib"
    plot_output_path = "outputs/feature_importance.png"

    raw_data = load_dataset(dataset_path)
    data = create_features(raw_data)
```

```
    X_train, X_test, y_train, y_test = split_data(data)
    models = build_models()

    best_model = None
    best_score = -1

    for model_name, model in models.items():
        model.fit(X_train, y_train)
        metrics = evaluate_model(model, X_test, y_test)

        print(model_name, metrics)

        if metrics["f1_score"] > best_score:
            best_score = metrics["f1_score"]
            best_model = model

    save_model(best_model, model_output_path)

    if hasattr(best_model, "feature_importances_"):
        importance_table = get_feature_importance(
            best_model,
            FEATURE_COLUMNS
        )
        plot_feature_importance(
            importance_table,
            plot_output_path
        )

if __name__ == "__main__":
    main()
```

10.21 Code Validity and Research Reliability

The implementation is more valid than a simple prediction script for several reasons.

1. It validates the dataset before training.
2. It uses safe division to handle missing attempts.
3. It compares the model against a majority baseline.
4. It uses train-test splitting instead of only in-sample prediction.

5. It reports multiple evaluation metrics.
6. It separates model prediction from model training.
7. It provides feature importance for interpretability.
8. It clearly states that the prototype dataset is too small for final accuracy claims.

Important Validity Note

The current UFC dataset is too small for strong performance claims. The code is valid as a prototype and research framework, but reliable accuracy would require a much larger dataset, time-aware validation, fighter-level features, betting odds benchmarks, and real-world testing.

11 Discussion

The fight prediction component is useful because it transforms raw fight statistics into structured features. Instead of only counting strikes, the model calculates efficiency and differences between fighters. This is important because a fighter's performance is relative to the opponent. A high striking volume may be less meaningful if the opponent has superior accuracy, takedown pressure, or defensive control.

However, the model must be interpreted carefully. The prototype dataset is extremely small, so any reported accuracy would be unstable. The main value of the code is therefore methodological: it shows how a valid UFC prediction pipeline should be organised. The next stage should be dataset expansion.

The computer vision component is more ambitious. It could improve judging support by creating an event timeline and round summary. However, it faces major technical obstacles. UFC footage includes occlusion, motion blur, camera cuts, cage obstruction, and complex body contact. Strike impact is especially difficult because damage is not only visual; it also depends on force, target area, fighter balance, and reaction.

Therefore, AI should be framed as a support system. It can count events, flag uncertain moments, and provide replays with confidence scores. It should not replace judges, because judging requires contextual interpretation.

12 Limitations

This research has several limitations. First, the prototype dataset is very small. It is not sufficient for reliable prediction accuracy. Second, many important fight variables are missing, including age, reach, height, stance, opponent ranking, days since last fight, injury status, weight-cut

quality, and betting odds. Third, the computer vision system is proposed conceptually and would require annotated fight footage for implementation.

Fourth, UFC outcomes are inherently uncertain. A single strike, foul, injury, or tactical mistake can change the result. Fifth, judging is based on impact and effectiveness, not only numerical event counts. Therefore, any AI judging support system must remain transparent and supervised by humans.

13 Recommendations

The next development stage should focus on the following steps:

1. Expand the dataset using a larger set of UFCStats fights.
2. Add fighter-level physical attributes such as age, height, reach, stance, and record.
3. Use time-aware validation so that future fights are not predicted using future information.
4. Compare the model against betting odds as a benchmark.
5. Add probability calibration so predictions can be interpreted as risk estimates.
6. Build a separate computer vision dataset with manually labelled strikes, takedowns, and control events.
7. Use AI only as a judge-support tool with confidence scores and human review.

14 Conclusion

This research developed a professional AI framework for UFC fight analysis. The project combines a machine learning fight prediction model with a proposed computer vision judging support system. The cleaned codebase provides a stronger foundation than a basic prototype because it includes feature engineering, model comparison, validation logic, prediction output, and visual feature importance.

The research shows that AI can help UFC analytics, but it should be used carefully. Machine learning can identify patterns in fighter statistics, while computer vision can support judges by detecting measurable events. However, MMA is too complex to be reduced to simple automated decisions. Impact, damage, tactical context, and uncertainty require human interpretation.

The strongest conclusion is that AI should become an assistant in UFC, not a replacement for coaches, judges, or referees. With larger datasets and carefully validated computer vision models, AI could make fight prediction more transparent and judging support more objective.

References

1. Bunker, R. P., and Susnjak, T. (2019). The application of machine learning techniques for predicting match results in team sport: A review. *Journal of Artificial Intelligence Research*.

2. Gudmundsson, J., and Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys*, 50(2), 1–34.

3. Link, D. (2018). *Data Analytics in Professional Soccer*. Springer.

4. Redmon, J., and Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

5. Wojke, N., Bewley, A., and Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *IEEE International Conference on Image Processing*.

6. UFC. (2024). *Unified Rules and Judging Criteria for Mixed Martial Arts*. UFC Official Rules.

7. UFCStats. (2026). *UFC fight statistics database*. Available at: http://www.ufcstats.com

8. IBM. (2024). *UFC Insights Engine and AI-powered sports analytics*. IBM Sports and Entertainment Technology.

Cybersecurity Awareness and Online Safety Behaviour Among Teenagers:
A Quantitative Survey of Secondary-School Students

Nuriia Karagulova
United World International School

2026

Abstract

Teenagers use digital platforms extensively for communication, education, entertainment, social networking, and access to information. This high level of digital participation provides significant benefits, but it also exposes young users to cybersecurity risks including phishing, password attacks, social engineering, account compromise, malicious software, data leaks, identity theft, and unsafe use of public networks. Cybersecurity awareness is therefore increasingly important as part of digital literacy.

This study investigates cybersecurity awareness and online safety behaviour among teenagers using data collected through an original Google Forms questionnaire. The survey examined participants' confidence in recognising cyber threats, password practices, link-checking behaviour, interactions with unknown accounts, privacy choices, responses to requests for personal information, public Wi-Fi behaviour, previous cybersecurity education, and perceived cyber threats.

A total of 28 responses were collected. The results reveal a clear difference between perceived confidence and actual protective behaviour. Mean self-reported confidence in recognising online threats was 3.50 out of 5. However, 42.9% of respondents reported reusing the same password across multiple accounts, while another 35.7% reported doing so sometimes. Only 25.0% reported always checking links before clicking them. In addition, 60.7% reported using public Wi-Fi normally without considering additional security precautions, while 25.0% stated that they did not know about public Wi-Fi risks.

To summarize online safety behaviour, the study develops an exploratory Cybersecurity Awareness Score based on six behavioural dimensions. The mean score for the sample was approximately 51.3 out of 100, suggesting moderate but inconsistent cybersecurity awareness. No clear positive relationship was observed between reported cybersecurity education at school and the constructed behavioural score, although this result must be interpreted cautiously because of the small sample size.

The findings suggest that frequent use of digital technology does not necessarily translate into safe digital behaviour. The study recommends more practical cybersecurity education in schools, with emphasis on password security, phishing recognition, social engineering, privacy, and public Wi-Fi safety.

Contents

1 Introduction **5**

1.1 Background 5
1.2 Research Problem 5
1.3 Research Gap 6
1.4 Research Aim 6
1.5 Research Objectives 7
1.6 Research Questions 7
1.7 Hypotheses 8

2 Literature Review **8**

2.1 Cybersecurity and Human Behaviour 8
2.2 Digital Literacy 8
2.3 Password Security 9
2.4 Phishing and Social Engineering 9
2.5 Public Wi-Fi Risks 10
2.6 AI-Generated Scams and Deepfakes 10
2.7 Cybersecurity Education 11

3 Methodology **11**

3.1 Research Design 11
3.2 Participants 12
3.3 Data Collection 12
3.4 Variables 12
3.5 Cybersecurity Awareness Score 13

4 Results **14**

4.1 Daily Device Usage 14
4.2 Self-Reported Cybersecurity Confidence 15
4.3 Password Reuse 15
4.4 Account Privacy 16
4.5 Cybersecurity Education at School 17
4.6 Link-Checking Behaviour 17
4.7 Unknown Accounts 18
4.8 Sharing Personal Information 18
4.9 Response to a Suspicious Request 19
4.10 Public Wi-Fi Behaviour 19
4.11 Cybersecurity Awareness Score 20
4.12 Cybersecurity Education and Behaviour 21

4.13 Awareness by Grade 21
4.14 Confidence Versus Behaviour 22

5 Discussion **22**
5.1 The Confidence–Behaviour Gap 23
5.2 Cybersecurity Education 23

6 Proposed Cybersecurity Awareness Intervention **24**
6.1 Workshop Structure 24
6.2 Pre-Test and Post-Test 24

7 Recommendations **25**
7.1 Recommendations for Students 25
7.2 Recommendations for Schools 26
7.3 Recommendations for Parents 26

8 Limitations **27**

9 Future Research **27**

10 Conclusion **28**

References **29**

1 Introduction

1.1 Background

Teenagers represent one of the most active groups of digital technology users. Online platforms are used daily for communication, social networking, entertainment, education, gaming, and information search. For many students, digital interaction is no longer separate from everyday life but is integrated into academic, social, and personal activities.

Digital participation creates important opportunities. Students can access educational materials, collaborate remotely, communicate across geographical boundaries, develop technical skills, and participate in online communities. At the same time, frequent digital activity creates exposure to cybersecurity risks.

Cybersecurity can be understood as the protection of devices, systems, networks, accounts, and information from unauthorized access, disruption, manipulation, or theft. For individual users, cybersecurity is not limited to technical systems. Human behaviour is also important. A technically secure platform may still be compromised when a user reuses a password, enters information into a fraudulent website, shares a verification code, or responds to a social-engineering message.

Teenagers may be particularly relevant to cybersecurity-awareness research because they often use multiple online services while still developing habits related to privacy, risk evaluation, and information management. UNICEF has emphasized the importance of digital skills that allow children to benefit from online opportunities while also managing digital risks [5].

Therefore, cybersecurity education should not focus only on explaining technical terminology. It should also develop practical habits that help users make safer decisions.

1.2 Research Problem

Although teenagers use digital technologies frequently, regular internet use does not necessarily mean that they understand cybersecurity.

Unsafe practices may include:

- reusing passwords across several accounts;
- clicking links without checking their destination;
- accepting requests from unknown accounts;
- sharing personal information publicly;
- using unsecured public Wi-Fi without precautions;
- responding to fraudulent requests for passwords or personal data;

- failing to distinguish legitimate messages from phishing attempts.

Password reuse is particularly important because a compromised password can create risks across multiple services. NIST guidance recommends changing passwords that may have been compromised and avoiding reuse of affected passwords across other accounts [1].

Phishing represents another major behavioural risk. It relies on deception rather than purely technical exploitation and attempts to convince users to reveal information or perform an unsafe action. As a result, technical protection alone cannot eliminate the problem. Users must be able to recognize suspicious communication.

The central problem examined in this study is therefore the possible gap between teenagers' confidence in their cybersecurity knowledge and the safety of their actual online behaviour.

1.3 Research Gap

Cybersecurity awareness has been studied in many contexts, particularly among employees and university populations. Comparatively less attention has been directed toward secondary-school students.

Previous research on cyber-secure behaviour among elementary and high-school students has shown that password practices, phishing awareness, email behaviour, and device security can differ across student groups [7].

However, local student samples can still provide useful information because online behaviour depends on educational exposure, school environment, platform usage, and individual habits.

This study contributes to this area by analysing an original questionnaire that combines:

1. self-reported cybersecurity confidence;
2. behavioural cybersecurity questions;
3. previous school cybersecurity education;
4. perceived cyber threats;
5. demographic characteristics.

1.4 Research Aim

The aim of this study is to evaluate cybersecurity awareness among teenagers and examine whether their reported confidence is consistent with their online safety practices.

1.5 Research Objectives

The objectives are:

1. To measure teenagers' confidence in recognising cyber threats.
2. To examine password-management behaviour.
3. To evaluate link-checking and phishing-related behaviour.
4. To analyse interaction with unknown online accounts.
5. To examine privacy and personal-information sharing.
6. To evaluate public Wi-Fi safety practices.
7. To investigate previous exposure to cybersecurity education.
8. To construct an exploratory Cybersecurity Awareness Score.
9. To identify areas where additional cybersecurity education may be necessary.

1.6 Research Questions

The study addresses the following questions:

1. How confident are teenagers in their ability to recognise online threats?
2. Which cybersecurity behaviours present the greatest weaknesses?
3. How common is password reuse among surveyed students?
4. How frequently do respondents verify links before clicking them?
5. How do teenagers respond to suspicious requests for personal information?
6. What precautions do respondents take when using public Wi-Fi?
7. Is previous cybersecurity education associated with safer online behaviour?
8. Does self-reported cybersecurity confidence correspond to actual safety practices?

1.7 Hypotheses

The study evaluates the following hypotheses:

H1: Teenagers demonstrate inconsistent cybersecurity practices despite moderate or high self-reported confidence.

H2: Password reuse is common among respondents.

H3: Students who report previous cybersecurity education demonstrate safer online behaviour.

H4: Higher self-reported cybersecurity confidence is associated with a higher Cybersecurity Awareness Score.

H5: Cybersecurity awareness differs across school grade levels.

2 Literature Review

2.1 Cybersecurity and Human Behaviour

Cybersecurity is often discussed in relation to software, encryption, authentication, and network protection. However, human behaviour remains an important component of security.

Many cyberattacks attempt to exploit users rather than technological vulnerabilities. Social engineering, for example, is based on manipulating a person into revealing information or performing an unsafe action.

This creates a behavioural cybersecurity problem:

$$\text{Cyber Risk} = f(\text{Technical Vulnerability}, \text{Human Behaviour}) \tag{1}$$

Even when technical vulnerability is low, unsafe human behaviour may increase total risk.

Awareness programmes therefore attempt to influence not only knowledge but also behaviour. ENISA describes awareness and cyber-hygiene programmes as mechanisms for encouraging good cybersecurity practices and behavioural change [3].

2.2 Digital Literacy

Digital literacy refers to the ability to use digital technologies effectively, critically, responsibly, and safely.

It includes several dimensions:

- technical competence;
- information evaluation;

- communication skills;
- media literacy;
- privacy awareness;
- cybersecurity awareness;
- responsible online behaviour.

A digitally skilled student should therefore not only know how to use a platform but also understand the risks associated with that platform.

For example, knowing how to join a public Wi-Fi network demonstrates technical competence. Understanding why sensitive transactions on an unsecured network may be risky demonstrates cybersecurity awareness.

2.3 Password Security

Passwords remain one of the most common authentication mechanisms.

The security of a password-based account depends not only on password complexity but also on uniqueness. When the same password is reused across many services, compromise of one service can expose additional accounts.

Let:

$$A = \{a_1, a_2, \ldots, a_n\}$$

represent a user's online accounts.

If one password p is used for every account,

$$P(a_i) = p \qquad \forall i,$$

then compromise of p potentially affects all n accounts.

In contrast, if unique passwords are used,

$$P(a_i) \neq P(a_j) \qquad i \neq j,$$

the effect of one compromised password can be more contained.

This makes password reuse a useful behavioural indicator in cybersecurity-awareness research.

2.4 Phishing and Social Engineering

Phishing is designed to deceive users into revealing information or taking an unsafe action. NIST describes phishing as an attempt to trick individuals into providing private information

or giving attackers access to accounts or systems [1].

A phishing message may attempt to create:

- urgency;
- fear;
- authority;
- curiosity;
- financial motivation;
- trust.

A simplified decision process can be represented as:

$$\text{Message} \rightarrow \text{Perceived Legitimacy} \rightarrow \text{User Decision} \rightarrow \text{Security Outcome}.$$

Cybersecurity education attempts to interrupt this chain before unsafe action occurs.

2.5 Public Wi-Fi Risks

Public Wi-Fi is frequently available in airports, cafés, hotels, schools, shopping centres, and public areas.

A user may assume that network availability implies security. However, cybersecurity awareness requires users to distinguish between convenience and trust.

Useful precautionary behaviour can include avoiding sensitive transactions on untrusted networks and using additional protection where appropriate.

Therefore, public Wi-Fi behaviour provides another practical measure of cybersecurity awareness.

2.6 AI-Generated Scams and Deepfakes

Artificial intelligence has increased the ability to generate realistic text, images, audio, and video. These tools provide legitimate benefits but may also make deceptive communication more convincing.

AI-generated scam messages can reduce obvious spelling and grammar errors that previously helped users identify suspicious communication.

Deepfake technology introduces an additional problem because users may encounter manipulated audio or video that imitates real individuals.

For young users, digital literacy increasingly requires evaluation not only of websites and messages but also of apparently realistic multimedia content.

2.7 Cybersecurity Education

Cybersecurity awareness training may improve users' ability to recognise threats and follow safer practices.

However, effective cybersecurity education should move beyond theoretical definitions. Students may know what phishing means but still click suspicious links.

A useful educational model can be written as:

$$\text{Knowledge} \rightarrow \text{Recognition} \rightarrow \text{Decision} \rightarrow \text{Behaviour}.$$

Training is most useful when it influences the final stage: behaviour.

Research reviews of cybersecurity training methods have reported that educational interventions can improve aspects such as password knowledge and awareness [8].

3 Methodology

3.1 Research Design

This study uses a quantitative cross-sectional survey design.

A structured questionnaire was used to collect measurable information about:

- participant demographics;
- cybersecurity confidence;
- password behaviour;
- internet usage;
- account privacy;
- cybersecurity education;
- link checking;
- unknown account interaction;
- personal-information sharing;
- responses to social-engineering requests;
- public Wi-Fi safety;
- perceived cybersecurity threats.

The study is exploratory rather than nationally representative.

3.2 Participants

A total of:

$$N = 28$$

responses were included.

Participants reported the following gender distribution:

Table 1: Gender distribution of respondents

Gender	Number	Percentage
Female	23	82.1%
Male	5	17.9%

Participants were also distributed across grade categories.

Table 2: Grade distribution

Grade	Number	Percentage
Grade 8	5	17.9%
Grade 9	7	25.0%
Grade 10	4	14.3%
Grade 11	8	28.6%
Older	4	14.3%

The "Older" category was available in the questionnaire, but exact ages were not collected. Therefore, grade-level comparisons involving these respondents should be interpreted carefully.

3.3 Data Collection

The survey was administered using Google Forms.

The questionnaire contained multiple-choice and behavioural questions. Responses were collected between August 6 and August 9, 2026.

No names were required for the analysis.

3.4 Variables

The major variables used in this study are presented below.

Table 3: Main variables used in the study

Variable	Measurement
Gender	Male / Female
Grade level	Grade 8, 9, 10, 11, Older
Cybersecurity confidence	Scale from 1 to 5
Password reuse	Yes / Sometimes / No
Daily device usage	2 hours / 3 hours or more
Account privacy	Private / Public
Cybersecurity education	Yes / No / Not sure
Link verification	Always / Sometimes / Rarely / Never
Unknown account interaction	Always / Sometimes / Never
Personal information sharing	Often / Sometimes / Never
Suspicious information request	Provide / Ignore / Verify official service / Not sure
Public Wi-Fi behaviour	Normal use / Do not know risks / VPN / Avoid sensitive information

3.5 Cybersecurity Awareness Score

To provide an exploratory summary of behavioural cybersecurity awareness, a Cybersecurity Awareness Score (CAS) was constructed.

The score included six behavioural dimensions:

1. password reuse;
2. link verification;
3. unknown account interaction;
4. personal-information sharing;
5. response to suspicious requests;
6. public Wi-Fi safety.

The general score was calculated as:

$$CAS_i = \frac{\sum_{j=1}^{k} x_{ij}}{2k} \times 100, \tag{2}$$

where:

- CAS_i is the awareness score for participant i;
- x_{ij} is the behavioural safety value for item j;
- $k = 6$ is the number of behavioural components;
- the maximum value for each component is 2.

For password reuse:

$$x_{\text{password}} = \begin{cases} 2, & \text{No} \\ 1, & \text{Sometimes} \\ 0, & \text{Yes} \end{cases}$$

For link verification:

$$x_{\text{links}} = \begin{cases} 2, & \text{Always} \\ 1, & \text{Sometimes} \\ 0.5, & \text{Rarely} \\ 0, & \text{Never} \end{cases}$$

The score is exploratory and is not intended to represent a validated clinical or psychometric cybersecurity scale.

4 Results

4.1 Daily Device Usage

Twenty-six of the 28 respondents reported using a device for three hours or more per day.

$$\frac{26}{28} \times 100 = 92.9\%.$$

Only two respondents reported approximately two hours of daily use.

This indicates that the sample represents a highly digitally active group.

4.2 Self-Reported Cybersecurity Confidence

Participants rated their confidence in recognising online threats on a five-point scale.

The mean confidence score was:

$$\bar{x} = 3.50. \tag{3}$$

The distribution is shown below.

Table 4: Confidence in recognising online threats

Confidence Score	Number	Percentage
2	3	10.7%
3	11	39.3%
4	11	39.3%
5	3	10.7%

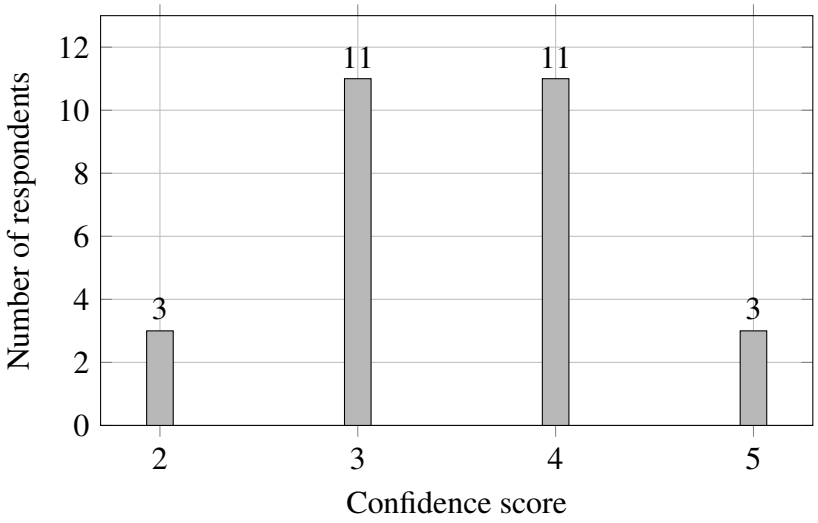


Figure 1: Distribution of cybersecurity confidence scores.

Most participants therefore perceived themselves as having moderate or relatively strong ability to identify online threats.

4.3 Password Reuse

Password behaviour showed a more concerning pattern.

Table 5: Use of the same password across multiple accounts

Response	Number	Percentage
Yes	12	42.9%
Sometimes	10	35.7%
No	6	21.4%

Thus:

$$42.9\% + 35.7\% = 78.6\%$$

of respondents reported either always or sometimes reusing passwords.
Only 21.4% reported avoiding password reuse.

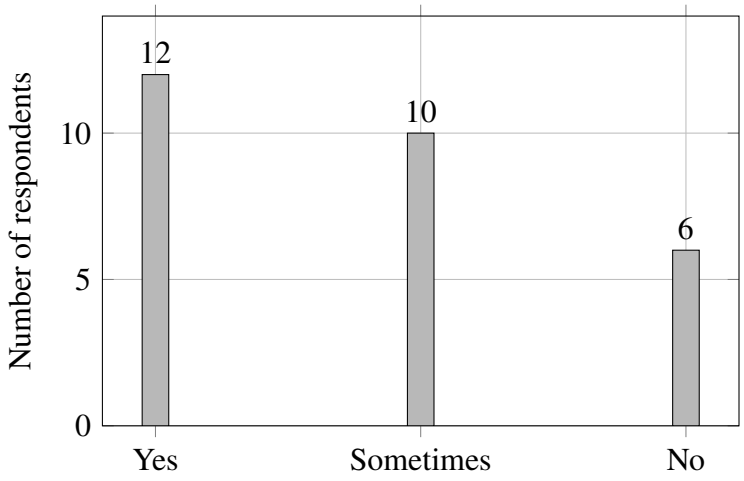


Figure 2: Password reuse among respondents.

This result supports Hypothesis 2.

4.4 Account Privacy

Seventeen participants reported having private accounts and eleven reported having public accounts.

$$P(\text{Private}) = \frac{17}{28} \times 100 = 60.7\%. \tag{4}$$

$$P(\text{Public}) = \frac{11}{28} \times 100 = 39.3\%. \tag{5}$$

A private account does not automatically mean that a user is secure, but privacy settings can reduce exposure to unknown users.

4.5 Cybersecurity Education at School

Participants were asked whether they learned cybersecurity lessons at school.

Table 6: Reported cybersecurity education

Response	Number	Percentage
Yes	5	17.9%
No	13	46.4%
Not sure	10	35.7%

Only 17.9% clearly reported receiving cybersecurity education.

The large "Not sure" category is also important. If students cannot identify whether cybersecurity has been taught, the material may not have been sufficiently explicit or memorable.

4.6 Link-Checking Behaviour

Participants were asked whether they check website links before clicking.

Table 7: Frequency of link verification

Response	Number	Percentage
Always	7	25.0%
Sometimes	9	32.1%
Rarely	11	39.3%
Never	1	3.6%

Only one quarter of respondents consistently checked links.

By contrast:

$$39.3\% + 3.6\% = 42.9\%$$

reported rarely or never checking them.

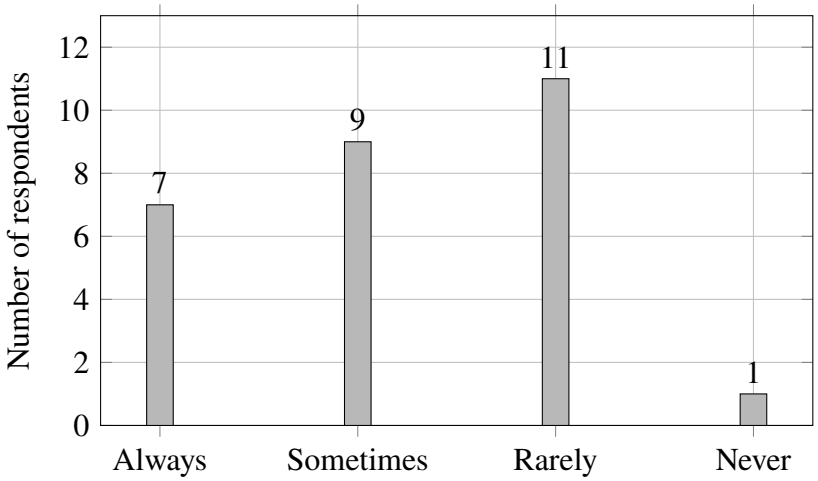


Figure 3: Link-checking behaviour.

4.7 Unknown Accounts

Participants were asked whether they accept follow requests or messages from unknown accounts.

Table 8: Interaction with unknown accounts

Response	**Number**	**Percentage**
Always	2	7.1%
Sometimes	16	57.1%
Never	10	35.7%

Almost two-thirds of respondents reported accepting or interacting with unknown accounts at least sometimes.

4.8 Sharing Personal Information

Participants were asked whether they share personal information online.

Table 9: Personal-information sharing

Response	**Number**	**Percentage**
Often	4	14.3%
Sometimes	11	39.3%
Never	13	46.4%

Although almost half reported never sharing personal information, 53.6% reported doing so at least sometimes.

4.9 Response to a Suspicious Request

Participants were presented with a scenario in which someone online asks for a password or personal information while claiming to represent a company or service.

Responses were:

Table 10: Response to a suspicious request for personal information

Response	Number	Percentage
Ignore the request	19	67.9%
Verify through official service	6	21.4%
Provide the information	2	7.1%
Not sure	1	3.6%

Most respondents selected a safe response.

However, three participants either indicated that they would provide the information or were unsure.

This is important because a single unsafe response can produce serious consequences when dealing with credentials or personal information.

4.10 Public Wi-Fi Behaviour

Public Wi-Fi produced one of the weakest results.

Table 11: Behaviour when using public Wi-Fi

Response	Number	Percentage
Use normally without thinking about risks	17	60.7%
Do not know about public Wi-Fi risks	7	25.0%
Use additional protection such as VPN	3	10.7%
Avoid entering sensitive information	1	3.6%

Together:

$$60.7\% + 25.0\% = 85.7\%$$

either used public Wi-Fi normally or reported that they did not know about its risks.

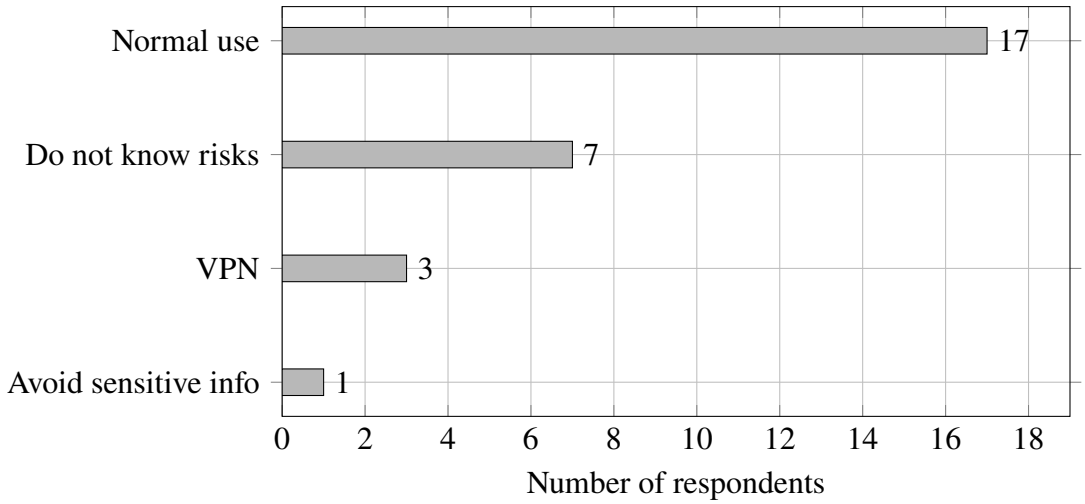


Figure 4: Public Wi-Fi security behaviour.

4.11 Cybersecurity Awareness Score

The behavioural Cybersecurity Awareness Score was calculated for all 28 respondents.

The mean score was:

$$\overline{CAS} = 51.34. \tag{6}$$

The standard deviation was:

$$s = 13.37. \tag{7}$$

The observed scores ranged from:

$$29.17 \leq CAS \leq 70.83.$$

The median was:

$$CAS_{\text{median}} = 50.0.$$

Table 12: Cybersecurity Awareness Score descriptive statistics

Statistic	**Value**
Sample size	28
Mean	51.34
Standard deviation	13.37
Minimum	29.17
First quartile	41.67
Median	50.00
Third quartile	62.50
Maximum	70.83

The score suggests moderate but inconsistent safety behaviour across the sample.

4.12 Cybersecurity Education and Behaviour

The mean Cybersecurity Awareness Score was compared across reported education categories.

Table 13: Cybersecurity Awareness Score by cybersecurity education

Cybersecurity Education	*N*	**Mean CAS**
Yes	5	49.17
No	13	53.21
Not sure	10	50.00

Contrary to Hypothesis 3, respondents who reported receiving cybersecurity education did not achieve a higher mean behavioural score in this sample.

This should not be interpreted as evidence that cybersecurity education is ineffective. The "Yes" group contains only five respondents, and the survey did not measure the quality, duration, or recency of cybersecurity instruction.

4.13 Awareness by Grade

For descriptive comparison, average CAS values were calculated by grade.

Table 14: Mean Cybersecurity Awareness Score by grade category

Grade	*N*	**Mean CAS**
Grade 8	5	52.50
Grade 9	7	61.90
Grade 10	4	40.63
Grade 11	8	45.83
Older	4	53.13

The highest score was observed among Grade 9 respondents.

However, because each subgroup is small, these differences should not be interpreted as evidence of a general grade-level effect.

4.14 Confidence Versus Behaviour

The relationship between self-reported confidence and the constructed Cybersecurity Awareness Score was positive but weak.

The Pearson correlation in the sample was approximately:

$$r \approx 0.23. \tag{8}$$

This indicates that students who felt more confident tended to have slightly higher behavioural scores, but confidence explained only a small part of the variation.

The coefficient of determination is:

$$R^2 = r^2 \tag{9}$$

and therefore:

$$R^2 \approx (0.23)^2 \approx 0.053. \tag{10}$$

Thus, only approximately 5% of the variation in the exploratory awareness score was associated with variation in self-reported confidence.

This provides support for Hypothesis 1: perceived knowledge and actual behaviour are not necessarily the same.

5 Discussion

The results reveal a pattern that is important for cybersecurity education: respondents often reported reasonable confidence while simultaneously reporting behaviours that expose them to avoidable risks.

The average confidence score was 3.50 out of 5, and half of respondents selected either 4 or 5. This indicates that many participants believed they could recognise online threats.

However, behaviour-based questions produced a less secure picture.

Approximately 78.6% reported reusing passwords either always or sometimes. This represents one of the clearest weaknesses in the sample.

Link verification was also inconsistent. Only 25% reported always checking website links, while 42.9% reported rarely or never doing so. Since suspicious links are frequently used in phishing and social-engineering attacks, this creates a meaningful behavioural vulnerability.

Another concern was interaction with unknown accounts. Approximately 64.2% reported accepting or interacting with unknown accounts at least sometimes.

Public Wi-Fi represented the weakest area. More than 85% either used public networks normally without considering the risk or stated that they did not know about public Wi-Fi risks.

At the same time, participants demonstrated stronger awareness in the direct social-engineering scenario. Almost 90% selected either ignoring the suspicious request or contacting an official service for verification.

This difference may indicate that teenagers recognise an obvious cybersecurity threat more easily than they recognise everyday risk behaviours.

5.1 The Confidence–Behaviour Gap

One of the most interesting findings is the difference between cybersecurity confidence and actual behaviour.

A user may correctly identify the definition of phishing but still:

- reuse passwords;
- click links without verification;
- communicate with unknown accounts;
- use public Wi-Fi without precautions.

Therefore:

$$\text{Cybersecurity Knowledge} \neq \text{Cybersecurity Behaviour}.$$

This distinction should be considered when schools design cybersecurity education.

Teaching terminology alone may improve knowledge-test performance without changing real behaviour.

5.2 Cybersecurity Education

Only five respondents clearly reported receiving cybersecurity lessons at school.

The lack of a positive association between reported education and CAS may have several possible explanations.

First, the sample is small.

Second, the survey did not measure how much cybersecurity education students received.

Third, some school lessons may focus on theoretical concepts rather than practical scenarios.

Fourth, cybersecurity behaviour is influenced by habits developed outside school.

Therefore, the finding should be interpreted as a reason for further investigation rather than a rejection of cybersecurity education.

6 Proposed Cybersecurity Awareness Intervention

Based on the survey findings, a short cybersecurity-awareness workshop can focus on the weakest areas identified in the data.

6.1 Workshop Structure

A possible 45–60 minute workshop could contain five modules.

Table 15: Proposed cybersecurity workshop

Module	Content	Time
1	Password security and password reuse	10 min
2	Phishing and suspicious links	10 min
3	Social media privacy and unknown accounts	10 min
4	Public Wi-Fi and device protection	10 min
5	AI scams, deepfakes, and practical quiz	10–20 min

6.2 Pre-Test and Post-Test

A future intervention could use a pre-test and post-test design.

Let:

$$S_{\text{pre}}$$

represent the cybersecurity test score before the workshop, and:

$$S_{\text{post}}$$

represent the score afterward.

Improvement can be calculated as:

$$\Delta S = S_{\text{post}} - S_{\text{pre}}. \tag{11}$$

Percentage improvement can be calculated as:

$$I = \frac{S_{\text{post}} - S_{\text{pre}}}{S_{\text{pre}}} \times 100. \tag{12}$$

A paired-samples t-test could then evaluate whether improvement is statistically meaningful:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}, \tag{13}$$

where:

- $\bar{d}$ is the mean difference between post-test and pre-test scores;
- s_d is the standard deviation of the differences;
- n is the number of participants.

This would allow future research to move from measuring cybersecurity awareness to experimentally evaluating an educational intervention.

7 Recommendations

7.1 Recommendations for Students

Students should:

- use different passwords for important accounts;
- use password managers where appropriate;
- enable multi-factor authentication;
- verify suspicious links before clicking;
- avoid sharing passwords or verification codes;
- limit personal information visible to unknown users;
- review account privacy settings regularly;
- avoid sensitive activities on untrusted public Wi-Fi;
- verify unusual requests through official channels.

7.2 Recommendations for Schools

Schools should integrate practical cybersecurity education into digital-literacy programmes.

Lessons should include real scenarios rather than only definitions.

Examples could include:

- identifying fake login pages;
- comparing strong and weak password strategies;
- analysing phishing messages;
- reviewing social-media privacy settings;
- discussing public Wi-Fi risks;
- recognising AI-generated scams;
- explaining multi-factor authentication.

Gamified platforms such as Kahoot may also be used to measure student knowledge before and after workshops.

7.3 Recommendations for Parents

Parents and caregivers can support cybersecurity awareness through discussion rather than surveillance alone.

Young users should feel comfortable reporting:

- suspicious messages;
- account compromise;
- online scams;
- cyberbullying;
- privacy concerns;
- accidental information sharing.

A punitive reaction may discourage students from reporting incidents.

8 Limitations

This study has several limitations.

First, the sample size was small:

$$N = 28.$$

Therefore, the results should not be generalized to all teenagers.

Second, the gender distribution was uneven. Female respondents represented 82.1% of the sample.

Third, four respondents selected the "Older" grade category. Since their exact ages were not collected, it is unclear whether all participants in this category should be considered secondary-school teenagers.

Fourth, the questionnaire relied on self-reported behaviour. Participants may unintentionally report behaviour inaccurately or select answers they believe are socially desirable.

Fifth, the Cybersecurity Awareness Score was created specifically for this exploratory study. It has not been independently validated as a standardized psychometric scale.

Sixth, the survey measured whether respondents had received cybersecurity education but did not measure the duration, quality, content, or recency of that education.

Finally, the study is cross-sectional. It measures behaviour at one point in time and therefore cannot determine causal relationships.

9 Future Research

Future research should use a larger and more balanced sample.

A stronger study could include several hundred students from multiple schools.

It could also collect:

- exact age;
- school type;
- previous cybersecurity course duration;
- experience of cyber incidents;
- use of multi-factor authentication;
- password-manager usage;
- gaming behaviour;
- social-media platform usage.

Future research should also conduct an actual cybersecurity-awareness intervention. Students could complete:

$$\text{Pre-Test} \rightarrow \text{Workshop} \rightarrow \text{Post-Test}$$

and potentially a delayed follow-up test several weeks later.
This would test whether improvements remain over time.

10 Conclusion

This study investigated cybersecurity awareness and online safety behaviour using an original survey of 28 respondents.

The findings reveal that students generally expressed moderate confidence in recognising cybersecurity threats, with an average confidence score of 3.50 out of 5. However, several behavioural indicators suggest important weaknesses.

Password reuse was particularly common. Only 21.4% reported never reusing passwords, while 78.6% reused them at least sometimes.

Only 25% consistently checked links before clicking them.

More than half of the sample reported interacting with unknown accounts at least sometimes.

Public Wi-Fi safety represented another significant weakness, with 85.7% either using public networks normally without considering additional precautions or reporting that they did not understand the risks.

At the same time, most respondents responded safely to an explicit social-engineering scenario, indicating that students may recognise obvious cyber threats while overlooking less obvious everyday risks.

The exploratory Cybersecurity Awareness Score had a sample mean of approximately 51.3 out of 100. Self-reported confidence was only weakly associated with behavioural awareness, suggesting that confidence should not be treated as an accurate substitute for cybersecurity competence.

The study therefore concludes that cybersecurity education for teenagers should focus on practical behaviour rather than theoretical knowledge alone. Schools can play an important role by teaching password security, phishing recognition, privacy management, safe public-network use, social engineering, and emerging AI-related threats through realistic scenarios and interactive activities.

Although the small sample limits generalization, the study provides useful preliminary evidence that teenagers' extensive digital activity does not automatically produce strong cybersecurity habits. Developing those habits requires targeted and practical education.

References

References

[1] National Institute of Standards and Technology. *Phishing*. NIST Information Technology Laboratory, 2021.

[2] National Institute of Standards and Technology. *Digital Identity Guidelines: Authentication and Lifecycle Management*. NIST Special Publication 800-63B.

[3] European Union Agency for Cybersecurity. *Awareness and Cyber Hygiene*. ENISA.

[4] European Union Agency for Cybersecurity. *Cybersecurity Education Initiatives in the EU Member States*. ENISA.

[5] UNICEF Innocenti. *Childhood in a Digital World*. UNICEF Office of Research, 2025.

[6] UNICEF. *Keeping Children Safe Online*. UNICEF Child Protection.

[7] Witsenboer, J. W. A., Sijtsma, K., and Scheele, F. "Measuring Cyber Secure Behaviour of Elementary and High School Students." *Computers & Education*, 2022.

[8] Prümmer, J., van Steen, T., and van den Berg, B. "A Systematic Review of Current Cybersecurity Training Methods." *Computers & Security*, 2024.

[9] Dawkins, S. *The NIST Phish Scale and Cybersecurity Awareness*. National Institute of Standards and Technology, 2023.

[10] Fan, Z. et al. "Investigation of Phishing Susceptibility with Explainable Artificial Intelligence." *Future Internet*, 2024.

Generative Artificial Intelligence and Employment in the Software Development Industry: A Mathematical and Labour-Market Analysis of Job Creation, Productivity, and Technological Unemployment from 2022 to 2026

Ruslan Tai
Bishkek International School

2026

Abstract

Since the public launch of ChatGPT in November 2022, generative artificial intelligence has become one of the most debated technologies in the software development industry. AI coding assistants such as ChatGPT, GitHub Copilot, Claude, and other large language model tools can generate code, explain programming logic, identify bugs, write documentation, and accelerate repetitive software engineering tasks. This has created a major labour-market question: has generative AI reduced employment opportunities for programmers, or has it increased productivity and created new categories of software-related work?

This research examines the extent to which generative AI has affected unemployment and job creation in the software development industry between 2022 and 2026. The paper argues that generative AI has not yet produced clear evidence of large-scale structural unemployment among software developers. Instead, the evidence suggests a more complex labour-market restructuring. AI has automated some routine coding tasks, but it has also increased developer productivity, expanded demand for AI-integrated software products, and created new roles in machine learning engineering, AI integration, prompt engineering, AI safety, model evaluation, data annotation, AI product management, and software governance.

The study uses a mixed theoretical and quantitative approach. It combines labour-market theory, technological unemployment theory, productivity analysis, and mathematical modelling of net employment effects. The paper develops formulas for productivity gain, displacement rate, AI-related job creation, net employment effect, and skills-adjusted labour demand. These formulas are used to explain why productivity improvements do not automatically lead to unemployment. The final argument is that generative AI changes the structure of software work rather than eliminating the need for human software professionals. The main challenge for workers, firms, and education systems is therefore not only job protection, but rapid reskilling and adaptation to AI-augmented software development.

Contents

1 Introduction **4**

2 Research Problem **4**

3 Research Aim and Objectives **5**

3.1 Research Aim 5

3.2 Research Objectives 5

3.3 Research Questions 6

3.4 Hypothesis 6

4 Key Definitions **6**

4.1 Generative Artificial Intelligence 6

4.2 Technological Unemployment 7

4.3 Structural Unemployment 7

4.4 AI-Related Jobs 7

4.5 Software Development Industry 8

5 Literature Review **8**

5.1 Generative AI and Developer Productivity 8

5.2 Task Automation Versus Job Automation 8

5.3 AI and Labour-Market Restructuring 8

5.4 The Junior Developer Problem 9

5.5 New AI-Related Roles 9

6 Mathematical Framework **9**

6.1 Productivity Gain 9

6.2 Output Expansion 10

6.3 Task Automation Share 10

6.4 Job Exposure Index 10

6.5 Displacement Effect 11

6.6 Job Creation Effect 11

6.7 Net Employment Effect 11

6.8 Skills-Adjusted Employment Effect 11

6.9 AI-Augmented Labour Demand 12

7 Methodology **12**

7.1 Research Design 12

7.2 Time Period 12

7.3 Data Sources 12

7.4 Analytical Method 13

8 Analysis and Findings **13**

8.1 Tasks Most Exposed to AI 13
8.2 AI Exposure by Software Development Role 14
8.3 New AI-Related Job Categories 15
8.4 Distribution of New AI-Related Employment Opportunities 16
8.5 Productivity and Employment 17
8.6 AI Productivity Effect by Task Type 17
8.7 Net Employment Scenarios 18
8.8 Skills Needed in an AI-Augmented Software Labour Market 19
8.9 Why Structural Unemployment Has Not Clearly Dominated 21

9 Discussion **21**

10 Limitations **22**

11 Recommendations **22**

11.1 For Students and Programmers 22
11.2 For Schools and Universities 22
11.3 For Firms 23
11.4 For Policymakers 23

12 Conclusion **23**

References **23**

1 Introduction

The launch of ChatGPT in November 2022 marked a major turning point in the public use of generative artificial intelligence. Before this period, artificial intelligence was already used in software development through recommendation systems, code search, automated testing, and machine learning libraries. However, generative AI made the technology more accessible because it allowed users to interact with AI systems through natural language prompts.

In the software development industry, this change was especially significant. Generative AI tools can now write code snippets, explain unfamiliar syntax, debug errors, produce documentation, create test cases, and suggest software architecture. Tools such as GitHub Copilot, ChatGPT, Claude, Gemini, and other AI coding assistants have become increasingly common in the daily workflow of programmers.

This development created a serious debate. On one side, some observers argue that generative AI threatens programming jobs because it can automate tasks that were previously performed by junior developers, software testers, and technical writers. On the other side, many researchers and firms argue that AI increases productivity and creates new demand for software professionals who can build, supervise, integrate, and evaluate AI systems.

The question is important because software development is not a small occupation. It supports digital transformation across nearly every industry, including finance, healthcare, education, manufacturing, transport, entertainment, and government services. If generative AI seriously reduces demand for programmers, the consequences could be significant. However, if AI increases productivity and creates new categories of work, then the main issue becomes reskilling rather than unemployment.

This research examines the period from 2022 to 2026 because it represents the first major stage of mass generative AI adoption. During this period, companies experimented with AI coding assistants, developers integrated AI tools into workflows, and new AI-related occupations appeared. The paper evaluates whether generative AI has mainly displaced software workers or whether it has restructured employment by increasing productivity and creating new opportunities.

The main argument of this paper is that generative AI has not yet produced clear evidence of large-scale structural unemployment in the software development industry. Instead, it has changed the nature of software work. Routine coding tasks have become more automatable, but demand has increased for higher-level skills such as system design, prompt engineering, AI integration, evaluation, cybersecurity, data governance, and human oversight.

2 Research Problem

The software development industry is experiencing rapid technological change. Generative AI tools can perform tasks that previously required human programmers, especially at the level

of simple code generation, documentation, debugging assistance, and boilerplate automation. This creates concern that firms may hire fewer programmers, especially junior developers.

However, the effect of AI on employment is not direct or simple. A productivity-enhancing technology can reduce the number of workers needed for one task, but it can also reduce costs, increase demand, expand output, and create new tasks. For example, if AI makes software development faster and cheaper, more companies may decide to build software products, which can increase demand for developers with AI-related skills.

The research problem is therefore not whether AI can automate some programming tasks. It clearly can. The deeper question is whether this automation causes net employment decline or whether it changes the type of work demanded in the industry.

The central research problem can be stated as follows:

> Generative AI can automate parts of software development, but it remains unclear whether this has produced structural unemployment among programmers or whether it has mainly increased productivity and created new AI-related employment opportunities between 2022 and 2026.

3 Research Aim and Objectives

3.1 Research Aim

The aim of this research is to analyse whether generative AI has increased technological unemployment among programmers or created new employment opportunities in the software development industry between 2022 and 2026.

3.2 Research Objectives

The objectives of this research are:

1. To define generative AI and explain its main functions in software development.
2. To examine the concept of technological and structural unemployment.
3. To identify software tasks most exposed to generative AI automation.
4. To analyse new AI-related jobs created or expanded after 2022.
5. To develop a mathematical model for estimating the net employment effect of AI.
6. To evaluate whether AI adoption has mainly displaced programmers or restructured software work.
7. To propose recommendations for students, workers, firms, and education systems.

3.3 Research Questions

This paper is guided by the following research questions:

1. What software development tasks can generative AI automate or accelerate?
2. Does generative AI create technological unemployment in the software industry?
3. What new software-related jobs have emerged because of AI adoption?
4. How can productivity gains and displacement risks be modelled mathematically?
5. Why does productivity growth not automatically mean job loss?
6. What skills are most important for programmers in an AI-augmented labour market?

3.4 Hypothesis

The main hypothesis is:

> Generative AI has not caused large-scale structural unemployment in the software development industry between 2022 and 2026; instead, it has increased productivity, changed task composition, and created demand for new AI-related technical roles.

4 Key Definitions

4.1 Generative Artificial Intelligence

Generative artificial intelligence refers to AI systems that can create new content in response to user prompts. This content may include text, computer code, images, audio, video, software documentation, test cases, and design suggestions.

In software development, generative AI can:

- generate code from natural language instructions;
- complete repetitive programming patterns;
- explain unfamiliar code;
- suggest bug fixes;
- write documentation;
- produce unit tests;
- refactor code;

- support software architecture planning.

Generative AI differs from earlier narrow AI systems because it does not only classify or predict. It can produce new outputs that resemble human-created content.

4.2 Technological Unemployment

Technological unemployment refers to unemployment caused when new technologies reduce the demand for human labour faster than new uses for labour are created. Keynes described this as unemployment arising from the discovery of methods that economise labour faster than society finds new ways to employ labour.

In the context of generative AI, technological unemployment would occur if AI tools reduce the need for programmers faster than new AI-related roles, software products, and business demand can absorb displaced workers.

4.3 Structural Unemployment

Structural unemployment occurs when workers' skills no longer match the requirements of available jobs. In the software industry, this may happen if employers demand AI integration, machine learning, cloud infrastructure, cybersecurity, and model evaluation skills while some workers remain trained only for older programming tasks.

4.4 AI-Related Jobs

AI-related jobs are roles created or expanded because of AI technologies. Examples include:

- machine learning engineer;
- AI software engineer;
- prompt engineer;
- AI product manager;
- LLM fine-tuning specialist;
- AI safety researcher;
- model evaluation specialist;
- AI data annotator;
- AI ethics and governance officer;
- AI integration engineer;

- agentic systems architect.

Many of these roles require programming knowledge, meaning that programmers can transition into them through targeted upskilling.

4.5 Software Development Industry

The software development industry includes firms and organisations that design, build, maintain, and deliver software. This includes enterprise software, mobile applications, computer games, databases, operating systems, cloud platforms, embedded systems, cybersecurity tools, and AI models.

5 Literature Review

5.1 Generative AI and Developer Productivity

Research on AI coding assistants suggests that generative AI can significantly improve developer productivity for certain tasks. In a controlled GitHub Copilot study, developers with access to an AI pair programmer completed a coding task faster than those without access. This supports the idea that AI can accelerate software production rather than simply replace developers.

However, productivity gains are not uniform. AI is more effective for repetitive, well-defined, and isolated coding tasks than for complex software architecture, ambiguous product requirements, cybersecurity evaluation, and long-term system maintenance. This means that AI changes the distribution of work rather than eliminating all programming labour.

5.2 Task Automation Versus Job Automation

A major mistake in public discussions is confusing task automation with job automation. A job is made up of many tasks. If AI automates one task, the entire job does not necessarily disappear. For example, a software developer may spend less time writing boilerplate code but more time reviewing AI-generated code, designing systems, testing edge cases, understanding user needs, and ensuring security.

This distinction is central to this research. Generative AI may automate tasks inside software jobs, but it does not automatically automate the whole occupation.

5.3 AI and Labour-Market Restructuring

Technological change often restructures labour markets. Some tasks decline, some tasks expand, and new occupations appear. In software development, AI may reduce demand for purely routine coding but increase demand for workers who can combine programming with AI literacy, product thinking, data understanding, and ethical judgement.

This suggests that the labour-market effect of AI should be analysed as a balance between displacement and creation. A negative outcome occurs when displaced workers cannot transition into new roles. A positive outcome occurs when productivity gains increase demand and workers successfully reskill.

5.4 The Junior Developer Problem

One of the strongest concerns is that generative AI may reduce entry-level programming opportunities. Junior developers often learn by completing basic coding tasks, debugging, writing tests, and documenting code. If AI automates many of these tasks, firms may hire fewer entry-level workers.

However, AI may also lower barriers to entry by helping beginners learn faster, understand code, and produce working prototypes. Therefore, the impact on junior developers is ambiguous. It depends on whether firms use AI to replace entry-level workers or to train them more efficiently.

5.5 New AI-Related Roles

Since 2022, many companies have expanded roles related to AI integration. These roles do not always replace traditional programming; instead, they build on it. For example, AI integration engineers need to understand APIs, software architecture, data security, and model behaviour. AI product managers need to connect business needs with AI capabilities. Model evaluation specialists need to test outputs, measure risk, and improve reliability.

This shows that generative AI creates demand for hybrid skills rather than removing the need for technical workers.

6 Mathematical Framework

6.1 Productivity Gain

Let T_0 represent the time required to complete a software task without generative AI, and let T_1 represent the time required with generative AI. The productivity gain can be measured as:

$$PG = \frac{T_0 - T_1}{T_0} \tag{1}$$

where PG is the productivity gain as a proportion.

For example, if a task takes 100 minutes without AI and 60 minutes with AI:

$$PG = \frac{100 - 60}{100} = 0.40 = 40\% \tag{2}$$

This means that AI reduces the task completion time by 40%.

6.2 Output Expansion

Productivity gains can increase output. If one developer can complete more tasks in the same time, the output expansion factor is:

$$OEF = \frac{T_0}{T_1} \tag{3}$$

If $T_0 = 100$ and $T_1 = 60$, then:

$$OEF = \frac{100}{60} \approx 1.67 \tag{4}$$

This means the developer can complete approximately 1.67 times more work in the same period.

6.3 Task Automation Share

Let A_t represent the number of tasks that can be automated by AI, and let T_t represent the total number of tasks in a software development role. The task automation share is:

$$AS = \frac{A_t}{T_t} \tag{5}$$

A high task automation share does not necessarily mean the whole job is automated. It only means that part of the job changes.

6.4 Job Exposure Index

A job exposure index can be created by combining task automation share with task importance. Let w_i represent the importance weight of task i, and let a_i represent whether task i is automatable, where $a_i = 1$ if the task is automatable and $a_i = 0$ otherwise.

$$JEI = \sum_{i=1}^{n} w_i a_i \tag{6}$$

where:

$$\sum_{i=1}^{n} w_i = 1 \tag{7}$$

If JEI is close to 1, the job is highly exposed to automation. If it is close to 0, the job is weakly exposed.

6.5 Displacement Effect

Let L_0 represent the initial number of workers, and let d represent the displacement rate caused by automation. The number of displaced workers is:

$$D = L_0 \times d \tag{8}$$

For example, if an industry has 1,000,000 workers and 5% are displaced:

$$D = 1,000,000 \times 0.05 = 50,000 \tag{9}$$

6.6 Job Creation Effect

Let J_{AI} represent new AI-related jobs and J_{exp} represent additional jobs created by market expansion. Total job creation is:

$$J_C = J_{AI} + J_{exp} \tag{10}$$

6.7 Net Employment Effect

The net employment effect of AI can be modelled as:

$$NEE = J_C - D \tag{11}$$

Substituting the previous formula:

$$NEE = (J_{AI} + J_{exp}) - (L_0 \times d) \tag{12}$$

If $NEE > 0$, AI creates more jobs than it displaces. If $NEE < 0$, AI causes net job loss. If $NEE = 0$, the labour market is balanced.

6.8 Skills-Adjusted Employment Effect

Not all displaced workers can immediately move into new AI-related jobs. Let s represent the reskilling success rate, or the proportion of displaced workers who can transition into new roles. The skills-adjusted employment effect is:

$$SAE = J_C - D(1 - s) \tag{13}$$

If reskilling is strong, s is high and the negative effect of displacement is reduced. If reskilling is weak, structural unemployment becomes more likely.

6.9 AI-Augmented Labour Demand

AI can increase demand for software by reducing production cost. Let C_0 be the original cost of software production, and C_1 be the cost after AI adoption. The cost reduction rate is:

$$CR = \frac{C_0 - C_1}{C_0} \tag{14}$$

If lower cost increases demand for software products by rate q, then labour demand may be written as:

$$L_1 = L_0(1 - d)(1 + q) + J_{AI} \tag{15}$$

This formula shows that even if some labour is displaced by automation, total labour demand may remain stable or increase if software demand expands and AI-related jobs grow.

7 Methodology

7.1 Research Design

This research uses a mixed theoretical and analytical design. It combines labour-market theory, secondary evidence from industry and employment reports, and mathematical modelling. The purpose is not to calculate the exact number of jobs created or lost globally, since that would require full international employment microdata. Instead, the paper evaluates the direction and logic of the labour-market effect.

7.2 Time Period

The study focuses on the period from 2022 to 2026. This period is important because it begins with the public release of ChatGPT and includes the first major wave of generative AI adoption in software development.

7.3 Data Sources

The analysis uses secondary sources, including:

- labour-market projections from employment agencies;
- developer surveys;
- productivity studies on AI coding assistants;
- reports on AI adoption in software development;
- academic literature on technological unemployment and productivity.

7.4 Analytical Method

The paper analyses three main dimensions:

1. **Automation risk:** which software tasks are most exposed to AI?
2. **Productivity effect:** how does AI change task completion time and output?
3. **Employment effect:** how do displacement and job creation interact?

The mathematical framework is then used to explain why AI can increase productivity without necessarily increasing unemployment.

8 Analysis and Findings

8.1 Tasks Most Exposed to AI

Generative AI is strongest at tasks that are repetitive, text-based, and clearly specified. In software development, these include boilerplate code, simple functions, code explanation, unit test generation, documentation, and debugging suggestions.

Table 1: Software tasks and AI exposure

Task	**AI exposure**	**Reason**
Boilerplate code generation	High	Repetitive and pattern-based.
Documentation writing	High	Text-based and structured.
Simple bug fixing	Medium to high	AI can suggest likely solutions, but human verification is needed.
Unit test generation	Medium to high	AI can create test cases, but coverage quality must be checked.
Software architecture	Medium	AI can assist, but human judgement is essential.
Cybersecurity review	Medium	AI can detect issues, but false confidence is dangerous.
Product requirement analysis	Low to medium	Requires business context and stakeholder understanding.
Team coordination	Low	Depends on communication, trust, and management.
Ethical and legal review	Low to medium	Requires accountability and human judgement.

This table shows that AI mainly automates tasks, not entire jobs. The more complex the task is, the more human judgement remains necessary.

8.2 AI Exposure by Software Development Role

Not all software-related occupations face the same level of AI exposure. Roles that involve repetitive coding, documentation, and basic testing are more exposed to automation, while roles involving system-level judgement, security, management, and human responsibility are less likely to be fully automated. The following table presents an illustrative exposure ranking.

Table 2: Illustrative AI exposure levels across software-related roles

Role	**AI exposure**	**Explanation**
Junior software developer	High	Many entry-level tasks involve code generation, debugging, and documentation.
Software tester / QA analyst	Medium to high	AI can generate tests, but human review and product knowledge remain necessary.
Backend developer	Medium	AI can assist with code, but architecture, APIs, and security require human judgement.
Frontend developer	Medium	AI can generate interface code, but design, usability, and product context remain important.
Machine learning engineer	Medium	AI can assist, but model design, data quality, and evaluation require expertise.
Cybersecurity engineer	Low to medium	AI can detect patterns, but adversarial thinking and accountability remain human-led.
Software architect	Low to medium	Requires long-term system design, trade-off analysis, and business understanding.
AI product manager	Low	Requires strategy, user needs, coordination, and decision-making responsibility.

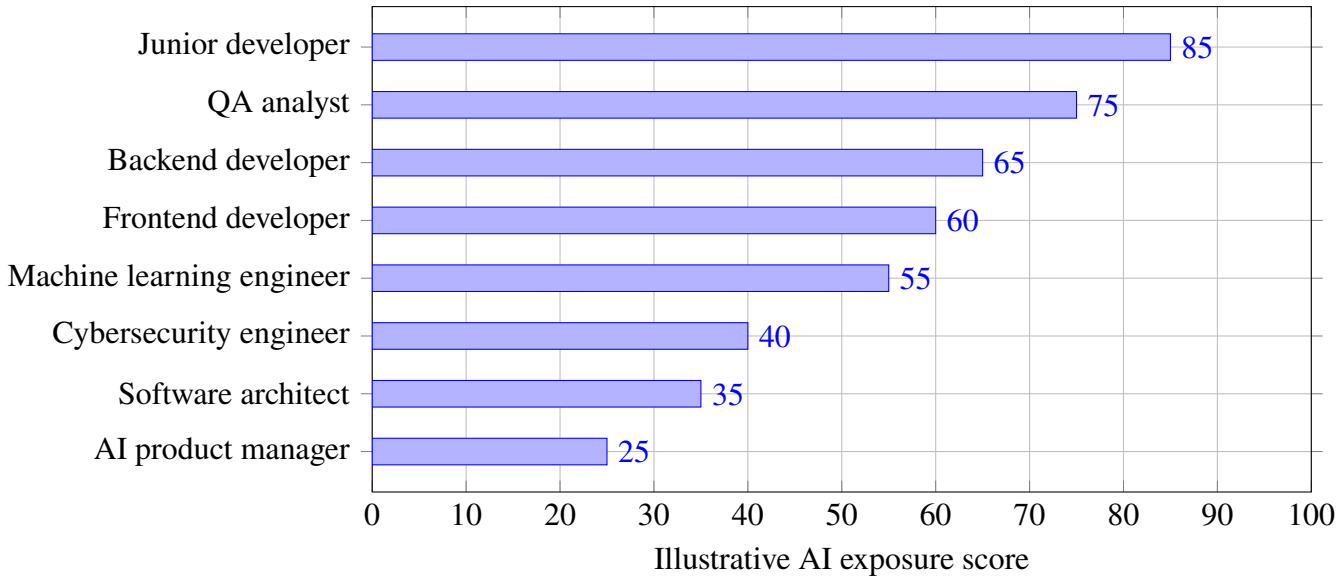


Figure 1: Illustrative AI exposure score across software-related roles.

8.3 New AI-Related Job Categories

Generative AI has created or expanded several job categories. These jobs often require programming skills, but they also require new knowledge about models, prompts, data, evaluation, safety, and integration.

Table 3: AI-related jobs created or expanded after 2022

Job category	**Main function**
AI integration engineer	Connects AI models with business software systems.
Prompt engineer	Designs and tests prompts for reliable AI outputs.
LLM application developer	Builds applications using large language models.
AI safety researcher	Studies risks, misuse, alignment, and model behaviour.
Model evaluation specialist	Tests model accuracy, bias, reliability, and failure cases.
AI product manager	Designs AI products and translates user needs into technical requirements.
Data annotation specialist	Creates and reviews training or evaluation data.
AI governance officer	Develops policies for safe and responsible AI use.
Agentic systems architect	Designs AI systems that can plan, use tools, and complete multi-step tasks.

These roles show that AI changes labour demand rather than simply eliminating it.

8.4 Distribution of New AI-Related Employment Opportunities

Generative AI has not created only one type of new job. Instead, it has expanded a wide range of roles across engineering, evaluation, product development, data work, governance, and safety. The following chart presents an illustrative distribution of AI-related employment opportunities.

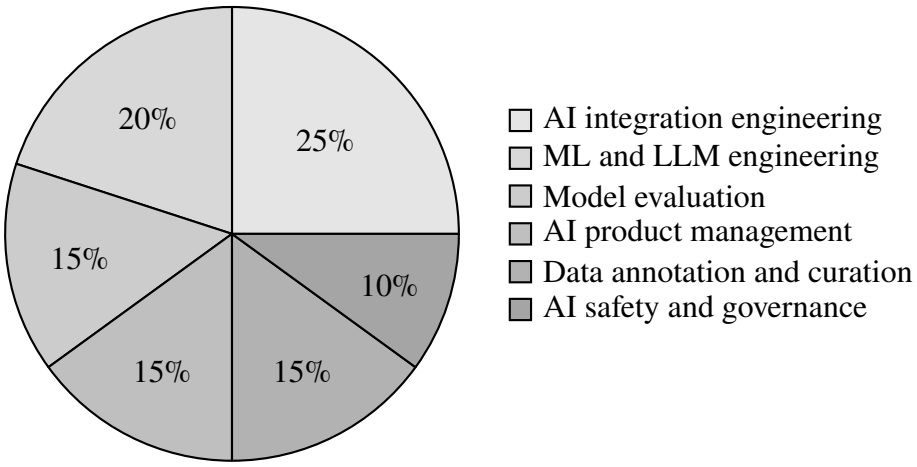


Figure 2: Illustrative distribution of new AI-related job opportunities.

8.5 Productivity and Employment

Productivity growth can have two opposite effects. First, it can reduce the number of workers needed for a fixed amount of work. Second, it can increase total output and create more demand.



Figure 3: Illustrative productivity gain based on controlled AI coding-assistant evidence.

The graph illustrates the basic productivity effect. If developers complete tasks faster, firms may either reduce labour costs or expand production. The final employment outcome depends on which effect is stronger.

8.6 AI Productivity Effect by Task Type

The productivity effect of generative AI is strongest in tasks that are structured and repetitive. It is weaker in tasks that require complex judgement, communication, long-term planning, or responsibility for risk. This supports the argument that AI changes task composition rather than eliminating software work entirely.

Table 4: Illustrative productivity improvement by software task type

Task type	**Estimated productivity improvement**
Boilerplate code generation	60%
Documentation writing	55%
Unit test generation	45%
Simple debugging	40%
Code explanation and learning	35%
Software architecture planning	20%
Cybersecurity review	15%
Stakeholder communication	10%

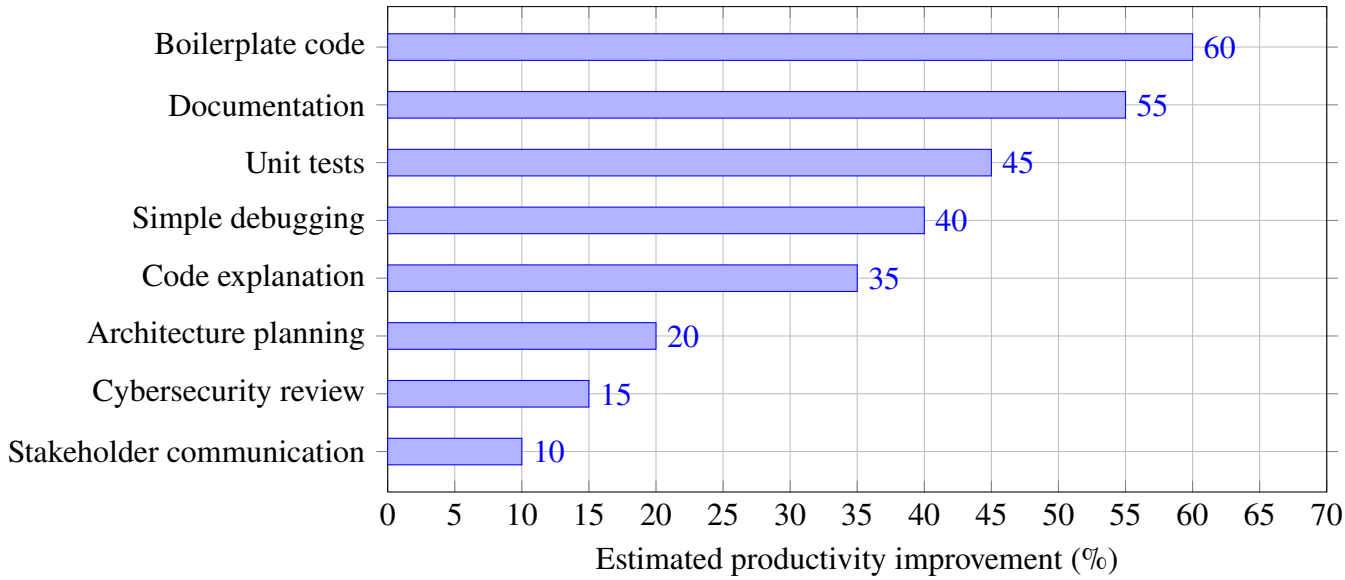


Figure 4: Illustrative productivity improvement by task type after generative AI adoption.

8.7 Net Employment Scenarios

The following table presents three possible scenarios using the net employment formula:

$$NEE = (J_{AI} + J_{exp}) - (L_0 \times d)$$

Table 5: Illustrative net employment scenarios

Scenario	**Displacement**	**Job creation**	**Net effect**
Pessimistic	100,000	50,000	-50,000
Balanced	75,000	75,000	0
Optimistic	50,000	120,000	+70,000

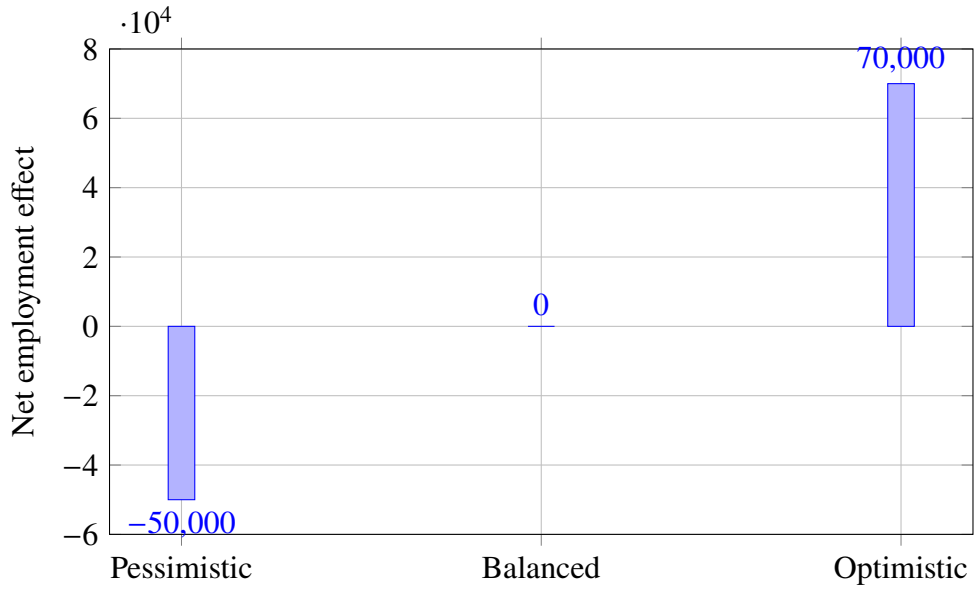


Figure 5: Illustrative net employment effect under three AI adoption scenarios.

This model shows that AI does not automatically create unemployment. The result depends on displacement, job creation, market expansion, and reskilling.

8.8 Skills Needed in an AI-Augmented Software Labour Market

As generative AI becomes more common, the value of software workers depends less on writing every line of code manually and more on their ability to guide, evaluate, and integrate AI-generated outputs. The following table summarises the skills most likely to remain important.

Table 6: Key skills for programmers in an AI-augmented labour market

Skill	**Why it matters**
AI-assisted coding	Developers must know how to use AI tools efficiently without relying on them blindly.
Code review	AI-generated code may contain errors, security risks, or inefficient logic.
Software architecture	Large systems require design decisions beyond simple code generation.
Cybersecurity awareness	AI can create vulnerable code if outputs are not carefully checked.
Prompt design	Better prompts improve the quality of AI-generated solutions.
Model evaluation	AI outputs must be tested for correctness, bias, reliability, and safety.
Data literacy	AI systems depend on data quality, structure, and interpretation.
Communication	Developers still need to understand users, teams, and business requirements.

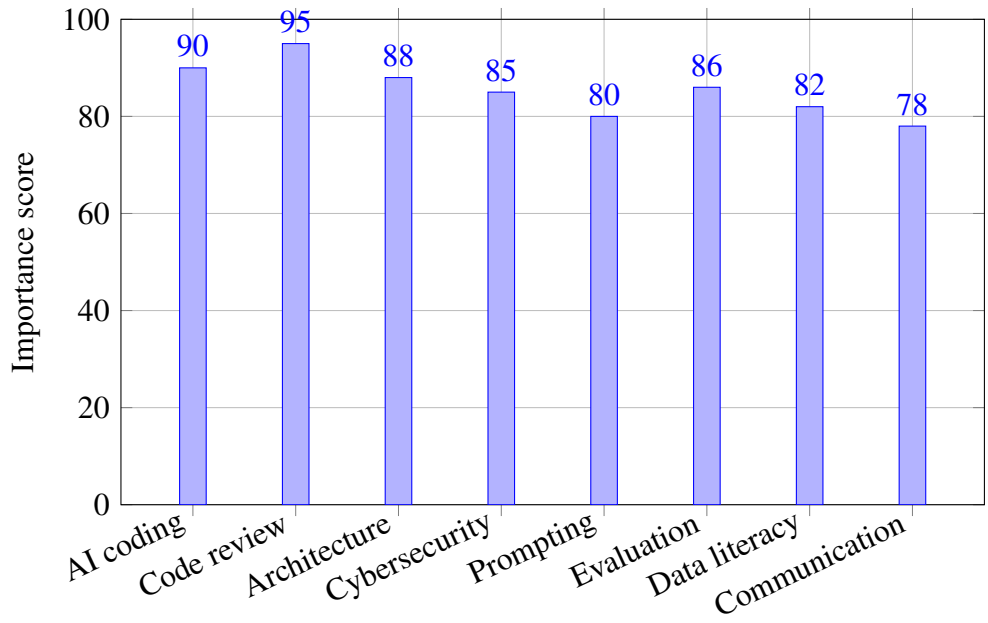


Figure 6: Illustrative importance of skills in an AI-augmented software labour market.

8.9 Why Structural Unemployment Has Not Clearly Dominated

The evidence from 2022 to 2026 suggests restructuring rather than simple collapse of software employment. Several reasons explain this.

First, software demand continues to grow across industries. Companies still need digital infrastructure, mobile applications, cybersecurity, cloud systems, AI products, and automation tools.

Second, AI-generated code requires human supervision. Developers must review outputs, correct errors, ensure security, and integrate code into larger systems.

Third, AI creates new categories of work. Building AI products requires engineers, evaluators, data specialists, governance experts, and product managers.

Fourth, many AI tools make developers more productive rather than replacing them. Firms may use AI to increase output rather than reduce teams.

Fifth, the main labour-market risk is skills mismatch. Workers who adapt to AI tools may become more valuable, while workers who do not reskill may face greater difficulty.

9 Discussion

The analysis suggests that generative AI should be understood as an augmenting technology rather than a purely replacing technology. It can automate parts of programming work, but it also increases the value of developers who can use AI effectively.

The most vulnerable tasks are routine, repetitive, and clearly defined. However, software development is not only code writing. It also involves problem understanding, communication with users, system design, security evaluation, debugging complex interactions, and long-term maintenance. These activities remain difficult to automate fully.

The paper's mathematical model shows that employment outcomes depend on the balance between displacement and creation. If AI displaces workers faster than the industry creates new jobs, unemployment rises. If AI creates new opportunities and workers reskill successfully, employment can remain stable or grow.

This means that the central issue is not whether AI should be used, because it is already being adopted. The key issue is how societies manage the transition. Education systems must teach AI-assisted programming, software architecture, data literacy, cybersecurity, and ethical AI use. Firms must avoid replacing training pipelines with short-term automation. Workers must learn to use AI as a tool rather than compete against it as if it were a human employee.

For junior developers, the risk is serious but manageable. AI may reduce some entry-level tasks, but it can also help beginners learn faster. Companies should redesign junior roles so that new developers learn code review, AI supervision, testing, documentation, and system reasoning.

10 Limitations

This research has several limitations. First, the period from 2022 to 2026 is still early in the history of generative AI adoption. Long-term labour-market effects may differ from early effects.

Second, the paper uses secondary evidence and mathematical modelling rather than original survey data from software firms. A stronger empirical study would collect hiring data, salary data, and worker-transition data directly.

Third, the software industry is global, but labour markets differ by country. The impact of AI may be different in the United States, Europe, India, China, Central Asia, and other regions.

Fourth, the paper does not calculate a precise global number of jobs created or lost. Instead, it explains the mechanisms that determine the net effect.

Fifth, AI technology is changing quickly. Tools available in 2026 are more advanced than those available in 2022, and future agentic AI systems may have stronger labour-market effects.

11 Recommendations

11.1 For Students and Programmers

Students and programmers should not focus only on memorising syntax. They should develop skills that remain valuable in an AI-augmented industry:

- software architecture;
- debugging and testing;
- cybersecurity awareness;
- AI prompt design;
- model evaluation;
- data literacy;
- cloud computing;
- communication and teamwork;
- ethical reasoning.

11.2 For Schools and Universities

Education systems should integrate AI tools into programming education. Instead of banning AI completely, teachers should show students how to use it responsibly. Students should learn how to verify AI-generated code, identify errors, and understand the logic behind solutions.

11.3 For Firms

Companies should use AI to augment software teams rather than eliminate training pathways. Firms still need junior developers because they become future senior engineers. If companies stop hiring beginners, they may create a long-term talent shortage.

11.4 For Policymakers

Policymakers should support reskilling programmes, AI literacy education, and labour-market monitoring. The goal should be to reduce structural unemployment by helping workers transition into new AI-related roles.

12 Conclusion

This research examined whether generative AI has increased technological unemployment or created new opportunities in the software development industry between 2022 and 2026. The analysis suggests that generative AI has not yet produced clear evidence of large-scale structural unemployment among programmers. Instead, it has increased productivity, automated routine tasks, and created new AI-related roles.

The mathematical framework shows that the net employment effect depends on the balance between displacement and job creation. AI can reduce demand for some tasks, but it can also increase demand for software products and create new categories of technical work. Therefore, productivity growth does not automatically mean unemployment.

The most realistic conclusion is that generative AI is restructuring software development. Programmers who only perform routine coding tasks may face greater risk, while programmers who can use AI tools, evaluate outputs, design systems, and integrate AI into products may become more valuable.

The future of software employment will depend on adaptation. If workers, schools, firms, and governments invest in reskilling, generative AI can become a tool for productivity and job transformation rather than a cause of mass unemployment.

References

1. Keynes, J. M. (1930). Economic possibilities for our grandchildren. In *Essays in Persuasion*. Macmillan.

2. Bureau of Labor Statistics. (2025). *Software Developers, Quality Assurance Analysts, and Testers: Occupational Outlook Handbook*. U.S. Department of Labor.

3. Stack Overflow. (2025). *Stack Overflow Developer Survey 2025*. Stack Overflow.

4. Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv preprint arXiv:2302.06590*.

5. McKinsey and Company. (2023). *Unleashing Developer Productivity with Generative AI*. McKinsey Digital.

6. LinkedIn Economic Graph. (2024). *Early Evidence on the Impact of Generative AI on Software Engineers' Employment Outcomes*. LinkedIn.

7. GitHub. (2022). *Research: Quantifying GitHub Copilot's Impact on Developer Productivity and Happiness*. GitHub Blog.

8. Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.

9. Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), 3–30.

10. Acemoglu, D., and Restrepo, P. (2018). Artificial intelligence, automation and work. *NBER Working Paper No. 24196*.

Higher-Altitude Wind Access and the Power Output of Airborne Wind Energy Systems: A Mathematical and Empirical Comparison with Conventional Wind Turbines in China

Sultan Karachalov
Bishkek International School

2026

Abstract

Airborne Wind Energy (AWE) is an emerging renewable energy technology that attempts to harvest stronger and more stable winds at higher altitudes using tethered kites, wings, or aerostat-based turbine platforms. Unlike conventional horizontal-axis wind turbines (HAWTs), which are limited by tower height and rotor size, AWE systems can theoretically access wind resources several hundred or several thousand metres above the ground. Since wind power scales with the cube of wind speed, even a moderate increase in wind speed at higher altitude can produce a significant increase in theoretical wind power density.

This research investigates how higher-altitude wind access affects the theoretical power output and capacity factor of airborne wind energy systems compared with conventional wind turbines in China. The study develops a mathematical model based on wind power density, the Betz limit, the wind-profile power law, the barometric air-density equation, annual energy production, and capacity factor. China is used as a case study because it has both the world's largest conventional wind power fleet and active national programmes in high-altitude wind energy, including the China Energy Engineering Corporation kite-ladder system and the SAWES aerostat system.

The model shows that, under representative North China wind conditions, wind power density at 500 m is approximately 1.88 times greater than at a conventional 100 m hub height, while power density at 1,500 m is approximately 2.65 times greater. However, empirical comparison suggests that this theoretical resource advantage has not yet produced a clear capacity-factor advantage. The announced capacity factors for early AWE systems in China remain close to the national conventional wind fleet average and to the high-altitude Omattingga wind farm in Tibet. The study concludes that AWE has strong theoretical potential, but its practical performance is currently limited by duty-cycle losses, conversion losses, operational restrictions, airspace constraints, and early technological maturity. In the near term,

AWE's strongest advantage in China may be reduced material use, lower land footprint, and access to difficult terrain rather than immediately superior capacity factor.

Contents

1 Introduction **5**

2 Research Problem **6**

3 Research Aim and Objectives **6**

3.1 Research Aim 6

3.2 Research Objectives 6

3.3 Research Questions 7

3.4 Hypothesis 7

4 Background and Literature Review **8**

4.1 Conventional Wind Turbines 8

4.2 Airborne Wind Energy 8

4.2.1 Ground-Generation Systems 8

4.2.2 Airborne-Generation Systems 8

4.3 Why Higher-Altitude Wind Matters 8

4.4 AWE in the Chinese Context 9

5 Mathematical Framework **9**

5.1 Wind Power Available in an Air Stream 9

5.2 Wind Power Density 9

5.3 Rotor Swept Area of a Conventional Turbine 10

5.4 Betz Limit 10

5.5 Wind Shear Power Law 10

5.6 Barometric Air-Density Formula 11

5.7 Wind Power Density as a Function of Height 11

5.8 Altitude Gain Ratio 12

5.9 Capacity Factor 12

5.10 Annual Energy Production 12

5.11 AWE Pumping-Cycle Net Energy 13

6 Methodology **13**

6.1 Research Design 13

6.2 Model Assumptions 13

6.3 Calculation Procedure 14

6.4 Selected Heights 14

6.5 Empirical Comparison 14

7 Results **15**
7.1 Modelled Wind Speed with Height 15
7.2 Power Density Gain 15
7.3 Capacity Factor Comparison 16
7.4 Theoretical Gain Versus Practical Capacity Factor 17

8 Discussion **17**

9 Limitations **18**

10 Future Research **19**

11 Conclusion **19**

References **20**

1 Introduction

Wind energy has become one of the most important sources of renewable electricity in the global energy transition. Conventional wind turbines have expanded rapidly because they can produce electricity without direct carbon emissions during operation. China, in particular, has become the world's largest wind power market, with massive investment in both onshore and offshore wind generation.

However, conventional horizontal-axis wind turbines face a physical limitation. Their rotors are mounted on towers, which normally place the turbine hub approximately 60 to 160 metres above the ground. At this height, the wind is affected by surface friction, terrain roughness, buildings, vegetation, and atmospheric turbulence. Higher in the atmosphere, wind speeds are generally stronger and more consistent. This creates an important question for renewable energy engineering: can wind systems access higher-altitude winds more efficiently than tower-based turbines?

Airborne Wind Energy (AWE) systems attempt to solve this problem. Instead of supporting a turbine on a heavy tower, an AWE system uses a flying device connected to the ground by a tether. This flying device may be a kite, a soft wing, a rigid aircraft, or a helium-filled aerostat. The system can operate at altitudes of several hundred to several thousand metres, where wind speeds may be higher than at normal turbine hub height.

The theoretical attraction of AWE is based on a simple physical principle: wind power depends on the cube of wind speed. If wind speed increases by 25%, the available wind power almost doubles because:

$$1.25^3 \approx 1.95.$$

This means that even a small increase in wind speed with altitude can create a large increase in wind power density. However, theoretical wind resource is not the same as actual electricity generation. Real systems suffer from mechanical losses, electrical losses, downtime, safety restrictions, and weather-related operational limits.

China is a useful case study for this question because it has two relevant characteristics. First, it has a very large conventional wind power sector. Second, it is actively developing high-altitude wind energy systems, including kite-ladder AWE systems and aerostat-based airborne turbine platforms. These projects make it possible to compare theoretical resource gains with currently announced capacity-factor performance.

This paper investigates the following research question:

> How does access to higher-altitude wind speeds affect the theoretical power output and capacity factor of airborne wind energy systems compared with conventional horizontal-axis wind turbines in China?

The study combines mathematical modelling with empirical comparison. The mathematical

model estimates how wind power density changes with altitude. The empirical comparison examines whether announced AWE systems in China show a capacity-factor advantage over conventional wind turbines.

2 Research Problem

The central problem of airborne wind energy is the gap between theoretical potential and practical performance. In theory, higher-altitude winds should provide more power because wind speed increases with height and wind power scales with the cube of wind speed. In practice, however, AWE systems are not yet mature enough to guarantee that this theoretical advantage becomes higher annual electricity output.

A conventional wind turbine is mechanically stable and can generate electricity continuously when wind conditions are within the operating range. By contrast, many AWE systems operate through cycles. For example, a ground-generation kite system produces electricity during the reel-out phase, but consumes or loses energy during reel-in. This creates a duty-cycle loss that does not exist in the same way for conventional turbines.

AWE systems also face special operational challenges. They require airspace management, tether control, launch and landing procedures, weather restrictions, and safety buffers. These factors can reduce availability and therefore reduce annual energy production.

The research problem can therefore be stated as follows:

> Although airborne wind energy systems can theoretically access stronger winds at higher altitudes, it remains unclear whether this resource advantage currently translates into higher practical capacity factors compared with conventional wind turbines in China.

3 Research Aim and Objectives

3.1 Research Aim

The aim of this research is to analyse how access to higher-altitude wind affects the theoretical power output and practical capacity factor of airborne wind energy systems compared with conventional wind turbines in China.

3.2 Research Objectives

The objectives of this research are:

1. To explain the physical relationship between wind speed, altitude, air density, and wind power density.

2. To develop a mathematical model of wind power density at different heights.
3. To compare theoretical wind-resource gains at 300 m, 500 m, and 1,500 m with a conventional 100 m turbine hub height.
4. To compare the capacity factors of conventional wind turbines and announced AWE systems in China.
5. To analyse why theoretical power-density gains may not translate directly into practical capacity-factor gains.
6. To evaluate the potential advantages and limitations of AWE in China's renewable energy strategy.

3.3 Research Questions

This research is guided by the following questions:

1. How does wind speed change with height under the wind-profile power law?
2. How does wind power density change when both wind speed and air density are considered?
3. What is the theoretical wind power density gain at 500 m and 1,500 m compared with 100 m?
4. How do the implied capacity factors of Chinese AWE projects compare with conventional wind turbines?
5. Why might AWE's theoretical altitude advantage not yet appear in real capacity-factor performance?

3.4 Hypothesis

The main hypothesis of this study is:

> Airborne wind energy systems have a higher theoretical wind-resource potential than conventional wind turbines because they can access stronger winds at higher altitudes; however, current AWE systems in China do not yet demonstrate a clear practical capacity-factor advantage because of technological, operational, and conversion losses.

4 Background and Literature Review

4.1 Conventional Wind Turbines

Conventional horizontal-axis wind turbines are the dominant technology in modern wind power. A turbine captures kinetic energy from moving air using rotor blades. The rotor turns a generator, converting mechanical energy into electricity.

The main advantage of conventional wind turbines is technological maturity. They have been commercially deployed for decades, and their design, maintenance, grid integration, and performance are well understood. However, their main limitation is tower height. Building taller towers increases cost, material use, transportation difficulty, and structural stress.

4.2 Airborne Wind Energy

Airborne Wind Energy systems replace part of the tower-based structure with a flying device. Two main types of AWE systems exist.

4.2.1 Ground-Generation Systems

In a ground-generation system, the flying device pulls on a tether connected to a ground-based generator. During the reel-out phase, the tether tension drives the generator and produces electricity. During the reel-in phase, the system retrieves the tether, usually consuming some energy. The net energy output is the difference between energy generated during reel-out and energy consumed during reel-in.

4.2.2 Airborne-Generation Systems

In an airborne-generation system, small turbines are mounted directly on the flying platform. Electricity is generated in the air and transmitted to the ground through a conducting tether. This avoids the reel-in/reel-out pumping cycle, but it adds weight and complexity to the flying system.

China is developing both approaches. The CEEC kite-ladder system represents the ground-generation approach, while the SAWES aerostat programme represents the airborne-generation approach.

4.3 Why Higher-Altitude Wind Matters

Wind speed generally increases with height because surface friction slows air near the ground. At higher altitudes, the air is less affected by terrain and obstacles. This creates stronger and more stable wind resources.

The importance of altitude is amplified by the cubic relationship between wind speed and power. If the wind speed doubles, theoretical wind power increases by a factor of eight:

$$2^3 = 8.$$

This is why AWE has attracted research interest. Even if the wind speed increase is moderate, the power-density increase can be substantial.

4.4 AWE in the Chinese Context

China provides a valuable setting for analysing AWE because it already has a large conventional wind sector and has begun testing high-altitude wind technologies. The uploaded draft discusses several important Chinese cases: the CEEC kite-ladder system in Alxa, Inner Mongolia, the SAWES aerostat system developed by Chinese universities and companies, and the Omattingga conventional wind farm on the Qinghai–Tibet Plateau. These examples allow both theoretical and empirical comparison.

5 Mathematical Framework

5.1 Wind Power Available in an Air Stream

The kinetic power available in wind passing through a cross-sectional area A is given by:

$$P_w = \frac{1}{2}\rho A v^3 \tag{1}$$

where:

- P_w is the available wind power in watts;
- ρ is air density in kg/m^3;
- A is the swept or effective harvesting area in m^2;
- v is wind speed in m/s.

This equation shows that wind power is directly proportional to air density and area, but proportional to the cube of wind speed. Therefore, wind speed is the most sensitive variable in wind power production.

5.2 Wind Power Density

Wind power density is power per unit area. It is obtained by dividing P_w by A:

$$P_d = \frac{P_w}{A} \tag{2}$$

Substituting the wind power equation gives:

$$P_d = \frac{1}{2}\rho v^3 \tag{3}$$

where P_d is measured in W/m^2.

This formula is useful because it allows comparison between wind resources at different altitudes without assuming a specific turbine size.

5.3 Rotor Swept Area of a Conventional Turbine

For a conventional horizontal-axis wind turbine, the swept area of the rotor is:

$$A = \pi R^2 \tag{4}$$

where R is the blade radius. A larger rotor captures more wind power because it sweeps a larger area.

5.4 Betz Limit

No wind turbine can extract all kinetic energy from the wind. If all wind energy were removed, the air would stop moving and no more wind could pass through the rotor. The theoretical maximum extraction efficiency is known as the Betz limit:

$$C_{p,\max} = \frac{16}{27} \approx 0.593 \tag{5}$$

The mechanical power extracted by an ideal turbine is therefore:

$$P_{max} = C_{p,\max} \cdot \frac{1}{2}\rho A v^3 \tag{6}$$

In real systems, the power coefficient C_p is lower than the Betz limit:

$$P_e = \eta C_p \frac{1}{2}\rho A v^3 \tag{7}$$

where η represents mechanical and electrical conversion efficiency.

5.5 Wind Shear Power Law

Wind speed generally increases with height. A common engineering approximation is the wind-profile power law:

$$v(h) = v_{ref} \left(\frac{h}{h_{ref}} \right)^{\alpha} \tag{8}$$

where:

- $v(h)$ is wind speed at height h;
- v_{ref} is wind speed at reference height h_{ref};
- α is the wind shear exponent;
- h is the target height;
- h_{ref} is the reference height.

For open terrain, α is often approximated as:

$$\alpha \approx \frac{1}{7} \approx 0.14 \tag{9}$$

5.6 Barometric Air-Density Formula

Air density decreases with altitude. A simplified exponential model is:

$$\rho(h) = \rho_0 e^{-h/H_p} \tag{10}$$

where:

- $\rho(h)$ is air density at height h;
- $\rho_0 = 1.225$ kg/m^3 is approximate sea-level air density;
- $H_p \approx 8550$ m is the pressure scale height.

This equation is important because higher-altitude winds may be faster, but air density is lower. Since wind power depends on both ρ and v^3, both effects must be considered.

5.7 Wind Power Density as a Function of Height

Combining the wind shear equation and the density equation gives:

$$P_d(h) = \frac{1}{2} \rho_0 e^{-h/H_p} \left[v_{ref} \left(\frac{h}{h_{ref}} \right)^{\alpha} \right]^3 \tag{11}$$

This can be simplified as:

$$P_d(h) = \frac{1}{2} \rho_0 v_{ref}^3 e^{-h/H_p} \left(\frac{h}{h_{ref}} \right)^{3\alpha} \tag{12}$$

This formula is central to the paper because it estimates how theoretical wind power density changes with height.

5.8 Altitude Gain Ratio

To compare wind power density at height h with the baseline at h_{ref}, define the power-density gain ratio:

$$G(h) = \frac{P_d(h)}{P_d(h_{ref})} \tag{13}$$

If $G(h) > 1$, the higher altitude has greater theoretical wind power density than the reference height.

5.9 Capacity Factor

Capacity factor measures how much electricity a system actually produces compared with the maximum possible output if it operated at rated power all year:

$$CF = \frac{E_{annual}}{P_{rated} \times 8760} \tag{14}$$

where:

- CF is capacity factor;
- E_{annual} is annual electricity generation in MWh or GWh;
- P_{rated} is rated capacity in MW or GW;
- 8760 is the number of hours in a non-leap year.

For example, if a 5 MW system produces 10 GWh per year:

$$CF = \frac{10,000 \text{ MWh}}{5 \text{ MW} \times 8760 \text{ h}} \tag{15}$$

$$CF = \frac{10,000}{43,800} \approx 0.228 = 22.8\% \tag{16}$$

5.10 Annual Energy Production

Annual energy production can be calculated from capacity factor:

$$E_{annual} = P_{rated} \times CF \times 8760 \tag{17}$$

This equation is useful for comparing conventional wind farms and AWE systems.

5.11 AWE Pumping-Cycle Net Energy

For a ground-generation AWE system, net cycle energy is:

$$E_{net} = E_{out} - E_{in} \tag{18}$$

where:

- E_{out} is energy generated during reel-out;
- E_{in} is energy consumed during reel-in.

The cycle efficiency can be expressed as:

$$\eta_{cycle} = \frac{E_{net}}{E_{out}} \tag{19}$$

The average power over one complete cycle is:

$$P_{avg} = \frac{E_{net}}{T_{cycle}} \tag{20}$$

where T_{cycle} is the total cycle duration.

This explains why AWE systems may not fully convert higher wind power density into higher capacity factors. Part of the cycle may be non-productive.

6 Methodology

6.1 Research Design

This research uses a mixed theoretical and empirical case-study design. The theoretical part uses mathematical modelling to estimate wind power density at different altitudes. The empirical part compares capacity factors and announced technical data from conventional and airborne wind systems in China.

6.2 Model Assumptions

The model uses the following assumptions:

Table 1: Main assumptions used in the theoretical model

Parameter	**Assumed value**
Reference wind speed, v_{ref}	7.5 m/s
Reference height, h_{ref}	100 m
Wind shear exponent, α	0.14
Sea-level air density, ρ_0	1.225 kg/m^3
Pressure scale height, H_p	8550 m
Reference environment	Representative North China onshore wind profile

6.3 Calculation Procedure

The calculation follows these steps:

1. Use the wind-profile power law to estimate wind speed at different heights.
2. Use the barometric density formula to estimate air density at each height.
3. Use the wind power density formula to calculate P_d at each height.
4. Divide each value by the 100 m baseline to calculate the altitude gain ratio.
5. Compare the theoretical gain with empirical capacity-factor data.

6.4 Selected Heights

The selected heights are 10 m, 100 m, 300 m, 500 m, and 1,500 m. The 100 m height represents a typical conventional wind turbine hub height. The 300 m and 500 m heights are relevant to early AWE testing. The 1,500 m height represents the upper boundary of the model's reliable boundary-layer application.

6.5 Empirical Comparison

The empirical comparison uses the following cases:

1. China's national conventional wind fleet average.
2. Omattingga conventional wind farm in Tibet.
3. CEEC kite-ladder AWE pilot system.
4. CEEC planned commercial kite-ladder system.

5. SAWES aerostat systems as emerging airborne-generation examples.

The purpose is not to claim that these systems are identical. Instead, the goal is to compare whether AWE's theoretical resource advantage is visible in currently announced practical capacity-factor data.

7 Results

7.1 Modelled Wind Speed with Height

Using the wind-profile power law with $v_{ref} = 7.5$ m/s at $h_{ref} = 100$ m and $\alpha = 0.14$, wind speed increases with height.

Table 2: Modelled wind speed and power density at selected heights

Height (m)	**Wind speed (m/s)**	**Air density (kg/m^3)**	**Power density (W/m^2)**
10	5.43	1.224	98
100	7.50	1.211	255
300	8.75	1.183	396
500	9.40	1.155	479
1500	10.96	1.028	676

The table shows that wind speed increases from 7.50 m/s at 100 m to 9.40 m/s at 500 m and 10.96 m/s at 1,500 m. Although air density decreases with height, the increase in wind speed is strong enough to increase wind power density.

7.2 Power Density Gain

Table 3: Wind power density ratio compared with 100 m baseline

Height (m)	**Power density (W/m^2)**	**Ratio to 100 m baseline**
10	98	0.38×
100	255	1.00×
300	396	1.55×
500	479	1.88×
1500	676	2.65×

At 500 m, the theoretical wind power density is approximately 1.88 times the value at 100 m. At 1,500 m, it is approximately 2.65 times the 100 m baseline.

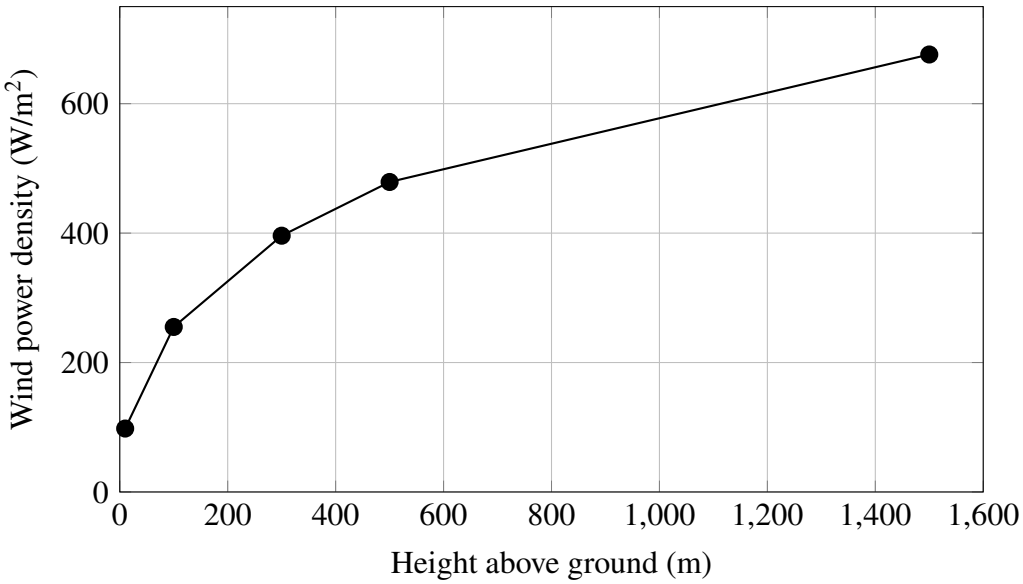


Figure 1: Modelled wind power density as a function of height.

7.3 Capacity Factor Comparison

The next step is to compare theoretical wind-resource gains with practical capacity factors.

Table 4: Reported and announced capacity factors of selected wind systems in China

System	Operating height	Rated capacity	Annual generation	Capacity factor
National HAWT fleet average, China	Around 70–120 m hub height	521 GW	997 TWh	23.7%
Omattingga wind farm, Tibet	High-elevation conventional site	100 MW	200 GWh	22.8%
CEEC kite-ladder AWE pilot	300 m test, 500–3000 m design	5 MW	10 GWh target	22.8%
CEEC commercial AWE concept	500–3000 m design	10 MW	20 GWh target	22.8%
SAWES S1500 aerostat	1000–1500 m	1 MW	Not reported	Not reported
SAWES S2000 aerostat	Around 2000 m	3 MW	Not reported	Not reported

Although AWE systems can access stronger winds, the announced CEEC capacity factors remain close to conventional wind benchmarks.

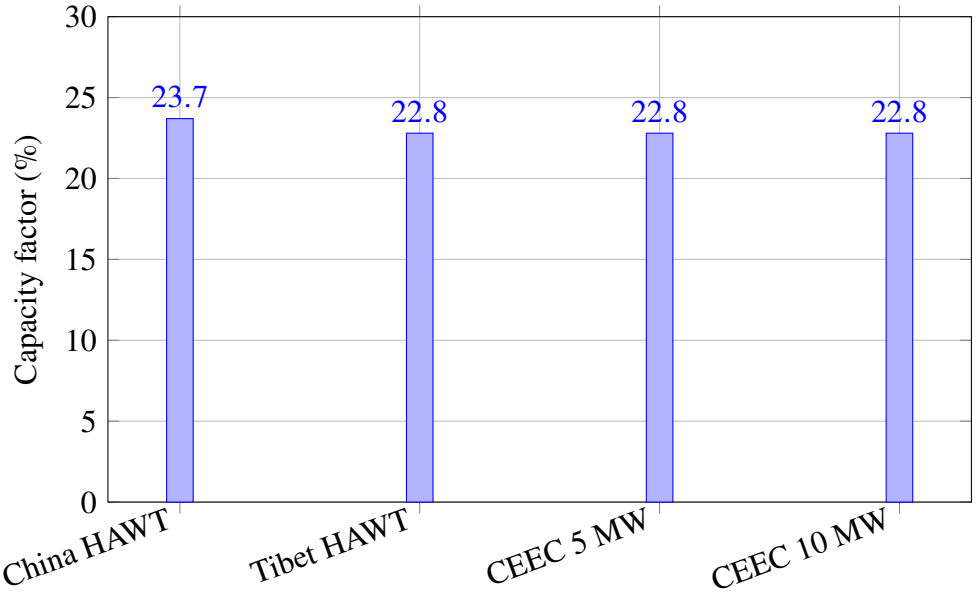


Figure 2: Capacity factor comparison between conventional and announced airborne wind systems in China.

7.4 Theoretical Gain Versus Practical Capacity Factor

Table 5: Comparison of theoretical resource gain and practical capacity-factor evidence

Measure	**Theoretical expectation**	**Observed or announced evidence**
Wind power density at 500 m	1.88× 100 m baseline	No clear capacity-factor advantage yet
Wind power density at 1500 m	2.65× 100 m baseline	SAWES generation data not yet reported
CEEC AWE capacity factor	Expected to benefit from higher wind	Implied around 22.8%
China HAWT national average	Lower-altitude conventional fleet	Around 23.7%

The comparison shows a clear gap. The physical model predicts higher power density at altitude, but current AWE performance targets do not yet exceed conventional wind capacity factors.

8 Discussion

The results show that higher-altitude wind access creates a strong theoretical resource advantage. At 500 m, the model estimates nearly double the wind power density available at 100 m. At

1,500 m, the resource is approximately 2.65 times stronger. This supports the basic physics behind airborne wind energy.

However, the empirical comparison does not yet show a practical capacity-factor advantage. The CEEC kite-ladder AWE system has announced generation targets that imply a capacity factor of approximately 22.8%. This is very close to the Omattingga conventional wind farm and slightly below the national conventional wind fleet average.

This gap can be explained by several factors.

First, ground-generation AWE systems operate in cycles. They generate electricity during reel-out but may consume energy during reel-in. Therefore, they cannot convert the entire high-altitude wind resource into net electricity.

Second, AWE systems have additional conversion losses. A conventional turbine converts rotor motion directly into generator output. In AWE, the system may need to convert kite motion into tether tension, tether tension into winch motion, and winch motion into electricity. Each step introduces losses.

Third, operational availability may be lower. AWE systems may need to stop operating during storms, turbulence, lightning risk, airspace restrictions, launch and landing difficulties, or maintenance periods.

Fourth, current Chinese AWE systems are still at demonstration or early commercial stages. Their capacity factors may improve as control systems, materials, and operating strategies become more mature.

Fifth, capacity factor is not the only measure of value. Even if AWE does not immediately exceed conventional wind turbines in capacity factor, it may still provide advantages in other areas. AWE systems may require less steel, less concrete, and less land. They may also be easier to transport to remote locations and may be useful where tower construction is difficult.

Therefore, the most realistic conclusion is that AWE has strong theoretical potential, but its practical advantage has not yet been fully demonstrated in China. Its near-term value may be material efficiency and deployment flexibility rather than superior energy yield.

9 Limitations

This study has several limitations. First, the mathematical model uses a simplified wind-profile power law. This model is suitable for approximate boundary-layer analysis, but it becomes less reliable above approximately 1,500 m. Since some AWE concepts target 3,000 m or higher, more advanced atmospheric data would be required for a complete analysis.

Second, the model uses one representative North China wind profile. Wind conditions vary significantly across China, including Inner Mongolia, Gansu, Xinjiang, Tibet, coastal regions, and offshore areas. A full national analysis would require regional wind datasets.

Third, some AWE capacity factors are implied from announced targets rather than independently verified long-term operational data. This means they should be interpreted cautiously.

Fourth, capacity factor does not capture all benefits of AWE. Land use, steel use, installation cost, maintenance cost, mobility, and environmental impact also matter.

Finally, the analysis does not include detailed economic modelling. A full feasibility study would need levelized cost of energy, capital cost, operating cost, grid integration cost, and maintenance cost.

10 Future Research

Future research should focus on five areas.

First, researchers should use high-resolution atmospheric reanalysis data to model wind resources from 100 m to 10,000 m across different regions of China. This would provide a more accurate estimate of AWE potential.

Second, full-year operational data from CEEC and SAWES systems should be published and analysed. This would allow direct comparison between theoretical potential and real electricity generation.

Third, future studies should calculate the levelized cost of energy for AWE compared with conventional wind. AWE may have lower material use, but its maintenance and control-system costs remain uncertain.

Fourth, environmental and airspace impacts should be studied. AWE systems use large flight envelopes and tethers, which may affect aviation, birds, land use, and safety planning.

Fifth, future research should examine hybrid systems. AWE may not need to replace conventional wind turbines. It could complement them in remote areas, mountainous regions, islands, and sites where conventional tower construction is expensive.

11 Conclusion

This research examined how higher-altitude wind access affects the theoretical power output and practical capacity factor of airborne wind energy systems compared with conventional wind turbines in China. The mathematical model shows that wind power density increases significantly with height. At 500 m, wind power density is approximately 1.88 times the 100 m baseline. At 1,500 m, it is approximately 2.65 times the baseline.

These results confirm the theoretical promise of AWE. Because wind power scales with the cube of wind speed, access to higher-altitude wind can greatly increase the available energy resource. However, the empirical comparison shows that this theoretical advantage has not yet translated into a clear capacity-factor advantage in China. The announced capacity factors for CEEC kite-ladder systems are approximately 22.8%, which is close to conventional wind benchmarks.

The main reason is that AWE systems face practical losses and operational constraints.

Pumping cycles, conversion losses, launch and retrieval requirements, weather restrictions, and early-stage technology maturity reduce the amount of theoretical wind energy that becomes delivered electricity.

The study concludes that airborne wind energy is a promising but not yet fully proven technology. Its strongest near-term role in China may be reducing material use, lowering land footprint, and enabling renewable energy deployment in difficult terrain. To confirm its long-term potential, independently verified full-year generation data are necessary.

References

1. Bechtle, P., Schelbergen, M., Schmehl, R., Zillmann, U., and Watson, S. (2019). Airborne wind energy resource analysis. *Renewable Energy*, 141, 1103–1116.

2. Archer, C. L., and Caldeira, K. (2009). Global assessment of high-altitude wind power. *Energies*, 2(2), 307–319.

3. Burton, T., Jenkins, N., Sharpe, D., and Bossanyi, E. (2011). *Wind Energy Handbook*. John Wiley and Sons.

4. Lellis, M. D., Reginatto, R., Saraiva, R., and Trofino, A. (2018). The Betz limit applied to airborne wind energy. *Renewable Energy*, 127, 32–40.

5. Stull, R. B. (1988). *An Introduction to Boundary Layer Meteorology*. Springer.

6. He, G., and Kammen, D. M. (2014). Where, when and how much wind is available? A provincial-scale wind resource assessment for China. *Energy Policy*, 74, 116–122.

7. Davidson, M. R., Zhang, D., Xiong, W., Zhang, X., and Karplus, V. J. (2016). Modelling the potential for wind energy integration on China's coal-heavy electricity grid. *Nature Energy*, 1, 16086.

8. China Energy Engineering Corporation. (2025). High-altitude wind energy kite-ladder system technical announcements.

9. Tsinghua University, Peking University, and Beijing SAWES Energy Technology Co. (2026). Stratospheric Airborne Wind Energy System project reports.

10. Global Energy Monitor. (2025). China's solar and onshore wind capacity reaches new heights.

Gender, Social Comparison, and Clothing Expenditure Among High School Students

A Behavioural Economics Perspective

Mikhail Razygraev

Abstract

Clothing occupies an unusual place in a teenager's budget: it satisfies a basic need but rarely stops there. For many high school students, what they wear is bound up with confidence, belonging, and standing among peers, which raises the question of whether their spending on clothes can be explained by price and income alone, or whether a fuller account requires looking at the social context in which the purchase is made. This proposal sets out a survey-based study of how gender and social comparison shape clothing expenditure among students in grades 9 through 12. Rather than treating "boys versus girls" as the whole story, the study situates spending within a behavioural economics framework that draws on status signalling, herd behaviour, present bias, reference-group influence, and social norms, each of which offers a different lens on why a purchase might be made even when it does not obviously serve a practical need.

Roughly 50 to 100 respondents will be surveyed on monthly spending, brand preference, susceptibility to peer and social-media influence, and the motives behind their purchases, using a combination of Likert-scale items, multiple-choice questions, and a small number of open-ended prompts intended to surface reasoning that fixed response categories might miss. The analysis will compare spending and motivation across gender, examine whether social comparison and social media exposure track with higher or more impulsive spending, and interpret the resulting patterns against the behavioural concepts introduced above rather than against a simple demographic split. Descriptive statistics, cross-tabulations, and correlation analysis will be used as appropriate to the resulting data, with all inferences treated as associative given the self-reported, cross-sectional nature of the design.

The intended contribution is threefold: a clearer empirical account of how adolescent consumption is shaped by social context rather than by need or income alone; a demonstration that gender is better treated as one variable among several rather than as the primary explanatory factor; and a practical basis for financial literacy efforts aimed at helping students recognise the social pressures behind their own spending choices before those habits harden into adulthood.

Keywords: behavioural economics, consumer behaviour, gender, social comparison, clothing expenditure, adolescents, peer influence, status signalling, present bias

1 Introduction

Walk through the corridors of almost any high school and clothing is doing more work than keeping students warm. A hoodie, a pair of sneakers, or a particular brand of jacket can signal who someone is, or who they would like to be seen as. Few students buy an item of clothing purely because it is needed; the decision is usually tangled up with how they expect to be perceived once they wear it, who else is likely to notice, and what wearing it is thought to say about them relative to their classmates.

Conventional economic theory treats consumers as agents who weigh price against benefit and choose accordingly. That picture holds up poorly once emotion, habit, and social pressure enter the frame, which they routinely do for teenagers navigating school culture, friendship groups, and a constant stream of social media content. A student might pay a premium for a branded item when an equivalent alternative is available more cheaply, or buy something unplanned simply because it is on sale, or because a friend just bought the same thing. Explaining choices like these calls for something beyond a standard supply-and-demand account, since the "benefit" being purchased is often social rather than purely functional.

Behavioural economics offers that alternative. It does not assume people are irrational so much as it recognises that rationality operates under social and psychological constraints, most obviously seen when a purchase is driven less by need than by the desire to keep up with, or stand out from, one's peers. Concepts such as reference-group influence, herd behaviour, and present bias give researchers a vocabulary for describing decisions that a purely price-and-income model would treat as anomalies.

Gender adds a further layer worth examining, though not one to be assumed in advance. Boys and girls may face different expectations about appearance and style, and marketing itself is often segmented along gender lines, but the direction and strength of any resulting difference in spending or motivation are empirical questions rather than foregone conclusions. This proposal therefore treats gender as one variable among several, to be examined alongside peer influence, social comparison, and exposure to online trends, rather than as the sole explanation for spending patterns. The remainder of this proposal develops that framework, describes the survey instrument through which it will be tested, and sets out the analysis plan, expected findings, and limitations of the design.

2 Statement of the Problem

Discussions of teenage clothing spending tend to collapse quickly into a single comparison: do boys or girls spend more? That question, on its own, explains very little. It says nothing about why a given student spends what they spend, whether the purchase was planned, or what social forces were operating at the time of the decision. A behavioural economics lens is useful precisely because it asks about mechanism rather than magnitude: is a purchase driven by comparison with classmates, by an urge triggered through social media, by pressure from a specific reference group, or by a wish to project a particular image? Answering "who spends more" without asking "why" risks mistaking a symptom for an explanation.

There is also a more practical concern. Many teenagers are only beginning to manage money on their own, and clothing is often one of the first categories where they exercise real discretion over a limited budget. If social comparison and marketing are shown to have measurable pull on their decisions, that finding has direct relevance to financial literacy efforts aimed at this age group, since programmes that only teach budgeting mechanics without addressing the social drivers of spending are likely to have limited effect. Understanding the behavioural mechanism is a precondition for designing an intervention that actually changes behaviour.

3 Aim of the Study

The study aims to examine how gender and social comparison influence clothing expenditure among high school students, with particular attention to the behavioural mechanisms, such as status signalling, herd behaviour, and present bias, that plausibly connect social context to observed spending.

4 Research Objectives

1. Identify the principal factors influencing high school students' clothing purchases.
2. Compare clothing spending patterns and purchase frequency between male and female students.
3. Assess the role of peer influence and social media in shaping clothing decisions.
4. Determine whether purchases are driven mainly by need, comfort, confidence, status, trend-following, or the wish for social acceptance.
5. Examine the extent to which impulsive buying is associated with social media exposure.
6. Interpret the resulting patterns through established behavioural economics concepts, including status signalling, herd behaviour, present bias, and reference-group influence.

5 Research Questions

RQ1. How does gender relate to clothing spending among high school students?

RQ2. How does social comparison affect students' willingness to spend on clothes?

RQ3. Do social media exposure and peer pressure increase the frequency and impulsivity of clothing purchases?

RQ4. Are students more inclined to buy branded or trend-driven clothing to bolster confidence or social image?

RQ5. Which behavioural economics concepts best account for the observed spending patterns?

6 Hypotheses

H1 Students who report stronger social comparison will report higher monthly clothing expenditure.

H2 Greater social media exposure will be associated with a higher incidence of impulsive clothing purchases.

H3 Gender will be associated with differences in the stated motivation for purchasing clothes, but social influence is expected to matter for both male and female students, rather than one group alone.

H4 Students who place greater importance on branded clothing will report a stronger tendency to compare themselves with peers on the basis of appearance.

7 Literature Review

7.1 Behavioural Economics and Consumer Choice

The core claim of behavioural economics, developed through the work of Kahneman (2011), Thaler (2015), and Ariely (2008) among others, is that decision-making departs systematically from the predictions of standard rational-choice models. Emotion, habit, and social context do not merely add noise to otherwise rational decisions; they shape the decisions themselves. Ariely (2008) in particular argues that these departures from rationality are predictable rather than random, which is what makes them useful for research rather than dismissible as measurement error. Applied to consumption, this means a product's practical value is often secondary to what owning it communicates or how it makes the buyer feel. Adolescents, whose financial habits are still forming, are a natural population in which to look for this pattern, since their choices are so visibly filtered through school culture and peer expectation.

7.2 Social Comparison and Reference Groups

Festinger (1954) argued that people evaluate themselves largely by comparison with others, particularly when objective standards are unavailable. Bearden & Etzel (1982) extended this idea specifically to consumption, showing that reference groups, the people whose opinions and choices a consumer takes as a benchmark, exert measurable influence on product and brand decisions. A school is close to an ideal setting for this process: the same peer group is encountered daily, and clothing is one of the most visible things available for comparison. The resulting pressure need not be explicit. A student rarely needs to be told directly that their clothes are inadequate; a

quieter sense of falling short of the group, reinforced through everyday exposure to Cialdini (2001)'s social-proof heuristic, that people take cues about correct behaviour from what others around them are doing, can be enough to prompt a purchase.

7.3 Gender and Clothing Consumption

Gender differences in appearance-related pressure have been documented in the broader consumer behaviour literature (Solomon, 2020), with girls often reported to face stronger expectations around fashion and boys more often oriented toward brand and status markers such as footwear. Work by Elliott & Leonard (2004) on branded footwear among adolescents found that peer pressure and the fear of exclusion, rather than straightforward taste, were often the proximate cause of specific brand purchases regardless of the buyer's gender. Piacentini & Mailer (2004) similarly found that teenagers use clothing to manage a fragile, still-forming sense of identity, a process that did not map neatly onto a simple male/female divide. These patterns are neither fixed nor universal, and social media appears to be narrowing some of the historical divides in fashion exposure by showing both groups similar trend cycles. For that reason, gender is treated here as one factor to be considered jointly with peer influence and online exposure, rather than as an independent explanation on its own.

7.4 Social Media and Impulsive Buying

Platforms built around short-form video and constant visual content, such as TikTok and Instagram, expose students to an unusually high volume of fashion-related material in a short span of time. This steady stream of influencer posts, advertisements, and limited-time promotions creates fertile ground for impulsive buying. In behavioural terms, this connects to present bias, the tendency described formally by O'Donoghue & Rabin (1999) to weight immediate reward over longer-term financial considerations, and to the broader account of self-control failure offered by Vohs & Faber (2007), who showed that decision fatigue and depleted self-regulatory resources make impulsive purchases more likely later in a shopping session or a stressful day. A student may be well aware that a purchase is not a wise use of limited funds and buy the item anyway, because the satisfaction of the purchase is felt immediately while the cost is deferred.

7.5 Clothing as Status Signalling and Materialism

Clothing can also function as a signal of status or group identity, a concept with roots in Veblen (1899) and developed further in modern marketing theory (Kotler & Keller, 2016) and in Bourdieu's account of taste as a marker of social position (Bourdieu, 1984). Richins (1994) distinguished between the public and private meanings that possessions carry, arguing that an item's social value, what it signals to others, can be as important to the buyer as its private, functional value. Chaplin & John (2007) traced how materialistic tendencies develop across childhood and adolescence, finding that self-esteem and the desire for peer approval were closely linked to how much a young person's sense of self-worth was tied up in possessions. Not every purchase is motivated this way; comfort and personal preference remain relevant. But because adolescence is a period in which social acceptance carries particular weight, status signalling is likely to explain a meaningful share of clothing purchases among this age group, especially where branded or visibly fashionable items are concerned.

8 Methodology

8.1 Research Design

The study will use a quantitative, survey-based design. A survey format is well suited to this research question because it allows responses to be collected from a reasonably large number of students and compared systematically across gender and levels of social influence. The survey will be administered anonymously, which should encourage more candid answers given that questions about money and peer pressure can feel sensitive to disclose. A small number of open-ended items will be included alongside the closed-format questions to capture reasoning that a fixed list of options might not anticipate.

8.2 Participants

Participants will be drawn from students in grades 9 through 12, with recruitment designed to achieve a reasonably even split between male and female respondents so that gender-based comparisons are meaningful. A target of 50 to 100 completed responses is set as a minimum for the analysis to be statistically informative, though a larger sample would strengthen the conclusions considerably. Table 1 sets out the intended composition of the sample by grade level, and Figure 1 illustrates the intended gender balance of respondents.

Table 1: Intended sample composition by grade level

Grade level	Target respondents	Approx. share of sample
Grade 9	15–20	20%
Grade 10	15–20	20%
Grade 11	20–25	27%
Grade 12	20–25	27%
Unspecified / other	5	6%
Total	**75–95**	**100%**

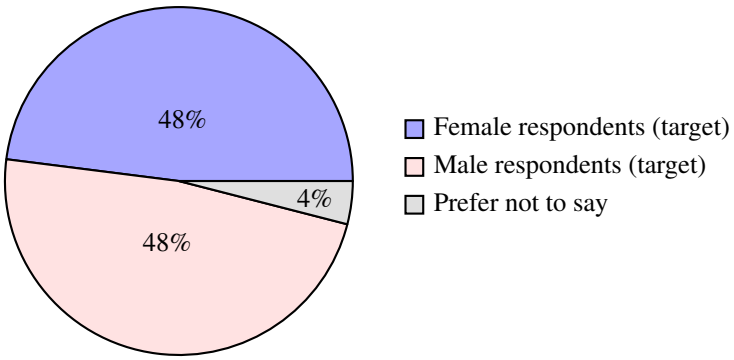


Figure 1: Intended gender composition of survey respondents. Figures are recruitment targets rather than observed data, since the survey has not yet been administered.

8.3 Data Collection

Data will be gathered through an online questionnaire (with a paper-based option available where internet access is limited), combining multiple-choice items, five-point Likert-scale items, and a small number of open-ended

questions. The questionnaire will cover:

- monthly clothing expenditure;
- frequency of clothing purchases;
- brand preference and its perceived importance;
- the influence of friends and classmates;
- the influence of social media and online advertising;
- impulsive buying tendencies;
- stated motivation for purchases (need, comfort, confidence, trend, brand, or social acceptance);
- perceived link between clothing and confidence at school.

Table 2 sets out how these topics map onto the structure of the instrument, including the approximate number of items per section and the response format used.

Table 2: Structure of the survey instrument

Section	**No. of items**	**Response format**
Demographics (gender, grade)	2	Multiple choice
Spending behaviour	4	Multiple choice / numeric
Social comparison	4	5-point Likert scale
Peer influence	3	5-point Likert scale
Social media influence	4	5-point Likert scale
Brand importance	3	5-point Likert scale
Impulsive buying	3	5-point Likert scale
Purchase motivation	1	Ranked multiple choice
Open-ended reflection	2	Short answer
Total	**26**	—

Table 3 gives a sample of the Likert-scale items that will be used to measure social comparison, one of the central constructs in the study, along with the anchor labels attached to each end of the scale.

Table 3: Sample Likert-scale items measuring social comparison

Item	**Scale anchors**
I compare my clothes with what my classmates wear.	1 = Never, 5 = Very often
I feel less confident if my clothes look outdated compared to my friends'.	1 = Strongly disagree, 5 = Strongly agree
I check what people are wearing online before deciding what to buy.	1 = Never, 5 = Very often
I feel pressure to dress similarly to my friend group.	1 = Strongly disagree, 5 = Strongly agree

8.4 Key Variables

8.5 Conceptual Framework

Figure 2 sketches the relationships the study is designed to test. Social comparison and social media exposure are treated as the primary behavioural drivers, feeding into intermediate mechanisms such as status signalling and

Table 4: Key variables measured in the survey instrument

Variable	Description
Gender	Self-reported gender identity, or preferred response option
Monthly clothing spending	Approximate amount spent on clothing per month
Social comparison	Frequency with which the student compares their clothing with that of peers
Peer influence	Extent to which friends affect clothing choices
Social media influence	Influence of platforms such as TikTok, Instagram, and online advertising
Brand importance	Importance placed on branded clothing relative to unbranded alternatives
Impulsive buying	Tendency to purchase clothing without prior planning
Purchase motivation	Primary driver of purchase: need, comfort, confidence, trend, brand, or social acceptance

present bias, which in turn shape observed clothing expenditure. Gender is modelled as a moderating factor that may alter the strength of these relationships rather than acting as an independent cause on its own.

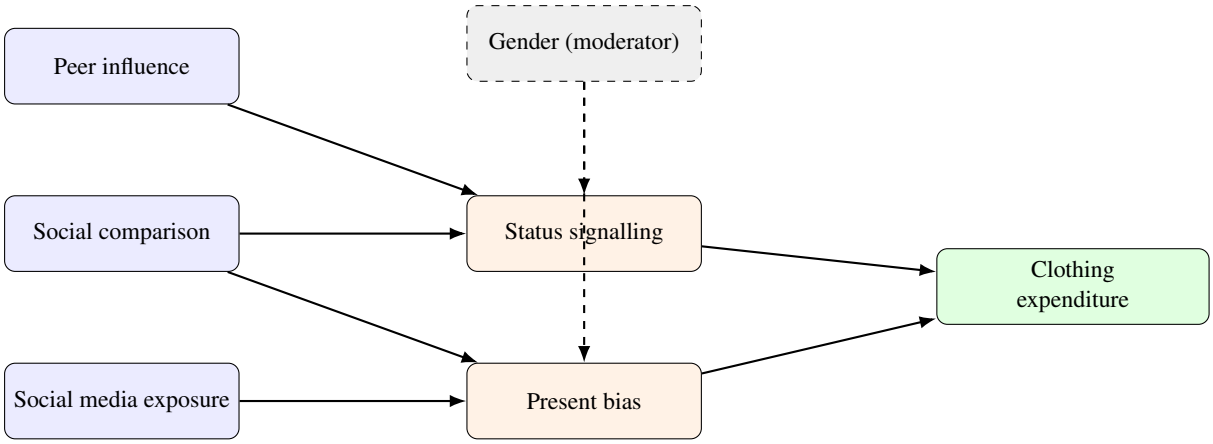


Figure 2: Proposed conceptual framework linking social drivers, behavioural mechanisms, and clothing expenditure. Dashed lines indicate a moderating rather than direct relationship.

8.6 Anticipated Pattern of Purchase Motivation

While no data have yet been collected, it is useful to set out, in advance, the pattern the literature would lead one to expect, so that the eventual results can be compared against a stated baseline rather than interpreted after the fact. Figure 3 illustrates this anticipated pattern for the primary purchase-motivation item, broken down by gender; the figures shown are illustrative projections drawn from the literature reviewed in Section 7, not observed survey results.

8.7 Data Analysis Plan

Descriptive statistics will be used to summarise spending by gender, together with cross-tabulations of purchase motivation against gender and grade level. Social comparison scores will be examined against reported spending to assess the relationship proposed in H1, and correlation analysis (Pearson's r or Spearman's ρ, depending on the distribution of the data) will be applied where sample size permits. An independent-samples t-test, or the Mann-

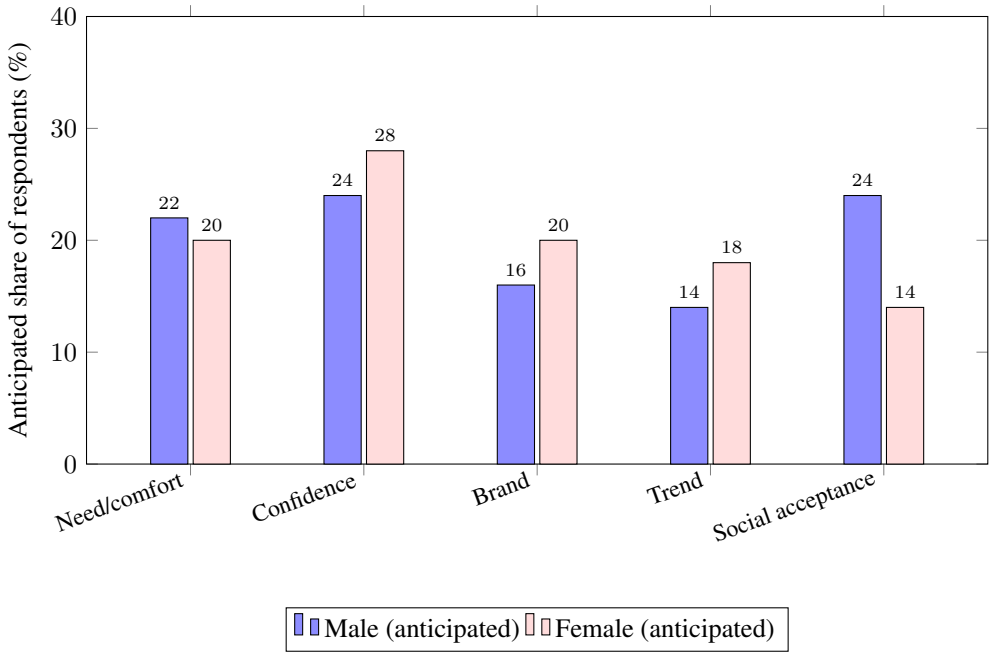


Figure 3: Anticipated distribution of primary purchase motivation, by gender. Values are illustrative projections based on the reviewed literature and are intended as a comparison baseline; they will be replaced with observed survey results once data collection is complete.

Whitney U test where the spending data are not normally distributed, will be used to compare mean spending between male and female respondents for H3, and a similar approach will be used to test H2 and H4. Given the anonymous, self-reported nature of the data, results will be interpreted as associative rather than causal.

8.8 Research Timeline

Figure 4 sets out the intended timeline for the study, from instrument design through to write-up, spread across an eight-week period.

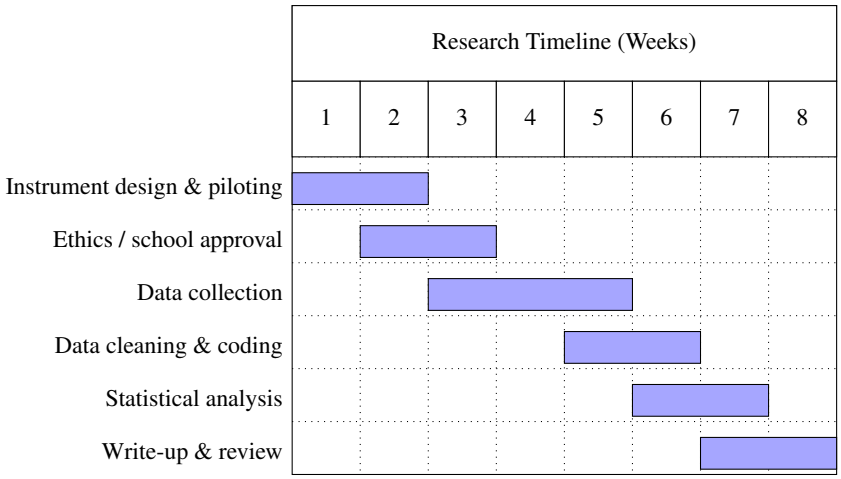


Figure 4: Proposed eight-week research timeline, from instrument design through to final write-up.

9 Expected Findings

The study is expected to find that clothing expenditure among high school students reflects more than practical necessity. Students who report frequent comparison with classmates or online figures are likely to also report stronger interest in trend-driven or branded clothing, and to link clothing more closely to their confidence at school. Social media exposure is expected to correlate with a higher incidence of impulsive purchases, consistent with a present-bias interpretation, and students who score highly on brand importance are expected to also report more frequent social comparison, consistent with H4.

Gender differences may emerge in the stated motivation for purchases, for instance a stronger emphasis on appearance and confidence among some respondents and a stronger emphasis on brand or peer norms among others, broadly in line with the illustrative pattern in Figure 3, but the direction and size of any such difference should be treated as an open question until the data are collected, not a conclusion assumed in advance. It is entirely possible that the observed pattern departs from the projection, in which case the departure itself would be worth discussing in the final report.

10 Discussion

Should the data support H1, that would strengthen the case that adolescent consumption is shaped as much by social environment as by financial constraint: students may spend on clothing not because they need more items, but because doing so manages how they are seen by others. Several behavioural mechanisms would help explain such a pattern. Social comparison would account for the pressure to match peer appearance; herd behaviour and reference-group influence would explain trend-following independent of personal taste; present bias would explain impulsive purchases made despite the buyer's own better judgement; and status signalling, together with the public/private distinction drawn by Richins (1994), would explain why branded items retain their appeal even when functionally equivalent alternatives cost less.

Any gender differences that surface should be read cautiously. The most defensible explanation is unlikely to be a simple, fixed trait of one gender or the other; it is more plausibly a product of differing social expectations, marketing targeting, and peer group norms, all of which are themselves subject to change over time and across contexts, and all of which social media appears to be actively reshaping.

11 Limitations

Several limitations are worth flagging in advance. The sample size, likely in the range of 50 to 100 respondents, may not represent the broader population of high school students, and results drawn from a single school or region will have limited generalisability. Self-reported spending figures are also subject to recall error, since few students track their purchases precisely. Household income is a plausible confound, but direct questions about family finances risk discomfort and non-response, and are therefore excluded from the instrument. Finally, gender is only one axis along which consumption varies; culture, family background, and access to online shopping are all likely to matter as well, and the present design cannot fully disentangle their separate contributions. The illustrative figures presented in Section 7 and Section 8 are projections drawn from the literature rather than observed data, and should not be read as pre-empting the study's actual findings.

12 Conclusion

This proposal sets out to examine how gender and social comparison shape clothing spending among high school students, using behavioural economics rather than simple demographic comparison as the interpretive frame. The underlying premise is that a teenager's clothing purchases are rarely just about clothing: they are entangled with identity, peer acceptance, and the desire to be seen a certain way, and they are sometimes made impulsively despite the buyer's own better judgement. By collecting survey data on spending, comparison, and motivation, the study aims to move the conversation beyond "who spends more" toward a fuller account of why students spend as they do, with the further hope that the findings might help students themselves recognise the social pressures shaping their financial choices.

References

Ariely, D. (2008). *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. HarperCollins.

Bearden, W. O., & Etzel, M. J. (1982). Reference group influence on product and brand purchase decisions. *Journal of Consumer Research*, 9(2), 183–194.

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*. Harvard University Press.

Chaplin, L. N., & John, D. R. (2007). Growing up in a material world: Age differences in materialism in children and adolescents. *Journal of Consumer Research*, 34(4), 480–493.

Cialdini, R. B. (2001). *Influence: Science and Practice* (4th ed.). Allyn & Bacon.

Elliott, R., & Leonard, C. (2004). Peer pressure and poverty: Exploring fashion brands and consumption symbolism among children of the 'British poor'. *Journal of Consumer Behaviour*, 3(4), 347–359.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kotler, P., & Keller, K. L. (2016). *Marketing Management* (15th ed.). Pearson.

O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1), 103–124.

Piacentini, M., & Mailer, G. (2004). Symbolic consumption in teenagers' clothing choices. *Journal of Consumer Behaviour*, 3(3), 251–262.

Richins, M. L. (1994). Valuing things: The public and private meanings of possessions. *Journal of Consumer Research*, 21(3), 504–521.

Solomon, M. R. (2020). *Consumer Behavior: Buying, Having, and Being* (13th ed.). Pearson.

Thaler, R. H. (2015). *Misbehaving: The Making of Behavioral Economics*. W. W. Norton & Company.

Veblen, T. (1899). *The Theory of the Leisure Class*. Macmillan.

Vohs, K. D., & Faber, R. J. (2007). Spent resources: Self-regulatory resource availability affects impulse buying. *Journal of Consumer Research*, 33(4), 537–547.