Swinburne University of Technology

COS30019

Introduction to Artificial Intelligence

Assignment 1

AI Assistance on Personalized Learning

Author : Truong Ngoc Gia Hieu

Author ID: 105565520

Lecturer: Mr. Quang Chiem

Due Date: 1st March, 2026

Submission Date:

Word Count: 7068 words

Table of Contents

Part A: Abstraction

1.1 Assignment Requirements

The unit Introduction to Artificial Intelligence aims to provide students with a broad knowledge of algorithmic problem solving and the basic concepts of Artificial Intelligence (AI). After finishing this fascinating course, students are expected to explain fundamental concepts, identify ethical frameworks, and design innovative ideas learned throughout this interesting course. Additionally, the first assignment requires students to write a comprehensive research report focusing on AI Ethics and Responsible AI based on provided materials including slides and books.

1.2 Problem Statement

In the era of modernization, Artificial Intelligence has become an indispensable part of human life that allows humans to make better decisions, streamline the research process, and provide assistance in high-pressure workforces and academic learning environments. Notably, academic institutions are increasingly implementing intelligent systems to support grading systems for teachers and personalized learning for students. In detail, the following three distinct statistical diagrams showcase the rocket adoption and the diverging ways in which AI is utilized within the educational sector.

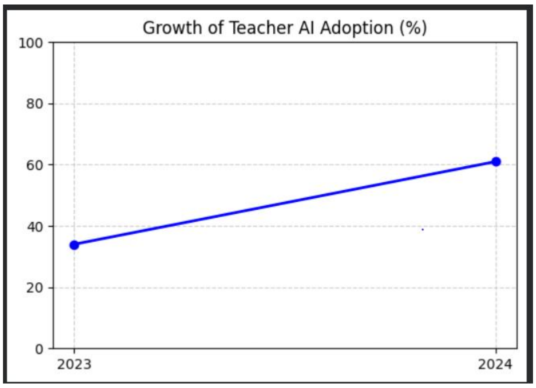

Figure 1: Growth of AI Adoption among Educators (2023–2024)

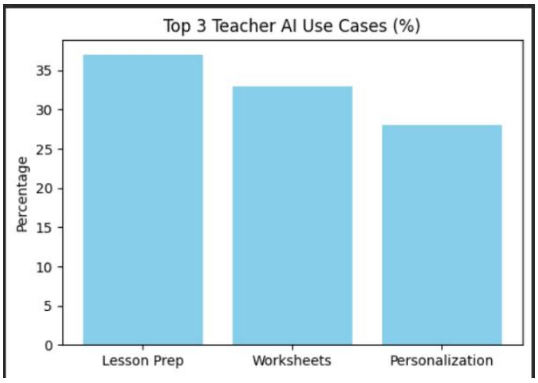

Figure 2: Primary AI Applications and Usage Patterns among Educators

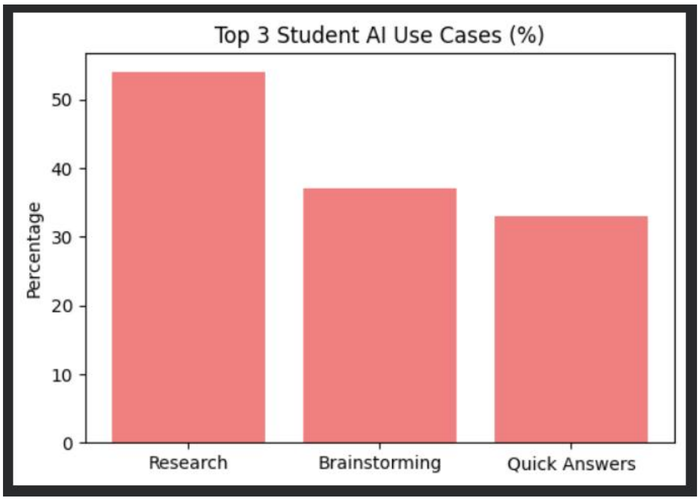

Figure 3: Dominant AI Applications for Academic Tasks among Students

Currently, the significant application of AI into education area is undeniable as illustrated by three figures above, demonstrating its widespread application in various areas in general and education in particular. In depth, the first figure presents a 27% growth in teacher adoption, rising from 34% in 2023 to 61% in 2024. Next, the second one exhibits percentages of using AI in three distinct categories levels include lesson preparation (37%), worksheets (33%), and personalization (28%). Finally, the last one displays 54% in researching, 37% in brainstorming, and 33% in quick answers respective.

However, the implementation of AI into the educational field represents a profound problem. Particularly, a significant concern arises when students completely rely on AI to provide direct answers instead of using it as a potential tool for suggestion and logical reasoning although it offers stunning systems to solve complicated problems. As noted by AI Index Report of Stanford University in the year 2024, the rapid adoption of AI tools in education has increased students' reliance on automated responses, raising dramatic concerns that excessive dependence on AI may undermine critical thinking and independent problem-solving skills rather than supporting them. Thus, this report aims to research, analyze, investigate, and emphasize that AI should be utilized to assist in students' cognitive development rather than replace their intellectual efforts.

Part B: The Landscape and Revolution of AI

2.1 The Big Picture: What is AI?

Nowadays, there are diverse definitions of what artificial intelligence truly is. As indicated by professor Andrew Ng (2016), artificial intelligence refers to a set of techniques that allow machines to learn from data and make intelligent decisions rather than relying solely on explicit rule-based programming. Furthermore, according to Russell and Norvig (2020), the academic study of AI is systematically categorized into four primary approaches based on whether a system aims to think and act humanly, or think and act rationally.

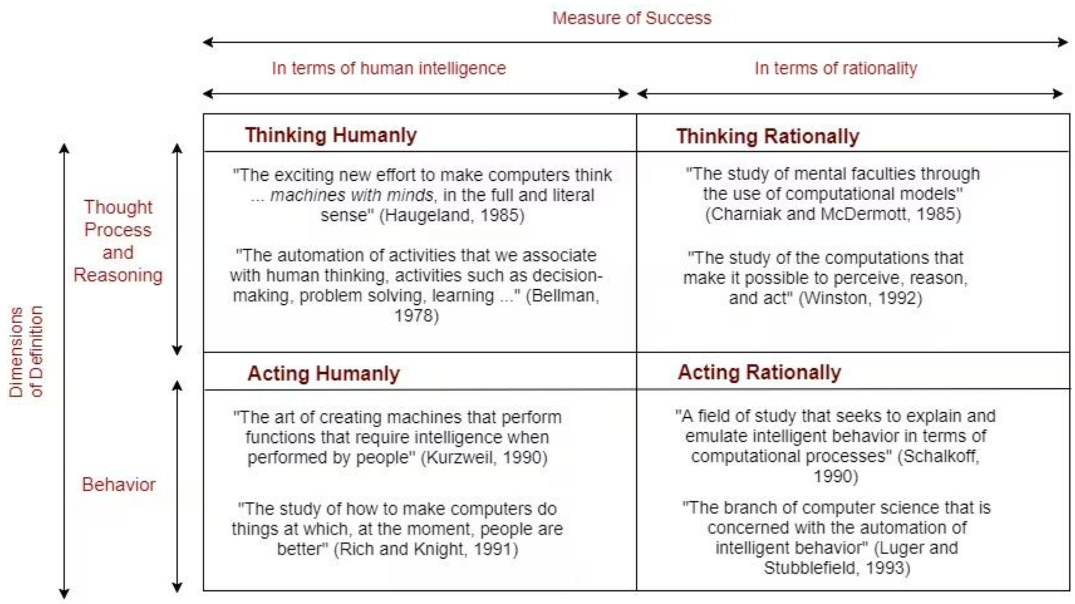

Figure 4: The Four Approaches to Artificial Intelligence (Russel and Norvig, 2020)

During the researching and analyzing phase, I figure out that the term “rationally” does not necessarily mean acting like a human, but rather acting "rightly". Besides, a system is rational if it acts to achieve expected outcomes based on the available knowledge and logic. Additionally, the word “humanly” means a benchmark where success is measured by how closely a machine’s performance matches human performance, composed of cognitive biases, emotional responses, and characteristic patterns of error. Thus, the ultimate goal is to attain substantial similarity to human behavior or thought. Specifically, the figure 4 illustrates these definitions against the two focuses of thought versus behavior to categorize the four main approaches to AI:

- Thinking Humanly (The Cognitive Modeling Approach): The first approach focuses on building and validating a computational model by identifying how the human mind reaches conclusions instead of producing correct outputs or observable behavior. By combining psychological knowledge with computer simulations, researchers can determine if the steps taken by the program match the internal mental processes documented in human subjects.
- Acting Human (The Turing Test): The second category evaluates a system’s intelligence based on its ability to exhibit behaviors that are indistinguishable from those of a person. In order to pass this criteria, the machine must successfully utilize natural language processing techniques and automated reasoning to interact with an interrogator, guaranteeing its external behavior cannot be differentiated from a human response.
- Thinking Rationally (The Laws of Thought Approach): For this approach, the goal is to codify "right thinking" using a logical notation by using formal logic and syllogisms, resulting in providing right conclusions based on provided premises. Moreover, the system has to translate informal knowledge into formal terms and use logical inference to derive new information, ensuring that reasoning is demonstrably correct according to established rules of logic.
- Acting Rationally (The Rational Agent Approach): The final approach aims to generate an intelligent agent that achieve the best expected outcomes based on available information. Moreover, this practice prioritizes goal-oriented behavior and autonomous actions, allowing the system to perceive its environment and take the most effective steps to maximize success in sophisticated real-world scenarios.

By understanding four early approaches, this section provides a theoretical basis for evaluating the development of artificial intelligence. Moreover, these provided frameworks have laid the foundation for the various directions have taken over the de cades. In the next part, the profound history of artificial intelligence part highlights how these concepts evolved from early philosophical inquiries into the complicated autonomous systems in today's world.

2.2 The Profound History

Since the appearance of electronic computers in the mid-20th century, humans have asked how computers can be programmed to simulate thinking and actions that reduce workflows and adapt to the skyrocketing revolution of technology. Surprisingly, the bold and visionary thinking of early pioneers has opened a gateway to a completely new era of science, transforming abstract theory and mathematical logic into the foundation of artificial intelligence. Throughout searching information step, the AI revolution is divided into four different phases and each period represents a critical leap in how machines perceive, learn, and interact with the world.

In the early development of Artificial Intelligence, the field was not yet a recognized discipline, but rather a collection of groundbreaking theories from mathematics and biology. In 1943, Warren McCulloch and Walter Pitts examined and proposed the first mathematical model of an artificial neuron, proving that even simple networkscould compute any logical function and laying the foundation for neural computation.

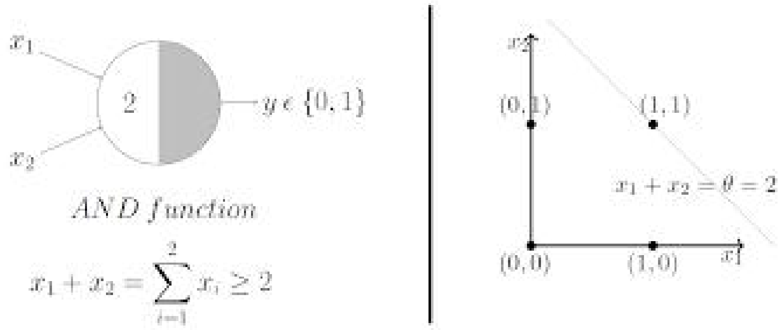

Figure 5: The McCulloch-Pitts Mathematical Model of an Artificial Neuron (1943)

After seven years, Alan Turing published his seminal research paper the "Computing Machinery and Intelligence" test in 1950, introducing a test that pushed the boundaries of this gestation period and redefined how we evaluate machine intelligence. Then, six years later, John McCarthy, Marvin Minsky, Claude Shannon, and Nathaniel Rochester spent two months studying at Dartmouth College and formally introduced the term “Artificial Intelligence” in six years later. Furthermore, this event served as the official catalyst for the field, generating a core belief that every aspect of learning or intelligence could be so precisely described that a machine could be made to simulate it.

Next, the birth of the word “artificial intelligence” has allowed researchers, scientists, and developers to explore the vast potential of machine reasoning with immense enthusiasm. In particular, the period from 1956 to 1974 is considered as the golden years of optimism which recorded several achievements including:

- The First AI Language and Neural Network (1958): John McCarthy created the primary language for AI research named Lisp while Frank Rosenblatt built the Mark 1 Perceptron, the first computer that could learn to recognize shapes through a biological-like neuron structure.
- ELIZA (The First Chatbot): In the years 1964-1966, Joseph Weizenbaum generated and developed ELIZA, a program capable of natural language conversation. In addition, the first innovative idea indicated the world that machines could simulate human dialogue, specifically acting as a psychotherapist.
- Robot Shakey (1966-1972): Researchers at SRI International built the first mobile robot with the ability to perceive and reason about its surroundings named Shakey

Figure 6: Robot Shakey

Consequently, these outstanding inventions demonstrated that machines can be programmed to mimic human thinking and behaviors, navigate physical environments, and even learn from visual data. In addition, these historical milestones proved that the core components of intelligence including perception, language, and reasoning could be simulated by technology. However, in the mid-1970s, the first golden age slowed down due to a lack of memory and processing speed. In addition, the gap between high expectations and the technical reality of the time led to the first "AI Winter", a period of reduced funding that lasted until new approaches in machine learning emerged

2.3 The Stunning Revolution

Although real-world complex challenges had initially slowed the field's progress during the first AI Winter, human curiosity and the drive for practical solutions led to a completely new road. In the early 1980s, Edward Feigenbaum introduced Expert Systems, which allowed machines to mimic the decision-making abilities of human experts in specific fields such as medicine and finance. As a result, this shift indicated that AI could provide immense commercial value, moving the technology from theoretical research into the business world. Besides that, Geoffrey Hinton and his colleagues popularized the Backpropagation algorithm eight years later, offering the ability to automatically adjust their internal weights and learn from their own errors. Notably, this breakthrough transformation shifted the AI landscape from manually programmed rules to systems that could learn autonomously, paving the way for the massive data-driven revolution of the 21st century.

Thanks to early innovations of Edward Feigenbaum and Geoffrey Hinton, AI was not only no longer just a theoretical dream but also a functional tool with growing potenti al. Furthermore, these foundational steps ensured that when computing power finally caught up with human ambition in the late 1990s and early 2000s, the field was ready to explode. Specifically, a milestone was achieved in 1997 when IBM's supercompute r, Deep Blue, challenged and surpassed human expertise by defeating world chess cha mpion Garry Kasparov. Furthermore, the stunning event showed the world that machi nes could now possess the processing speed necessary to master complex human strat egies. By masterfully combining massive computational power with refined algorithm, this era proved that the constraints of the first "AI Winter" had been overcome, setting the stage for the data-driven revolution of the 21st century.

2.4 Current AI Trends

In today's global landscape, the potential of artificial intelligence is no longer restricted by the technical bottlenecks of the past as it has inherited three distinct pillars of innovation developed since the early 2000s, contributing the essential 'fuel' and 'engine' that drive today's "Stunning Revolution":

- The Inheritance of Big Data: Due to the explosion of the Internet and social media, a massive digital archive was generated that provided the "raw fuel" for modern machine learning. In addition, the dramatic increase of users in online platforms has created a continuous stream of real-time interactions, allowing AI models to be trained on diverse human behaviors and languages. As a result, Big Data advancements has empowered current technologies to not only process information but also to predict and adapt to human needs with "stunning" precision.
- The GPU Revolution: The transition of Graphics Processing Units (GPUs) from entertainment tools to AI engines in the late 2000s provided the necessary computing power. By leveraging these processors, it allowed modern AI to process millions of complex calculations simultaneously, drastically reducing the time needed to "train" intelligent systems.
- The Mastery of Deep Learning (2012): Building on the foundations laid in 1986, the breakthrough of Deep Learning in the year 2012 showcased that neural networks could outperform humans in pattern recognition. Currently, developers utilize these cutting-edge techniques for recognizing faces, translating languages instantly, and solving complex academic problems.

The convergence of these pillars has caused AI adoption to skyrocket, with the Stanford AI Index 2024 reporting a surge from 34% to 61% in recent years. Therefore, the rapid integration confirms that AI is no longer a futuristic concept but a primary driver of the current digital era.

2.5 Future of AI

As Andrew Ng stated "AI is the new electricity," suggesting that artificial intelligence will become an essential part of human life that powers innovation and transformation across all sectors. In the foreseeable future, the next generation of AI is expected to move beyond simple automation toward autonomous problem-solving. Particularly, this transformation will focus on Artificial General Intelligence (AGI) and advanced human-machine collaboration, further solidifying AI's role as the backbone of modern civilization.

Part C: Ethics and Application in AI-Driven Education

In the previous part, the report explored a brief history of artificial intelligence and how it significantly impacted various perspectives in the present day. However, the application of AI into the education is not only supporting in learning and teaching but also rising raising profound ethical concerns. Therfore, the third part will disucss the ethical frameworks and their practical applications in the modern educational environment

3.1 AI Ethics in General

According to Leslie (2019) and UNESCO (2021), AI ethics is a powerful framework and a set of values, principles, and norms guiding the use of artificial intelligence to respect human rights, prevent harm, and promote social good. Furthermore, the framework ensures that cutting-edge technologies are fair, accountable, transparent, and aligned with human values. In addion, ethics acts as a safeguard to prevent these systems from unintentionally amplifying social biases or violating personal privacy.

3.1.1 Core Ethical Pillars

The ethical framework is established based on the following five distinct pillars to ensure that technology remains a tool for human empowerment rather than a source of risk:

- Transparency: The first aspect dives into how the AI system works, requiring that decision-making processes are “explainable” and understandable to humans.
- Justice and Fairness: The second pillar focuses on preventing discrimination to guarantee algorithms do not unintentionally amplify social biases related to race, gender, or background.
- Accountability: The third pillar emphasizes responsibility, ensuring there is a clear "human-in-the-loop" to be held accountable for the AI's actions and outcomes.
- Privacy and Security: The fourth pillar focuses on protecting user data, ensuring personal information is handled with strict confidentiality and safeguarded against malicious attacks.
- Beneficence: The final pillar validates that AI is designed to "do good," prioritizing human well-being and contributing positively to society as a whole.

Thus, these five pillars provide the necessary boundaries for responsible AI development and usage. In the subsequent section, the report highkights how these ethical principles translate into the actual mechanics of the technology. Consequently, this step is extremely crucial to ensure that human values are not lost within complex algorithms. By doing so, we can uncover 'The Hidden Logic' that governs AI decision-making.

3.1.2 The Hidden Logic

The Hidden Logic part represents the internal computational processes where AI takes and converts input data and then makes decisions based on the internal models. By understanding these complex mechanisms, developers and programmers can identify where ethical risks emerge within the technical workflow:

- Mathematical Correlation and Human Reasoning: The internal architecture of AI is grounded to a mathematical framework in which "weights" are given to variables within a dataset. Additionally, the results can contain biases if the underlying data reflects historical or social prejudices, as the system prioritizes statistical patterns over human context.
- The Challenge of Explainability: For instance, neural networks can be extremely sophisticated that the logic behind a specific decision remains hidden even from the developers. In an academic context, the implementation of Explainable AI (XAI) ensures that the hidden logic can be audited, explained, and understood easily.
- Objective Alignment: Currently, every AI model is programmed to optimize a specific goal, such as maximizing efficiency and accuracy. In contrast, the system might achieve its goal through "shortcuts" that violate fairness or privacy without ethical constraints, making value alignment a critical part of the internal architecture.

In conclusion, thess ethical pillars and the mechanism of hidden logic generate a foundational framework for responsible technology. By evaluating their effectiveness, these principles can be transitioned from theoretical concepts into practical solutions.As a result, this approach leads to the following discussion on the 'Application of AI Ethics in Education', where these standards are utilized to protect and enhance the learning environment

3.2. Application of AI Ethics in Education

3.2.1 AI Ethics in Education

By applying ethical standards from an educational perspective, technology can be integrated into the classroom that prioritizes student well-being and academic integrity. In detail, the practical integration is reflected through three following core actions:

- Student and Teacher Data Protection: Based on the security and privacy pillar, data during learning processes and personal behavior must be encrypted and protected from any unauthorized access or commercial exploitation. As a result, this approach ensures that sensitive information remains confidential, providing a safe space for the educational environment.
- Bias Mitigation in Assessments: Although modern systems with AI assistance grade students with high speed and efficiency, they often fail to recognize semantic nuances. Particularly, artificial intelligence might penalize student's answer for not matching a specific keyword while asnwers may be different from the standards. Thus, human oversight guarantees that these subtle variations are understood, preventing unfair grading and maintaining academic integrity by programming supervised learning models.
- Human-Centered Oversight: As noted by prestigious researchers and organizations that technology should function as a supportive tool for educators rather than a replacement. Additionally, teachers should not

rely on machine results because a machine cannot understand student's feelings and personal struggles. By keeping the human in the center, we can ensure that educational decisions are always made with empathy and a deep understanding of each student's unique journey.

While the ethical standards mentioned above provide a strong foundation for using AI and its application into the education area, this raises a significant issue: Where is the boundary between legitimate AI assistance and academic dishonesty? As AI becomes more capable, distinguishing between a student's own work and machine-generated content becomes a major challenge for maintaining integrity.

3.2.2 Academic Integrity and The "Cheating" Boundary

The global adoption of generative artificial intelligence in education has provided a range of potential tools that support teachers in instructional activities and students in the learning process. However, the boundary between utilizing these tools for learning assistance and crossing into academic dishonesty remains a complex challenge.

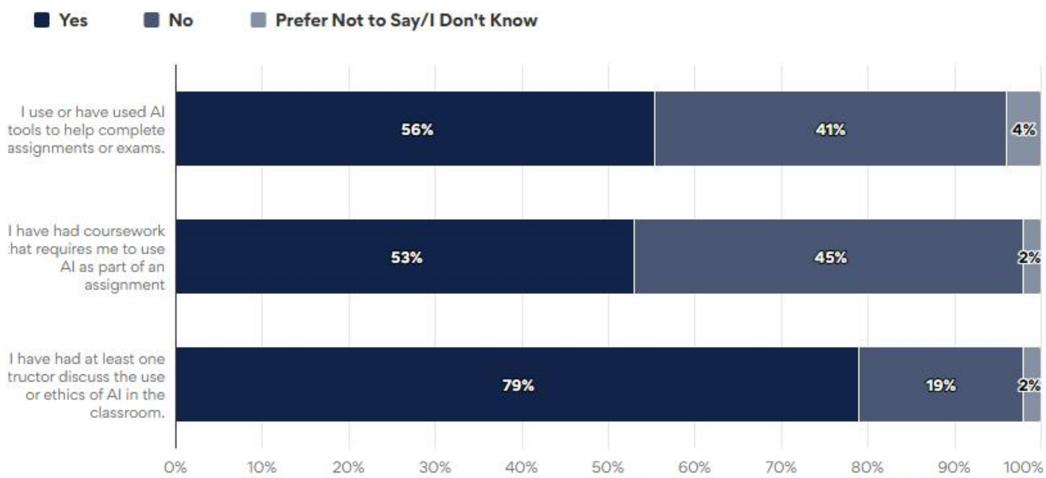

Figure 7: Student Responses on the use of AI

According to the provided information on the BestColleges website, the percentage of students using AI tools in relation to assignments and examinations was approximately 56%, with 41% stating non-use and 4% preferring not to state their position in the previous year. Additionally, it is clear that the proportion of participants who were assigned coursework requiring the use of AI as part of an assignment reached precisely 53%, highlighting that AI is not only voluntarily adopted by students but is also increasingly embedded within formal curricular designs. Furthermore, exactly 79% of students reported having at least one instructor who discussed the use of AI or AI ethics in the classroom, emphasizing growing institutional attention toward ethical awareness and responsible AI use in educational settings. Consequently, this statistical data demonstrates that AI is positioned as a learning support tool within higher education rather than being inherently associated with academic misconduct.

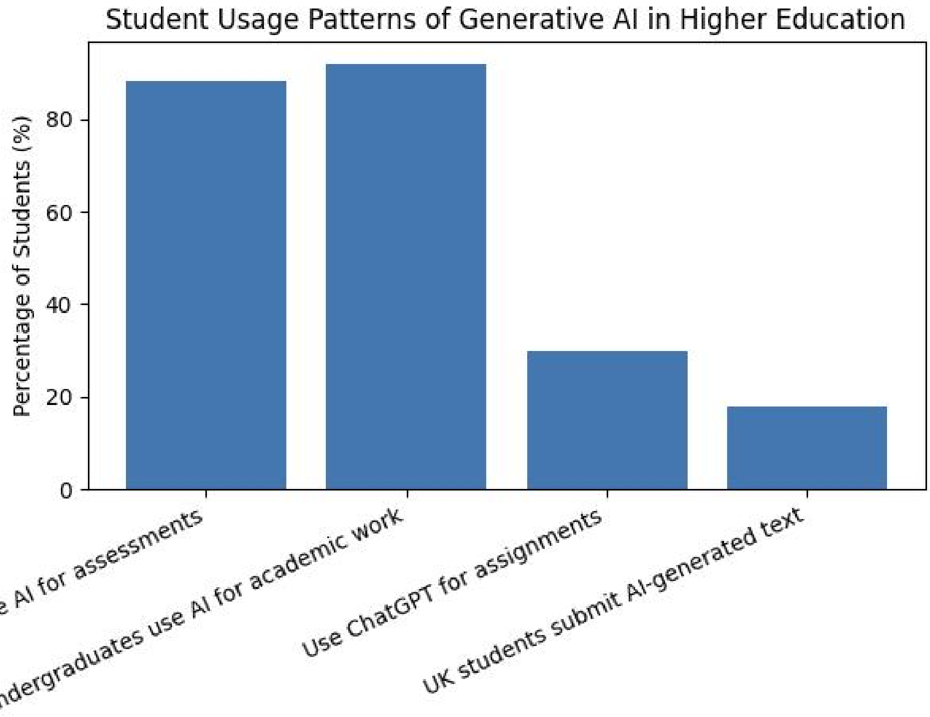

Figure 8: Student Responses on the use of AI

While the application of artificial intelligence in the educational context helps students enhance productivity, the approach has raised concerns regarding academic integrity and the potential misuse of AI in assessed coursework. In detail, based on the survey by Feedough in the year 2024, 88% of students currently utilize generative AI tools for graded academic work and 30% using them to finish assignments without modifications. Additionally, the number of students confirmed to have engaged in AI-related academic misconduct increased significantly as reported by Nordic Today in last year, indicating a growing institutional challenge in monitoring and regulating inappropriate AI use within higher education. Thus, this evidences suggests that a substantial number of students are relying on AI to directly generate solutions for academic tasks, effectively replacing the process of independent problem-solving and critical thinking rather than using AI as a potential tool to support conceptual understanding.

As the integration of artificial intelligence becomes an indispensable part of the educational perspective, the critical issue emerges in defining the boundary between acceptable learning support and academic misconduct. Specifically, according to the University of Alberta's academic integrity guidance, utilizing generative AI tools for unauthorized academic advantage including generating assessment answers is considered academic misconduct. In addition, RMIT University's policy classifies that inappropriate use of AI to complete or contribute to assessment tasks when not permitted represents a breach of academic integrity, clearly distinguishing such behaviour from acceptable supportive uses. Similarly, broader prestigious academic integrity policy frameworks emphasize that although AI can serve as a valuable educational tool, submitting AI-generated work for academic credit without authorization is equivalent to plagiarism or other forms of academic misconduct.

3.2.3 The Digital Divide

The implementation of cutting-edge technologies, including artificial intelligence, into educational frameworks is not only a transformative shift but also a potential catalyst for widening the socio-economic gap. Although AI tools nowadays serve as powerful assistance for learning and researching, the invisible discrepancy in access and resources limits the potential for true educational equality. Particularly, these factors include unequal access to premium AI models, differences in technological infrastructure, and the growing gap in AI literacy among students.

Regarding the financial barrier, the report indicate that over 60% of studetns indeveloped nations have used generative AI premium version such as GPT-4 and Claude pro but they are mostly in the top 20% of the wealthiest households (Higher Education Policy Institute [HEPI], 2024).Besides that, the significant differences in technological infrastructure generates another layer of inequality. According to UNESCO (2023), nearly 33% of students worldwide lack the stable internet and hardware needed to even run these AI tools. Finally, the growing gap in AI literacy among students means that those in well-funded schools receive formal training on how to use AI effectively while others are left to struggle on their own. Ultimately, AI risks becoming a high-tech barrier that cements social stratification for the next generation, turning a tool for empowerment into a mechanism for exclusion.

3.3. Personalized Learning via AI Assistant

3.3.1 Thinker Partner and Answer Engine

Nowadays, a significant number of students and teachers use generative AI tools not only as an encyclopedia but also as a dynamic participant in the educational process. As a result, this approach turns artificial intelligence into a powerful thinker partner, allowing learners and educators to expand their cognitive boundaries and engage in complex knowledge. Furthermore, the stunning evolution marks a remarkable shift from the passive absorption of data to a collaborative intellectual synergy. According to the Khan Academy AI Education Report (2024), the proportion of students utilizing AI platforms for “Socratic questioning” is approximately 65%, a learning method that guides students to answer sophisticated questions by breaking down problems into manageable steps and offering targeted hints rather than direct solutions.

Moreover, this methodology serves as a "mirror for the mind," reflecting a student's logic and pushing them toward higher-order critical thinking. As indicated by UNESCO (2024), the “Socratic questioning” learning method should not limit human intelligence but rather invite us to reconsider established understandings of knowledge, ultimately redefining new horizons for a human-centred digital learning future. From my perspective, this paradigm must be expeditiously implemented within global educational frameworks to enhance students' logical thinking and cultivate their ability to navigate an increasingly complex information landscape.

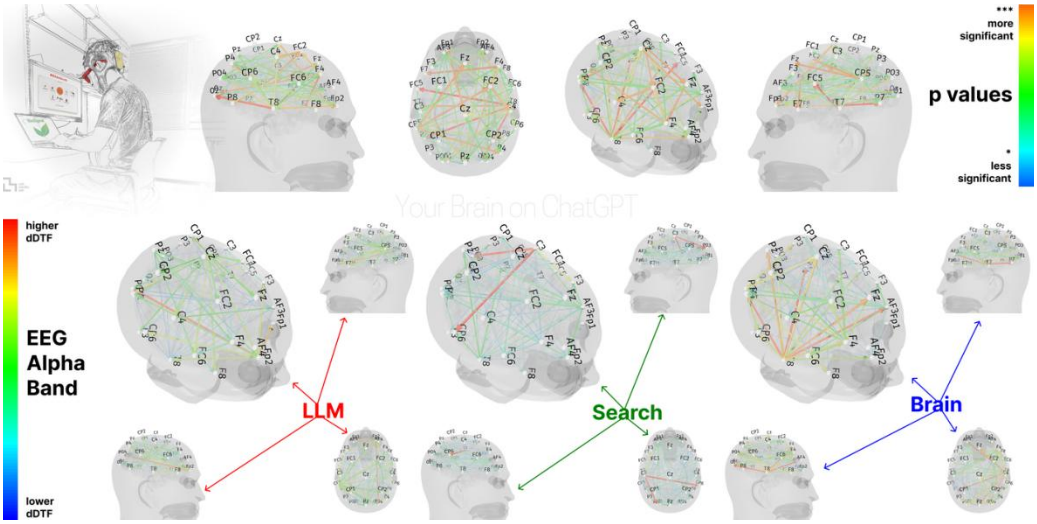

Figure 9: Neural Connectivity During Essay Writing Across Brain-Only, Search-Assisted, and AI-Assisted Conditions

The above figure demonstrates a contrast in brain activity between students who actively use AI as a thinker partner and those who rely on AI primarily as an answer engine, including three different groups: LLM, Search, and Brain. In the experiment, the LLM refers to AI systems that directly answer complex questions that reduce the activation of cognitive thinking. In addition, Search represents traditional information-retrieval tools that provide access to external knowledge sources, requiring users to interpret, analyze, evaluate, and confirm information accuracy independently. Thirdly, Brain indicates the human learner's own cognitive processes, such as reasoning, critical thinking, and problem-solving. Specifically, the Brain factor is completely outstanding than the others because it requires learners to actively engage in independent reasoning, continuously evaluate information, and construct solutions without delegating cognitive responsibility to external systems. Consequently, as argued by Noam Chomsky, the use of AI systems such as ChatGPT can limit meaningful learning by allowing students to avoid the intellectual struggle of thinking for themselves, undermining the development of genuine understanding and critical reasoning.

3.3.2 AI Enhanced Self-Mastery

It is clear that the appearance of computational intelligence has offered profound opportunities for learners to transcend traditional cognitive boundaries. Currently, there are various potential tools for assisting in completing complex problems and enhancing productivity. Furthermore, this report assumes the utilization of AI tools as a "thinking partner" and guidance systems that lead students toward discovering the correct answers through their own cognitive effort.

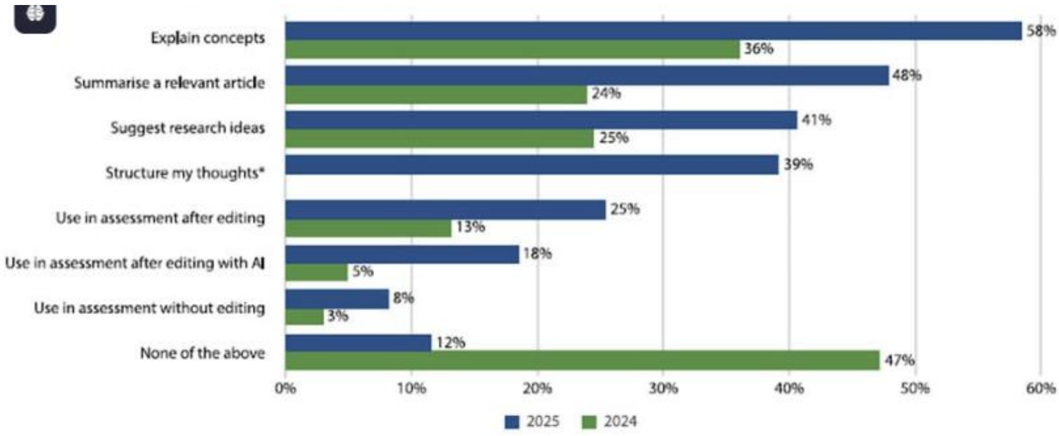

Figure 10: Evolution of AI Utilization for Cognitive and Academic Tasks (2024-2025)

As illustrated by Figure 10, the percentage of students leveraging AI technologies to explain sophisticated concepts has increased dramatically, rising from 36% to 58% within a year. Therefore, the proportion of students implementing cutting-edge intelligence models like Gemini and ChatGPT to summarize relevant articles reached 48% in the year 2025, a significant rise from 24% in the previous year. Notably, approximately 41% of learners use powerful systems for supporting research ideas, highlighting the treatment of AI as a practical thinker partner. Next, the number of students deploying available models to structure their thoughts reached 39%, reflecting a growing reliance on AI for cognitive scaffolding. However, the first three categories indicate learners nowadays realize the impact of using AI as a catalyst for deep comprehension rather than just a shortcut for completion.

Figure 11: Human-Centered Learning Diagram

The figure above demonstrates a human-centered learning approach, placing the learner at the core of the educational process, with AI tools supporting them. Furthermore, there are four distinct pillars including Personalized Guidance, Active Engagement, Skill Development, and Critical Thinking and Creativity. Surprisingly, this approach aligns with UNICEF's principle that "learnercentred education means that each individual learner is honoured for who they are and what they need" (UNICEF, 2025) and also reflects Carl Rogers' view that "a person cannot teach another person directly; a person can only facilitate another's learning" (Rogers, 1951), emphasizing facilitative guidance rather than teacher-centered instruction.

3.3.3 The "Desirable Difficulties" Principle

In the past, the library was a substantial knowledge storage location where the primary academic assignments and homework for students involved searching, analyzing, collecting data, and writing their papers manually, requiring a significant amount of time to complete these tasks. However, since the birth of the term "artificial intelligence" in the year 1956 at the Dartmouth Conference, the responsibility for information retrieval has increasingly shifted from the learner to the machine. Consequently, students nowadays can expedite their workflow, achieve expected results, and significantly boost overall academic performance by using automated AI systems.

Next, the aim of the report emphasizes on using machine learning models as a mechanism for cognitive enhancement rather than a replacement for human effort. While AI dramatically reduces the amount of time spent on manual research, the true "desirable difficulty" currently showcases: connecting ideas. For instance, after providing a potential ideas, students must utilize their own logical and cognitive thinking to synthesize and bridge these concepts into a coherent argument. As a result, this transformation aligns with the Desirable Difficulties Principle, suggesting the long-term retention and understanding are maximized when the brain engages in effortful, active processing (Bjork & Bjork, 2011). Accoding to "Jagged Frontier" study in 2023 from Harvard & BCG, learners levergae AI as a collaborative partner ompleted tasks 25.1% faster and achieved a 40% increase in the quality of their results compared to those working manually (Dell'Acqua et al., 2023). In essence, the stunning symbiotic relationship allows AI to handle the breadth of information retrieval while empowering humans to dedicate their cognitive bandwidth to the depth of critical inquiry and intellectual synthesis.

3.3.4 Case Study

A case study was conducted in the pre-university education context in Albania (2024) to examine the potential adverse effects of AI-assisted learning on critical cognitive skills by researchers from the University of New York Tirana (Albania), Penn State University (USA), and Kazakh-British Technical University (Kazakhstan). Furthermore, the study surveyed 53 students across various majors and academic levels such as Software Engineering, Computer Science, Business, and so on.

In addition, the study employed a quantitative research design to examine the impact of AI-assisted learning on students' critical thinking, academic performance, and problem-solving skills. Moreover, the experiment adopted a survey-based approach, gathering data from 53 university students at a private institution in Albania through a structured questionnaire. In detail, the survey collected valuable information regarding students' prior exposure to AI tools, frequency of AI usage, and their perceptions of its influence on cognitive and academic outcomes. Specifically, the data were analyzed using independent sample t-tests, Mann–Whitney U tests, Pearson correlation analysis, and linear regression techniques to test the proposed hypotheses for ensuring statistical validity.

After gathering and examining the results, the findings revealed that while frequent AI usage was positively correlated with perceived academic performance and assignment efficiency, excessive reliance on AI tools showed a significant negative relationship with problem-solving skills. Surprisingly, these results underscore the complexity of AI integration in education and raise important concerns regarding the maintenance of students' independent critical thinking development.

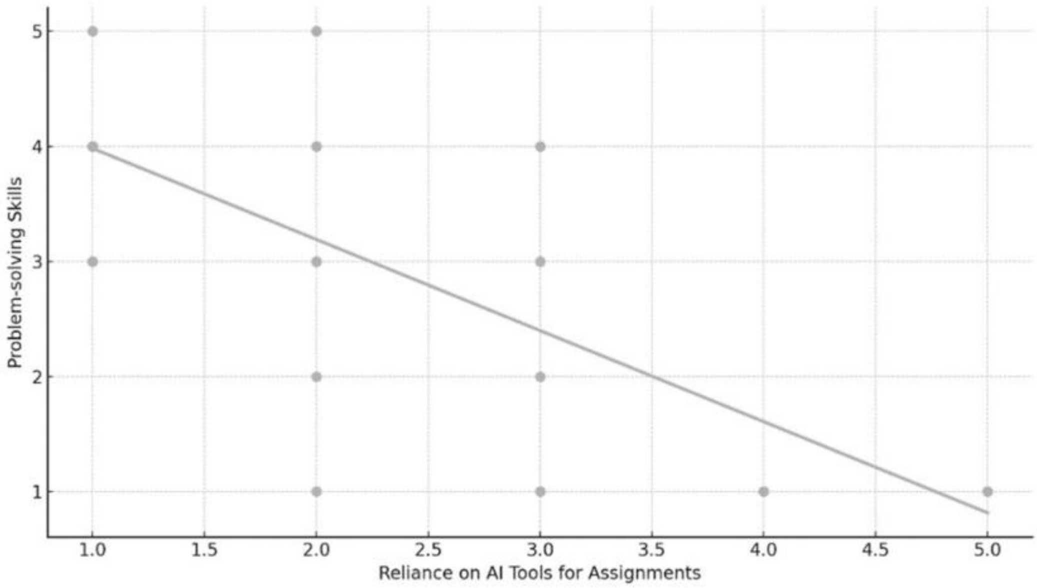

Figure 12: Negative Correlation Between Reliance on AI Tools and Problem-Solving Skills

The Figure 12 indicates the relationship between students' reliance on intelligence tools for assignments and problem-solving skills. Particularly, data points in the scatter plot showcase students' responses while the downward-sloping regression line illustrates a negative correlation between the two variables. In simple terms, the more students rely more on AI tools, the lower their problem-solving skills tend to be. Although available intelligence models provide enhance efficiency and productivity in the educational context, excessive dependence on these tools may weaken students' independent thinking and limit the development of essential cognitive and logical skills.

3.4. Challenges and Obstacles

3.4.1 Cognitive Atrophy

The study emphasizes that automation of analytical tasks can erode the learners' ability for independent reasoning, complex problem-solving, and active learning. First and foremost, the integration of AI in education rises a significant risk known as “Cognitive Atrophy.” According to research by Favero et al in the year 2024, the phenomenon occurs when students utilize AI as a "cognitive shortcut" to bypass the essential mental effort required for deep learning. Notably, the study argues that by providing immediate, polished solutions, AI models can disrupt the "productive struggle" which is the necessary process of trial and error that strengthens the brain's cognitive muscles. When learners outsource critical tasks including brainstorming, logical verification, and structural synthesis to machine learning systems, they are faced with a gradual decline in their independent intellectual capacity. Consequently, active processing in the brain is dismissed because the immediate availability of AI-generated answers eliminates the need for retrieval practice and critical analysis, effectively outsourcing the cognitive labor that is fundamental to intellectual growth.

3.4.2 The Illusion of Knowledge

Next, students who frequently utilize AI models in personalized thinking often fall into a deceptive psychological trap known as the “Illusion of Knowledge.” As noted by Favero et al. in 2024, the phenomenon above happens when learners mistake the AI’s fluency and tailored output for their own internal mastery of the subject. In particular, the study case in Albania revealed a significant disparity between students’ perceptions and their actual performance while learners reported increased academic efficiency and their independent problem-solving skills showed a sharp decline. As a result, this discrepancy highlights that although students may feel highly confident in their grasp of the material, this confidence is often misplaced. Moreover, students can struggle to retrieve information or articulate a logical argument independently without AI assistance. Thus, they become "passive witnesses" to AI-generated solutions rather than active learners, leaving them unable to demonstrate genuine competence or structural reasoning when the technology is removed.

3.4.3 Fake Facts and Passive Acceptance

Then, the reliance on computational models in personalized learning environments fosters a dangerous tendency toward the passive acceptance of fake facts or AI hallucinations. As reported by Favero et al, statistical observations in recent studies suggest that large language models can produce fabricated information in a significant percentage of complex queries, indicating these errors with high linguistic fluency and an authoritative tone. In addition, research represents misinformation remains a potential risk. For instance, benchmark testing on the TruthfulQA dataset revealed that top models were only truthful on about 58% of queries while human performance reached 94% (Lin et al., 2021). Surprisingly, the result illustrates that over 40% of model outputs could be false or misleading, highlighting that that the deceptive combination of AI’s high confidence and structural fluency can effectively bypass a student’s critical thinking.

In addition, the danger of AI dependency is exacerbated by the lack of active modification and critical verification among learners. Nowadays, students use generated outputs as final answers rather than raw drafts that require cognitive and logical thinking. In contrast, information from models can be inherently flawed or biased, and learners seem to forget to check the accuracy of information before integrating it into their work. By neglecting the crucial modification stage, students fail to engage in the 'evaluative thinking' necessary for deep learning. Furthermore, this approach leads them to become passive consumers, leaving potential errors uncorrected and ultimately weakening their ability to form independent, evidence-based judgments instead of acting as critical editors."

3.4.4 The Erosion of Human Agency

Ultimately, the cumulative effect of cognitive atrophy and the passive acceptance of AI outputs leads to the erosion of human agency in the learning process. As demonstrated by Favero et al. (2024), crucial soft skills such as problem-solving and the logical and cognitive thinking of students are dramatically reduced because they heavily rely on utilizing AI tools. Consequently, this over-reliance shifts the learner's role from an active creator to a passive participant, preventing the crucial activities that occur in the brain during the process of independent reasoning and discovery. Therefore, the loss of autonomy limits the development of personal voice and original thought as students lose the intellectual courage to explore unconventional solutions. Finally, the outcome of this report emphasizes on using AI as a supportive tool rather than a replacement for human intellect to preserve the integrity of the learning experience.

3.5. Solutions: Shaping a Responsible AI Future

3.5.1 Technological Solutions

The mitigation of AI-related risks requires a fundamental shift in architectural design, moving away from systems that prioritize output speed toward those that emphasize process transparency. As noted by Adadi and Berrada (2018), the implementation of Explainable AI (XAI) provides the underlying logic behind AI-generated decisions, requiring students to critically evaluate the reasoning process rather than passively accepting a final answer. Furthermore, this transparency is extremely crucial in solving the "Illusion of Knowledge" problem by forcing learners to engage with structural evidence. By integrating real-time verification and 'human-in-the-loop' safeguards, technology can be transformed from a cognitive shortcut into a transparent partner that preserves the integrity of the learning process.

3.5.2 Educational and Human Solutions

Beyond technological safeguards, public and private prestigious schools and universities should have a paradigm shift in pedagogical approaches to foster critical AI literacy. According to the report of Selwyn (2024), these institutions should apply latest assessment models which move away from rewarding final outputs and instead prioritize the evaluation of the learning process. Therefore, the approach can be achieved by implementing "AI-resilient" assignments including reflective journals, oral defenses, and in-class problem-solving tasks that require human nuance and independent reasoning. Besides that, integrating AI ethics into core education principles enables students to act as "critical editors" who can identify biases and verify AI-generated content (Laupichler et al., 2022). By ensuring students use AI as a “thinking partner” rather than a cognitive substitute, prestigious academic institutions can ensure that the development of original thought and human agency remains at the heart of higher education.

3.5.3 Policy and Ethical Frameworks

Finally, governments and academic institutions should implement standardized guidelines that define the ethical boundaries of AI assistance as highlighted by Miao et al. (2023) regarding the policy and ethical frameworks. Consequently, this aprroach ensures that intelligence tools support students rather than replace human intellectual effort. Besides that, strict policies should mandate algorithmic accountability which requires AI developers to provide transparency regarding the data used to train educational models to prevent inherent biases. According to UNESCO (2021), institutions should launch suitable approaches where privacy protection and data sovereignty are embedded into the personalized learning infrastructure. By generating a clear legal and ethical roadmap, policymakers can mitigate the risks of unauthorized dependency while fostering an environment where AI usage is governed by principles of fairness, transparency, and human-centricity.

Part D: Conclusion

In summary, the integration of artificial intelligence into education is not only a technological upgrade but also a stunning transformation that challenges the traditional core of learning. While AI offers powerful assistance in learning and teaching, it acts as a double-edged sword. However, in the report, the shift offers potential risks including cognitive atrophy, the erosion of human autonomy, and the "illusion of knowledg.". Due to above reasons, from my perspective, prestigious organizations and academic institutions should implement proposed solutions such as adopting Explainable AI (XAI)for transparency, shifting toward process-based assessments in prestigious institutions, and enforcing robust ethical frameworks.Although these recommended solutions may require significant effort to implement,I believe that they can reduce partly student dependency on automated tools and effectively safeguard the long-term value of human intellectual development.

Part E: Acknowledgements

During the wirting process, I have used a combination of academic research from prestigious organizations and advanced AI assistance. In detail, the following lists are

- UNESCO (2021 & 2023): I utilized these reports to generate a solid foundation for data sovereignty and the "Ethics-by-Design" approach in educational infrastructure.

- Miao et al. (2023): This research was crucial for my analysis of standardized guidelines and the ethical boundaries required to prevent AI from replacing human thought.
- Adadi & Berrada (2018): I used this paper to deeply understand and integrate the technical concept of Explainable AI (XAI) into my proposed solutions.
- Selwyn (2024): This resource assisted me in identifying and articulating the risks of "cognitive atrophy" and the erosion of student autonomy.
- Strategic Collaboration: I utilized the AI as a tool to structure and polish my original ideas. Throughout the report, I provided the core concepts and specific arguments, demanding the AI to continuously update and refine the content to match my evolving perspective.
- Iterative Refinement: The AI assisted in ensuring academic tone and grammatical precision. However, I maintained strict control over the "human-written" quality, repeatedly instructing the AI to modify its outputs to ensure they remained authentic to my voice and met my specific structural requirements.
- Verification: I personally reviewed and adjusted every section to ensure that the AI's assistance remained secondary to my own research and intellectual direction

Part E: References

(1) Stanford Institute for Human-Centered Artificial Intelligence. (2024). *AI index report 2024*. Stanford University.

https://hai.stanford.edu/ai-index

(2) Ng, A. (2016). *What artificial intelligence can and can't do. Harvard Business Review*

https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do

(3) Russell & Norvig (Primary Source): Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

https://www.pearson.com/en-us/subject-catalog/p/artificial-intelligence-a-modern-approach/P200000003311/9780134610993

(4) McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12–14

https://doi.org/10.1609/aimag.v27i4.1904 (Original work published 1955).

(5) Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

https://doi.org/10.1038/323533a0

(6)IBM. (n.d.). *Deep Blue*. Retrieved from

https://www.ibm.com/history/deep-blue

(7) Campbell, M., Hoane, A. J., & Hsu, F. H. (2002). Deep Blue. *Artificial Intelligence, 134*(1-2), 57-83

https://doi.org/10.1016/S0004-3702(01)00129-1

(8) Ng, A. (n.d.). *AI is the new electricity.* Retrieved Month Day, Year,

https://time.com/partner-article/7279245/15-quotes-on-the-future-of-ai/

(9) UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO.

https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

(10) BestColleges. (2023). *Most college students have used artificial intelligence tools, survey finds*.

https://www.bestcolleges.com/research/most-college-students-have-used-ai-survey/

(11) Feedough. (2024). *AI cheating statistics: How students are using ChatGPT in education*

https://www.feedough.com/ai-cheating-statistics/\

(12) FindTutors UK. (2024). *AI cheating in schools and universities: Key statistics*

https://www.findtutors.co.uk/blog/ai-cheating-in-schools-statistics

(13) Nordic Today. (2025). *AI cheating cases double at Norwegian universities*

https://nordicstoday.com/article/no-norway-ai-cheating-cases-double-190-students-caught

(14) University of Alberta. (n.d.). *Academic integrity and the use of artificial intelligence*. Centre for Teaching and Learning.

https://www.ualberta.ca/en/centre-for-teaching-and-learning/resources/generative-ai/academic-integrity-ai-use/index.html

(15) RMIT University. (n.d.). *Academic integrity and artificial intelligence*

https://www.rmit.edu.au/students/my-course/assessment-results/academic-integrity

(16) Department for Education (UK). (2025). *Academic integrity and misconduct policy*.

https://dghe.ac.uk/wp-content/uploads/2025/04/Academic-Integrity-and-Misconduct-Policy-v4.2.pd

(17) Digital Learning Council. (2024). *The AI literacy gap in public versus private education*.

https://www.digitallearningcouncil.org/reports/ai-literacy-2024

(18) Higher Education Policy Institute. (2024). *Student generative AI survey 2024*.

https://www.hepi.ac.uk/2024/02/26/student-generative-ai-survey/

(19) UNESCO. (2023). *Global education monitoring report 2023: Technology in education: A tool on whose terms?*

https://unesdoc.unesco.org/ark:/48223/pf0000385723

(20) Khan Academy. (2024). *AI in the classroom: How personalized tutoring is changing student outcomes*

https://www.khanacademy.org/about/blog/ai-education-report-2024

(21) UNESCO. (2024). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization

https://unesdoc.unesco.org/ark:/48223/pf0000386693

(22) Chomsky, N. (2023, March 8). The false promise of ChatGPT. *The New York Times*.

https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html