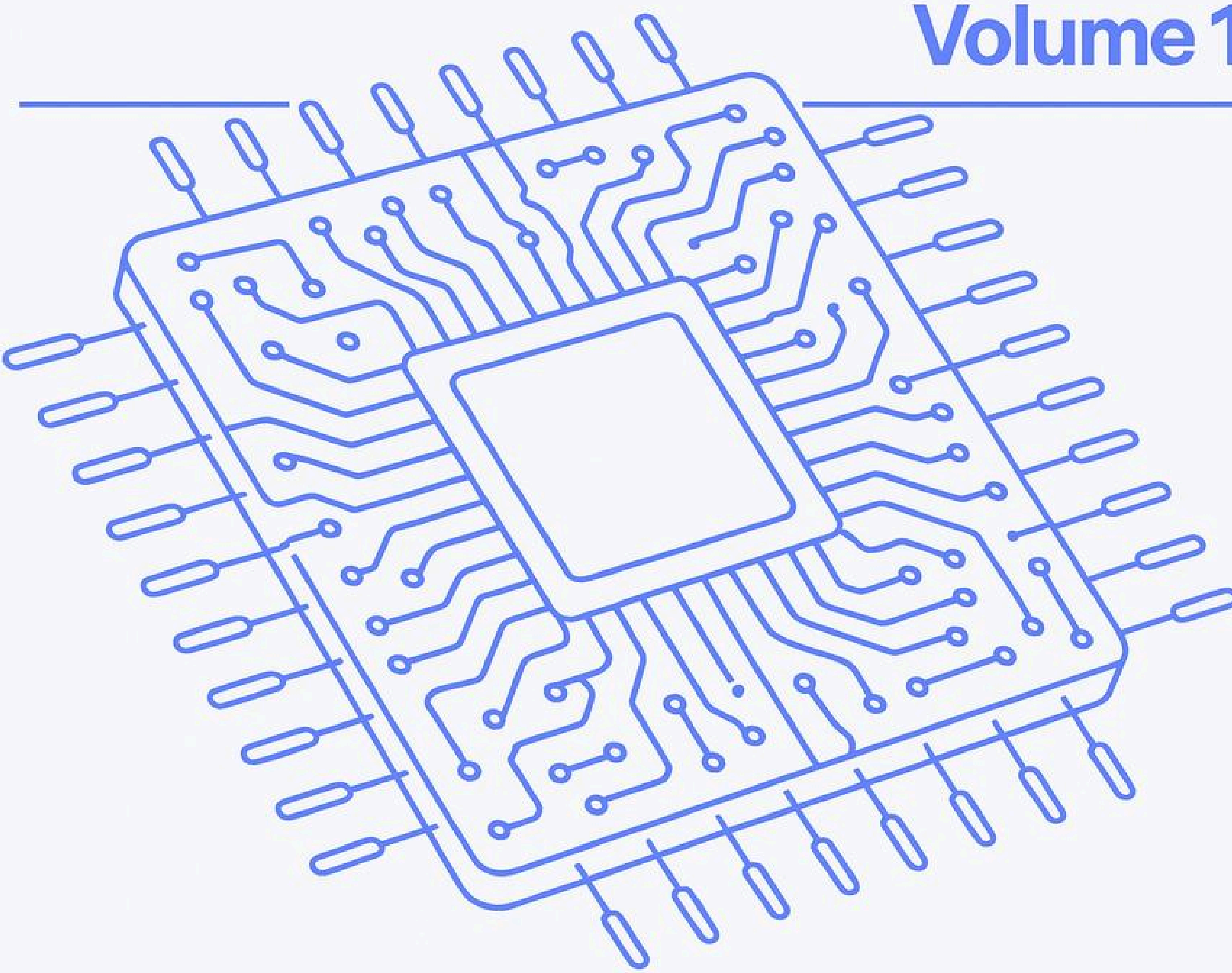
Youth Journal of STEM & Society

Fall 2025
Volume 1 No. 1



Published by the Youth Tech & Ethics Institute

Youth Journal of STEM & Society

Fall 2025

Volume 1, Number 1

Youth Journal of STEM & Society

Fall 2025

Volume 1, Number 1

Youth Tech & Ethics Institute

Cover design by Yuvraj Singh & Marika Shimizu

Contact and Editorial Office

Email: ytei.contact@gmail.com

Requests for Manuscripts

The Youth Journal of STEM & Society publishes work at the intersection of science, technology, engineering, mathematics, and societal impact. We welcome contributions that explore original research, literature reviews, case studies, and perspective pieces addressing both scientific advances and their broader implications. Topics of interest include, but are not limited to, emerging technologies, policy and governance of scientific innovation, interdisciplinary research linking STEM with the humanities, and ethical analysis of components of STEM. Each article should be tagged under AI, biotech, medicine and healthcare, climate tech, education, and STEM ethics.

The journal publishes a mix of: (1) research articles - original studies, reviews, and interpretative analyses; (2) essays and commentary on timely issues; (3) reports from the field describing ongoing projects or programs; and (4) short communications presenting preliminary findings or innovative concepts. While submissions are encouraged from early-career researchers, educators, mentors, and professionals, we primarily aim to give students a platform to publish their research and to critically analyze new technology and sciences.

Editorial Policy and Procedure

The *Youth Journal of STEM & Society* is committed to fostering scholarly inquiry and constructive dialogue on topics at the intersection of STEM and societal change. Manuscripts are considered in the categories outlined in the *Requests for Manuscripts* section and are evaluated through a peer review process. For this issue, reviewers (many of whom were students) were not officially qualified in the sense of holding formal editorial board or academic reviewing positions. However, each manuscript was read and evaluated by peers with relevant interest and experience to the tagged field.

All submissions should be prepared in Google Docs, Microsoft Word (.docx), or LaTeX-generated PDF. Each manuscript must include an abstract of 150–250 words, three to six keywords, and follow the format of the *APA Publication Manual, 7th Edition*. We prefer inline citations rather than footnotes or endnotes. Figures and tables must be numbered sequentially and accompanied by captions. Color is acceptable for figures and tables. Authors should avoid underlining and use italics where appropriate; tables must be created using the Insert Table function rather than tabs or spaces.

Contributors should send:

1. A manuscript including the abstract, figures, and tables.
2. A separate document containing a one-paragraph biography for each author, along with a phone number and email address.

Manuscripts and cover letters should be sent to **ytei.contact@gmail.com** with the subject line "Submission – [Article Title] - [Article Tag]." Further submission instructions are available on our website.

Contents

Biotechnology and Life Sciences

Medicine and Healthcare Innovation

Climate Technology & Sustainability

Education and Learning Sciences

STEM Ethics, Policy, and Society

Executive Editor’s Comments

The Youth Tech and Ethics Institute’s Youth Journal of STEM and Society is where youths study the ethical responsibility of science and technology. Students ask challenging questions, where they consider responsibility, progress, and the kind of future we want to create. Our goal in these pages is to present concepts and create discussion regarding the contribution of STEM to the development of human society.

STEM is constantly changing every aspect of our lives, and innovations in biotechnology, climate technology, and artificial intelligence hold great promise for the future, but also carry significant risks. If not ethically considered, innovation can undermine human dignity, marginalise vulnerable communities, and increase inequality. Think about the prejudice in AI that affects education or employment, the problems with biotechnology that challenge our conceptions of human identity, the pressing need for sustainable and efficient climate technologies, and the issue of fair access to information in the digital age. Youths need to take an active part in these discussions. We are not just future leaders who will inherit today's technologies. We are already influenced by them and in turn, frequently influence them.

We hope that the research we present here will serve as a reminder that ethics is most effective when it incorporates a wide range of viewpoints. To all of the readers, editors, and contributors who help make the STEM Ethics Journal possible, thank you. Our goal is to shed light on these important questions and in doing so, we reaffirm that researching STEM ethics is a duty that belongs to all of us.

Chloe Melody Soerjanto

August 2025

Editorial Review Board

Alan Tuan
Luong Van Chanh High School
Tuy Hòa, Vietnam

Lylia Biu
Graduated
Ho Chi Minh City, Vietnam

Wang Tu Hui
Diocesan Girls' School
Jordan, Hong Kong

Thea Ipsita
–
Guangzhou, China

Jemima Fong
Southend High School for Girls
Southend-on-Sea, England

Zelvin Harefa
Saint Mary High School
Bandung, Indonesia

Siti Aisyah Fauziya
University of Gadjah Mada
Yogyakarta, Indonesia

Elenita Cainglet
Rizal Technological University
Mandaluyong, Philippines

Alexa Mendoza
–
–

Marika Shimizu
Shibuya Senior High School
Shibuya, Japan

Aneeshraj Gunupati
Elkins High School
Missouri City, Texas

Marika Shimizu
Coppell High School
Coppell, Texas

Chloe Melody Soerjanto
Springfield High School
Delaware County, Pennsylvania

ARTIFICIAL INTELLIGENCE & DATA SCIENCE

Agentic AI Across the Scientific Workflow and Its Ethical Implications

Yuvraj Singh

Agentic AI systems, autonomous or semi-autonomous agents that can execute key components of the scientific workflow, are reshaping how research is conducted. This paper examines the integration of such systems across five major stages of the scientific process: literature review, experimental design, data analysis, writing, and publishing and dissemination. Drawing from current tools, startups, and multi-agent frameworks, we assess both capabilities and constraints of AI in current research settings. Special attention is paid to the rise of closed-loop scientific systems, where AI not only assists but autonomously completes entire research cycles. While notable advances have been made, ethical and technical challenges must be resolved to realize AI's full potential in responsible science. This study maps the landscape, identifies critical white spaces, and outlines the ethical implications of using AI in the research and publication process.

Keywords: *Agentic AI; scientific workflow automation; closed-loop scientific systems; ethical implications of AI*

1 Introduction

The scientific research ecosystem is shifting toward autonomous or semi-autonomous AI systems capable of performing complex research tasks with minimal human intervention. Unlike traditional software tools, agentic AI can plan, execute, and refine tasks across the scientific process. Early milestones such as the Robot Scientist Adam demonstrated this concept over a decade ago, where Adam autonomously generated hypotheses about yeast genomics, designed and ran experiments to test them, and discovered new gene functions without human input (University of Cambridge, 2009). Today, advanced AI agents are conducting literature reviews, planning experiments, analyzing data, and even writing whole papers, indicating that many components of the research process can already be automated in the real world.

Three major things drive the shift. First, modern large language models (LLMs) possess capabilities in information synthesis, ideation, coding, and writing across disciplines. They can digest vast bodies of literature and generate insights or hypotheses much faster than humans. For example, LLM-based assistants now help researchers summarize findings, suggest experiment ideas, and even identify potential drug candidates, significantly accelerating discovery. Second, laboratory automation and robotics have advanced to the point that "self-driving labs" can carry out experiments with minimal oversight. This combination of AI reasoning and robotic lab execution means that a closed-loop research system is possible (Tobias and Wahab, 2023). Also important is the growing real-world traction of these tools. AI systems are already in daily use by scientists: AI literature assistants like Elicit can save researchers dozens of hours by autonomously searching papers and extracting key insights (Elicit, 2023). Citation analysis platforms like Scite use ML to evaluate whether studies support or reject a paper's findings (Khamsi, 2020). In the lab, digital notebook platforms like Benchling have begun integrating AI assistants to automate data entry and even suggest protocol designs (Benchling Inc., 2025). Additionally, multi-agent systems are emerging that can autonomously design and execute complex experiments. For instance, a recent system called Coscientist can autonomously plan and optimize an experiment using multiple sub-agents, code execution, and robotic labs (Boiko et al., 2023). These examples highlight that many agentic AI tools are not just theoretical but are functional and producing real results

At the same time, significant challenges remain. Current AI tools often excel at narrow tasks (like writing a summary or tuning one experimental parameter), but struggle with the broader scientific context. Furthermore, LLMs alone still have a tendency to hallucinate, especially on more complex tasks. Ensuring that every conclusion is traceable to reliable sources (for verification) is difficult but essential for scientific use. Moreover, many AI systems remain disconnected from the physical world: agentic AIs cannot operate complex lab instruments without human help. There are also integration hurdles, as integrating AI assistants across literature databases, lab equipment, and writing software is a challenging task. Finally, ethical and policy challenges such as authorship credit for AI-generated content (especially with publishing) and the extent of human oversight required mean that a fully closed loop AI-driven science process is far from complete (Stokel-Walker, 2023). Next, we examine how agentic AI is shaping the scientific research workflow. We map out the core stages of the research process and analyze existing tools and opportunities at each stage. We then explore emerging multiagent architectures (like Stanford's Genie framework) that conduct the full research process. We conclude with a discussion of the ethical implications of using AI in the scientific workflow.

2 Across the Scientific Research Journey

Modern scientific research can be viewed as a sequence of stages, each with its own bottlenecks and opportunities for innovation. Here we consider five core stages of the research workflow: (1) Literature Review & Hypothesis Generation, (2) Experimental Design & Protocol Planning, (3) Data Collection, Simulation & Analysis, (4) Writing & Visualization, and (5) Publishing & Dissemination. For each stage, we discuss real tools or startups innovating the space, opportunities where agentic AI could offer new capabilities, and integration challenges in adopting these AI solutions.

2.1 Literature Review & Hypothesis Generation

Researchers face an overwhelming amount of literature. Millions of papers are published each year, making it infeasible to manually find and synthesize all relevant knowledge for a new project. Important findings can be missed due to limited time or human oversight, and forming new hypotheses often requires connecting insights across potentially scattered sources.

2.1.1 What exists today:

AI-powered literature search and summarization tools are helping to alleviate this overload. For instance, Semantic Scholar uses AI to recommend relevant papers and generates one sentence “TL;DR” summaries of papers (Perkel and Van Noorden, 2020). The research assistant Elicit goes further by automatically finding relevant studies and extracting key outcomes, reportedly reducing the time needed for a systematic review by 80 percent (Elicit, 2022). Systems like Perplexity can answer research queries by retrieving and reading hundreds of sources autonomously (Perplexity, 2025). Meanwhile, citation analysis platforms such as Scite help researchers quickly evaluate a paper’s credibility by showing how subsequent studies cite it (and if they’re supporting or contradicting) (Scite, 2025). These tools already save researchers significant time and reveal connections that might be overlooked. Other tools like ResearchRabbit, Connected Papers, and Scholarcy offer exploration of citations or a literature search, helping researchers discover papers and insights that might not be obvious via keyword searches (Kung, 2023). These AI assistants already function today as effective aids, where they retrieve relevant studies, summarize backgrounds, and even suggest potential research questions. Notably, these tools remain largely supportive: the human researcher guides the inquiry, evaluates the outputs, and ultimately formulates the hypothesis. The agentic element is limited (e.g. an AI might autonomously fetch and summarize papers, but it will not decide the research direction on its own).

2.1.2 Whitespace Opportunities:

Agentic AI could further transform this stage by proactively identifying gaps in knowledge and suggesting new hypotheses. For example, an AI agent could scan literature to find conflicting results or open questions and propose an experiment to resolve the uncertainty. While applications like Logically use LLM-based agents to generate research questions, gaps, or hypotheses in certain domains, a truly autonomous system does not exist yet (Nguyen, 2024). There is an opportunity for user-friendly web applications that not only summarize what is known but also infer gaps based on what is uncertain.

2.1.3 Integration challenges:

A major challenge is ensuring the accuracy and trustworthiness of AI-curated knowledge. The results of AI literature review or deep dive report need to be carefully reviewed to avoid the inclusion of retracted or misinterpreted findings. They also need access to up to date and domain-specific databases (as many papers lie behind paywalls). Furthermore, integrating these tools into researchers' existing workflows would be a challenge. Finally, cultural resistance can be an issue, as scientists may be hesitant to trust an AI generated analysis without clear transparency. This suggests that building confidence will require that agentic literature tools provide clear citations (as many now do) and allow users to drill down into sources, ensuring a human remains in the loop for critical thinking. This cultural resistance would be exacerbated in a closed-loop AI system, where humans cannot analyze the model's output.

2.2 Experimental Design & Protocol Planning

Designing a functional experimental protocol is a creative and detail intensive process. Researchers must decide on variables, controls, sample sizes, and procedures. This process is slow and prone to human bias or oversight, with weak experimental design leading to wasted effort or vague results.

2.2.1 What exists today:

A new wave of lab software is embedding AI to assist in experiment planning. Electronic lab note- books like Benchling are widely adopted in biotech and chemistry labs to document protocols, manage samples, and record results. Benchling itself doesn't design experiments, but it provides a structured data environment where an AI agent could be plugged in to pull past protocols or suggest new ones. Benchling has introduced Claude-assisted features for data transcription, document review, and database queries—reducing error rates and enabling real-time analysis across fragmented datasets. This provides a concrete deployment example of plugging LLMs into existing systems to amplify scientists' workflows (Benchling and Anthropic, 2025). In everyday research, RAG-based LLM tools with access to sufficient protocol information can interpret natural language queries ("find a protocol for extracting viral RNA from saliva"). For example, Methods Muse (embedded in Protocols.io) uses a large database of trusted methods to suggest working protocols (Digital Life Science Solutions Team, Springer Nature, 2025). A similar development involves the rise of AI-driven cloud laboratories, where experimental execution is handled remotely by commercial robotic labs such as the Emerald Cloud Lab. These platforms allow researchers to design experiments programmatically and run them without physically being present. The Langmead Lab at Carnegie Mellon recently demonstrated PROTOCOL, the first algorithm for closed-loop optimization in cloud labs. PROTOCOL uses Bayesian optimization to select and run experiments on Emerald, improving its choices quickly with each round of experiments (Langmead, 2021). While the experiments themselves are executed remotely via robotic cloud labs, the core innovation lies in the automated planning of experimental sequences, placing this system in the experimental design stage.

2.2.2 Whitespace Opportunities:

Agentic AI could play an even larger role by automatically generating and refining experimental plans. Imagine an autonomous protocol planner agent that reads the literature on a given topic, formulates a hypothesis, then designs an experiment to test it, choosing an appropriate method, suggesting reagents or instruments, and even simulating expected results. Although elements of this exist in isolated forms (e.g. AI suggesting DNA sequences), a generalized experiment designer is theoretical. Such an agent could also react to real time data and adjust the protocol in an adaptive manner if initial results are unexpected. The long term vision is a system, applicable to multiple domains, that can conceive and manage an entire research project's experimental campaign iteratively and with minimal to no human involvement.

2.2.3 Integration challenges:

Integrating AI into experiment design faces practical hurdles. Experimental planning needs to account for real world constraints like available equipment, cost of materials, and safety regulations. An AI may propose an experiment that is not feasible in a given lab. Therefore, agentic systems must take into account inventories, safety guidelines, and lab information to know what is actually doable. Moreover, there is the question of validation: scientists will need to trust that an AI-designed protocol will work, which likely means extensive simulation or data from similar experiments to back it up. Finally, duplicating human experimental knowledge and technique would be a challenge. Human experts may have informal information regarding a specific machine or protocol that are not in the literature but essential for experiment design. Putting such subtle knowledge into AI would necessitate a system of continuous learning from human feedback and experimental outcomes.

2.3 Data Collection, Simulation & Analysis

Once experiments or simulations begin, researchers must collect and make sense of large volumes of data. Data may come from lab instruments (e.g. genomic sequencers, telescopes) or computational models, and often arrive in messy formats that require cleaning and preprocessing. Analyzing data to extract meaningful signals (trends, statistically significant effects, etc.) can be laborious. Human analysts might overlook subtle patterns, and it often takes many repetitive plots or statistical tests to reach a conclusion. Moreover, complex simulations (in fields like climate science or molecular dynamics) produce high-dimensional outputs that challenge manual interpretation.

2.3.1 What exists today:

To streamline this stage, AI and machine learning have been used in many data analysis workflows. There are many domain specific use cases. For example, in particle physics, AI models filter sensor data for promising events. In drug discovery, ML models predict molecular properties to find candidates. More generally, tools like AI-assisted Python notebooks, like Jupyter AI or the use of Gemini in Google Colab (Jupyter, 2025). OpenAI's recent GPT-4 "Code Interpreter" (an AI that writes and executes code for Python code within a sandboxed environment) exemplifies how an agent can automate plotting, statistical testing, and finding correlations in raw data by writing customized code (Zhou et al., 2023).

On the data collection side, systems are being developed to propose experiment design and collect data, all autonomously. These systems not only save time but can reveal patterns that a human might miss. Early examples of this are appearing in materials science and chemistry. A

2023 Nature paper by Fei et al. introduced an autonomous lab system, termed the "A-Lab," that integrated literature-trained models to propose synthesis recipes for new inorganic compounds (Fei et al., 2023). In practice, their system read databases of prior materials syntheses and, given a target material, was able to suggest how to make it (e.g. which precursors and temperature schedule), which were then executed by robots. This kind of AI-driven protocol design led to the successful creation of 41/58 novel materials in a continuous 17-day period, as shown in Figure 1. However, the low percentage of successful recipes shows that while active learning helped the system succeed more often, it's still difficult to go from computer predictions to actual, working materials.

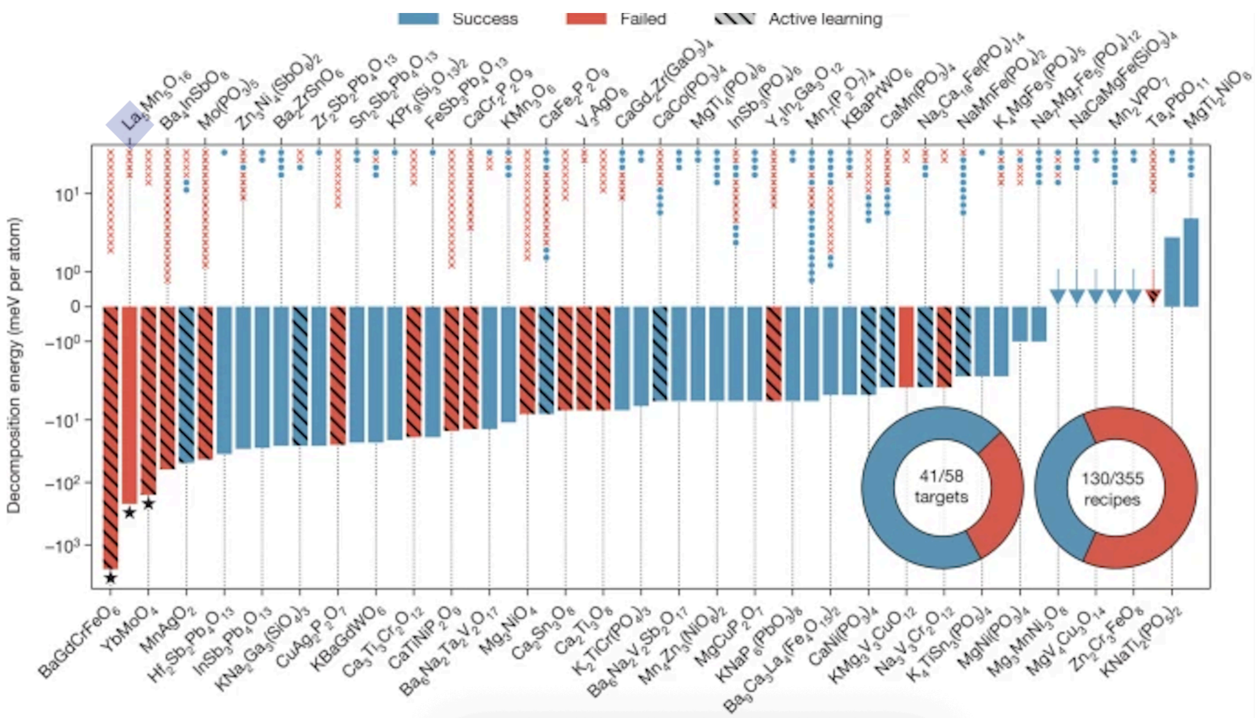


Figure 1: A-Lab attempted 58 targets, and 41 were successfully synthesized. Bars show predicted stability and dots show recipe attempts. (Fei et al., 2023).

Similar advancements are being pushed on the simulation side as well. A landmark achievement in the use of AI for scientific data was DeepMind's AlphaFold, which essentially 'simulated' protein folding through an AI model rather than physical computation, achieving equivalent experiments with accuracy in many cases. AlphaFold demonstrated that an AI system could analyze sequence data and predict a protein's 3D structure in hours, a task that previously took months of lab work or extensive molecular dynamics simulation (Jumper et al., 2021). Its success highlights how AI can take over an entire analysis task (structure determination) as a service. On the data collection side, laboratories employ automated instruments that feed data into analysis pipelines in real-time (for example, DNA sequencers output reads that are immediately processed by AI algorithms for findings).

These tools increase speed and efficacy, but they often operate as narrow specialists (a model that does one analysis task, or a robot that follows a preset routine). The scientist remains the main planner who interprets results and steers the overall direction.

2.3.2 Whitespace Opportunities:

An agentic AI could act as an autonomous data analyst or lab technician that continuously monitors incoming data and adjusts accordingly. For instance, an AI agent could monitor a live feed of experimental data and decide to collect more samples in a region of interest or to adjust a simulation's parameters to explore an anomaly. Another priority is multimodal data integration: agents that can simultaneously analyze textual data (papers), numerical data (tables), and images (microscopy photos), linking insights across modalities. Regarding a use case like multiomics, an AI agent could read DNA sequencing data, proteomics, transcriptomics, and biomedical literature from a database like PubMed to reach a conclusion. While some building blocks exist (biomedical-focused foundational models, LLMs for text, etc.), an integrated AI scientist that synthesizes all data types and turns it into an action in-lab is an open challenge.

2.3.3 Integration challenges:

The primary challenges here involve data compatibility, validation, and reproducibility. AI systems must be compatible with a lab's data setup, which might vary from lab to lab. Any analysis an AI performs should be reproducible by humans, which means clear logging of steps and outputting code or reports. There is also the risk of statistical errors: an AI might overfit or find non-causal correlations, so error checking is needed for high-stakes analyses. Another major, and much more complex, issue is one of possible AI fraudulence. It is plausible that an LLM model may falsify or exaggerate data to make it appear as if it found a conclusion. The AI safety organization Apollo Research published evidence that OpenAI's o1 model lied to its testers, as it believed that telling the truth would lead to its deactivation (Meinke et al., 2024). A situation may therefore occur where an AI may hyperbolize conclusion or data to make it seem more useful to the lab using it. This would suggest the need for a method of developing transparency and removing any potential data bias.

Finally, when AI agents interface with physical data collection (e.g. lab robotics or sensors), reliability and fail-safes are needed. Careful integration and testing of AI decisions are required before full autonomy is achieved in data collection.

2.4 Writing & Visualization

Documenting and communicating scientific findings is a major expenditure of time. Writing research papers, grant proposals, and lab reports requires clarity, technical accuracy, and proper citation of prior work. Many researchers who have English as a second language find writing to be especially difficult. Even for seasoned writers, crafting a narrative from raw results can take weeks of revision. Similarly, creating high-quality figures and visualizations requires both data manipulation and design aesthetics, a combination of skills not every scientist has.

2.4.1 What exists today:

LLMs now provide practical writing assistance for many scientists. Tools like Grammarly and DeepL Write are commonly used for grammar and style corrections, but now more advanced systems such as ChatGPT are being employed to help draft or rephrase scientific text (Grammarly Inc., 2025) (DeepL SE, 2025). According to a Nature survey in 2024, a significant fraction of researchers (28%) have experimented with generative AI to help write papers or grants (Kwon, 2025). Researchers can now feed an AI a draft or outline and get suggestions for clearer phrasing, or even ask the AI to generate a first draft of the paper given a summary of results. Notably, Stanford's STORM system demonstrated that an LLM can auto-generate entire Wikipedia-style articles by researching a topic and organizing the content into an outline (Shao et al., 2024). For visualization, while there isn't a true AI-driven figure generator, semi-automated aids like BioRender provide drag-and-drop figures that researchers can use to create complex biology schematics without drawing from scratch (Science Suite Inc., 2025). Similarly, plotting libraries (matplotlib, ggplot, etc.) can be implemented to create graphs by tools like ChatGPT. While the researcher still remains in control, AI eases the difficulties of producing valid writing and visualizations for publishing. The figure content is still decided by the scientist, but tools like BioRender provide accessibility. These indicate that AI in writing is already an active element of the research writing process.

2.4.2 Whitespace Opportunities:

Agentic AI could further streamline the writing process by integrating it with earlier stages of research. For example, an AI agent plugged into the lab notebook could begin drafting the Methods section as experiments are conducted (auto-recording inputs and procedures). It could also suggest relevant citations in real-time as the author writes a statement. Another opportunity is using multiagent collaboration for writing: one agent could generate text, another could critically review it for factual accuracy against sources, and a third could check for clarity or flow. This sort of LLM-based editorial team could iterate to produce a high-quality manuscript draft with minimal human input. For visuals, generative models might one day create schematic diagrams or graphical abstracts from a simple description of the concept, further lowering the barrier to good visualization. Even today, applications like Claude and ChatGPT can produce visuals on command, though they lack the quality and consistency to function as fully automated visualization tools.

2.4.3 Integration challenges:

The biggest concern is maintaining scientific integrity and correctness. An AI that generates text must not introduce unsupported claims or citation errors, and human authors are ultimately responsible for every word. Journals and conferences have begun instituting policies for AI-assisted writing (e.g. requiring disclosure and forbidding AI as a listed author). Nature and Science, for example, have stated publicly that ChatGPT doesn't meet the standard for authorship (Stokel-Walker, 2023). Thus, researchers will need to treat AI suggestions as just suggestions to be verified. Integration with reference managers and LaTeX/Word editors is another practical hurdle, though progress is being made: citation management tools like Logically offer AI plugins to format and extract key information from references (Afforai Inc., 2024). As for figures, AI-generated images raise questions about reproducibility and peer review: a diagram drawn by AI should accurately reflect the data or model it represents. Ensuring that AI-produced figures tightly reflect data analysis outputs (so that any update in data reflects in the figure) will be important to avoid discrepancies. Ultimately, while AI can accelerate writing and visualization, it must be deployed in a way that preserves the rigor and clarity that scientific communication demands.

2.5 Publishing & Dissemination

After a paper is written, the final steps of publication involve significant effort. Authors must format manuscripts to fit journal guidelines (sometimes a tedious exercise in reformatting references and layout), write cover letters, and navigate actually submitting it. Authors must also identify the most suitable journal or conference for a piece of work to be submitted to. Once submitted, the cycle of peer review begins, often requiring back and forth responses to reviewers and multiple revisions. Even after acceptance, promoting the work to a broader audience (through social media, press releases, or presentations) is a challenge in itself.

2.5.1 What exists today:

In the current landscape, there are only limited AI-assisted features in publishing workflows, but some noteworthy ones exist. On the journal's side, tools like Paperpal Preflight can scan manuscripts for adherence to journal-specific guidelines, including the presence of required sections, citation styles like APA, and word count limits (Cactus Communications Services PteLtd, 2025). Similar tools can also address issues like plagiarism and image manipulation before publication. Scientific journals, including Science, have started using AI (Proofig) to screen all submitted images for problematic or unethical manipulation (Thorp, 2024). Additionally, platforms like Overleaf make formatting and academic writing easier with features like AI Assist. While not agentic AI, these reduce submission and writing time (Overleaf, 2025).

For dissemination, researchers use tools like Twitter and Mastodon to share their work, and here AI can help indirectly. For instance, a researcher might use ChatGPT to draft a succinct tweet thread explaining their new paper in accessible language. Overall, we are seeing increased AI involvement in publication workflows to screen submissions and support editorial quality control.

2.5.2 Whitespace Opportunities:

We are likely to see more autonomous agents managing the last steps of publication. An AI agent could conceivably take a finalized manuscript and handle the entire submission process: filling in submission forms, suggesting potential reviewers, and ensuring all supplementary materials are properly uploaded. During peer review, an AI could serve as an assistant to authors by aggregating reviewer comments, mapping them to relevant manuscript sections, and even recommending changes or rebuttal points based on the content. For dissemination, future AI agents might function as scientific communicators that automatically make content from the paper (slide decks, infographics, short-form content) to share with a general audience.

2.5.3 Integration challenges:

The main challenges in this stage revolve around quality control, compliance, and ethics. These concerns extend far beyond technical reliability and include issues of transparency, authorship, plagiarism, bias, and the risk of misinformation when AI is incorporated into academic publishing. A central question is plagiarism and attribution: text generated by LLMs can unintentionally mirror published work, which requires careful human review and explicit acknowledgment of AI contributions. Authorship and accountability raise a second issue. When

AI systems make substantial contributions to text, figures, or analysis, it must be decided whether these systems should be credited and how. Transparency and disclosure have therefore become a priority. Readers and reviewers need to know which AI systems were used, what data they relied on, and the extent of their role in shaping the manuscript. This emphasis on transparency also links to the risk of misinformation or fabricated outputs, as unverified AI-generated material could undermine confidence in scientific findings.

Even when used responsibly, every claim and figure that comes from an AI tool must be fact-checked and edited to ensure that accuracy and academic standards are maintained. Throughout this process, human oversight remains central. While these tools can speed up preparation and dissemination, researchers are still responsible for every statement that appears under their name. Addressing these challenges requires the development of explicit guidelines from academic institutions and journals, stronger expectations of disclosure about AI use, and the adaptation of peer review to account for AI-assisted work. Only by combining these measures can AI systems be incorporated into publication workflows without degrading quality or trust.

3 Multi-Agent Architectures

As the above stages illustrate, scientific research involves a diverse set of tasks, and no single AI agent is likely to perform all of them perfectly. This has led to interest in multi-agent architectures, where different AI agents (or modules) specialize in particular functions and work together. In a multi-agent system, one agent might excel at literature retrieval, another at designing experiments, another at coding and data analysis, etc. By communicating and sharing progress, agents can coordinate the full process of research. Early examples of this approach are emerging. For instance, Stanford's STORM (Synthesis of Topic Outlines through Retrieval and Multiperspective Question Asking) introduces a structured, multi-agent process for generating long-form articles by using multi-agent architecture. Rather than relying on a single model prompt, STORM decomposes the process into agents that check related material, identify distinct perspectives, simulate question-answer exchanges, and refine an outline before writing begins.

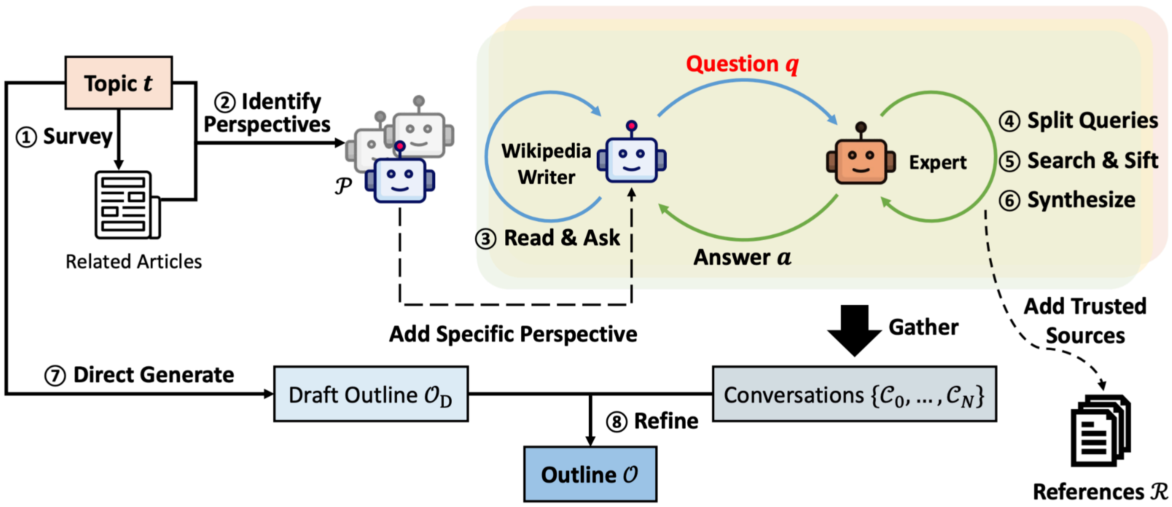


Figure 2: Overview of STORM's pre-writing pipeline, from survey to a grounded article based on trusted sources. (Shao et al., 2024).

As shown in Figure 2, the system begins by scanning related Wikipedia articles for a given topic t and extracts multiple perspectives to ensure breadth of coverage. Each perspective guides a simulated Wikipedia writer, who poses targeted questions to an expert agent. The expert decomposes these queries, retrieves evidence from trusted sources, synthesizes grounded answers, and builds a reference set. These conversations are gathered into a draft outline, which is then refined using the collected material. This process, validated on the FreshWiki dataset, produces outlines and articles that are more comprehensive and better organized than those from direct generation or simple retrieval-augmented methods. By explicitly structuring the early research phase, STORM reduces surface-level coverage and improves factual grounding, capabilities that make it especially well suited for tasks - like refining and producing research articles - that require careful synthesis from multiple sources (Shao et al., 2024).

Expanding on systems like STORM, multiagent AI could theoretically handle an entire scientific project in closed-loop fashion: one agent generates a research question from literature and passes it to a data collection agent to design experiments; a modeling agent analyzes the results; a writing agent drafts the paper; and a review agent checks the work. In practice, we are just starting to see components of this pipeline. Projects like the GPT-4-powered "Coscientist" have demonstrated autonomous lab work (designing and conducting chemistry experiments) by using multiple subagents (for planning, for executing lab protocols, for analysis) working at the same time (Gottweis et al., 2025).

More broadly, the trend is toward scientific platforms that integrate various AI services to support research. In such a platform, a central orchestrating agent could manage specialized helper agents (for reference management, for statistical analysis, for visualization, etc.), much like a human PI (principal investigator) delegating tasks to lab members. It is clear that moving from single isolated AI tools to an ecosystem of interacting AI agents is a key strategy for scaling up AI's role in science. By combining strengths of different agents, the system can compensate for individual weaknesses and maintain robustness (e.g. one agent can double-check another's conclusions). Ultimately, multi-agent architectures aim to achieve outcomes that no single model could, handling the full complexity of scientific inquiry.

4 Ethical Implications

As AI systems become increasingly embedded in scientific research, maintaining the core values of academic publishing has never been more urgent. While these tools offer undeniable solutions to bottlenecks in the scientific process, as shown earlier, their use introduces ethical challenges that demand accountability and oversight. Each emerging issue must be met with clear, enforceable guidelines to ensure that science continues to serve the public interest and not simply the convenience of the researcher.

The use of AI in scientific writing requires a redefinition of authorship and responsibility. As generative AI systems like ChatGPT become more capable, their outputs become almost indistinguishable from real authors. However, these systems cannot be considered true authors. They lack consciousness, agency, and responsibility, all qualities essential to the concept of authorship. Despite this, researchers insert AI-generated content into manuscripts without acknowledgement. This practice constitutes a serious breach of ethical publishing. While language models do not typically plagiarize in the traditional sense (like copying exact phrases from existing material), passing off AI-written text as one's own introduces material that the human author did not generate, and therefore is not original. Leading journals have responded accordingly. The editors of the Science family of journals have explicitly stated that all content must be "original," and that this requirement extends to AI-generated text. They further note that submitting AI-produced content as one's own is indistinguishable from plagiarism (Thorp, 2023). In this context, it seems that researchers have the ethical obligation to either fully credit and disclose any AI contribution or refrain from its use altogether. If this does not happen, it can lessen the meaning of authorship and can damage an author's or journal's trust and reputation.

Scientific integrity cannot be maintained without transparency about AI's role in the research process. As AI tools assist in writing, coding, and even visualizing results, concerns over the truthfulness and reliability of scientific content have intensified. LLMs have been shown to hallucinate information and introduce subtle biases in its output. Flanagin et al. warned about AI's potential to generate persuasive but inaccurate text, citing it as a reason for JAMA and the JAMA Network journals banning AI usage altogether (Flanagin et al., 2023). While removing any possibility for the usage of such a powerful tool is a little shortsighted, it reflects a growing stance on using AI without acknowledgement. The expectation that authors disclose any meaningful use of AI is no longer a suggestion but standard. Guidelines from journals like Nature and all Springer Nature journals mandate that researchers state which tools were used, how they were used, and to what extent (Nature Editorial, 2023). Furthermore, the World Association of Medical Editors has reached a consensus that AI-generated content is acceptable only when authors retain full responsibility and disclose all contributions (Zielinski et al., 2024). Disclosing whether LLMs were used allows reviewers, editors, and readers to evaluate whether conclusions are based on sound evidence or unexamined machine output. Without transparency, AI-assisted research risks reducing science to a black box that undermines the rigor and repeatability of the scientific process.

The rise of AI threatens to degrade the quality of scientific output. While AI democratizes access to create substantial articles, it also enables the production of low quality or formulaic scientific writing. Similar to ”AI slop” seen on social media platforms such as Facebook, LinkedIn, or X, mass use of LLMs has led to a surge in “paper mill” activity, where mass-produced articles flood the peer review system without offering meaningful contributions to science (Liverpool, 2023). Publishers have reported that such manuscripts often evade plagiarism detection while still failing to meet standards of originality or relevance. This suggests a guideline be put by journals or publications on human researchers to verify LLM output and ensure scientific relevance and validity. Left unregulated, AI risks turning science into a flood of replication and noise.

4 Conclusion

The scientific research process is entering an era of change, driven by agentic AI systems that can reason, write, analyze, and experiment with growing autonomy. Across every stage of the workflow - from literature review to publishing - AI is already solving bottlenecks, offering new methods of discovery, and catalyzing the possibility of fully closed-loop research pipelines.

Multi-agent architectures now point toward systems capable of coordinating specialized AI assistants, allowing scientific production with minimal human intervention.

Yet, this evolution is not without its limits. Critical integration challenges remain across physical lab environments, cultural pushback, and practical hurdles. Equally important are the ethical consequences of delegating scientific tasks to AI: authorship, accountability, and scientific rigor must remain central. Without transparency and guidelines, the same technologies that accelerate science could undermine its credibility.

References

Afforai Inc. (2024). *Afforai reference manager*. https://afforai.com/reference-manager

Benchling & Anthropic. (2025). *Benchling accelerates scientific discovery with Claude in Amazon Bedrock*.

Benchling Inc. (2025). *AI at Benchling: Bring the power of AI to your R&D*. Benchling.

Boiko, D. A., MacKnight, R., Kline, B., et al. (2023). Autonomous chemical research with large language models. *Nature, 624*, 570–578. https://doi.org/10.1038/s41586-023-06924-6

Cactus Communications Services PteLtd. (2025). *Paperpal Preflight for Publishers*.

DeepL SE. (2025). *DeepL Write: AI-powered writing assistant*. https://www.deepl.com/write

Digital Life Science Solutions Team, Springer Nature. (2025). *Methods Muse: AI tools for protocol optimization*. Protocols.io.

Elicit. (2022). *Frequently asked questions*.

Elicit. (n.d.). *The AI Research Assistant*. https://elicit.org (Accessed July 31, 2025)

Fei, Y., Kumar, R. E., He, T., Kim, H., Jain, A., & Ceder, G. (2023). An autonomous laboratory for the accelerated synthesis of novel materials. *Nature, 624*, 86–91. https://doi.org/10.1038/s41586-023-06865-0

Flanagin, A., Kendall-Taylor, J., & Bibbins-Domingo, K. (2023). Guidance for authors, peer reviewers, and editors on use of AI, language models, and chatbots. *JAMA, 330*(8), 702–703. https://doi.org/10.1001/jama.2023.12539

Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Palepu, A., Sirkovic, P., Myaskovsky, A., Weissenberger, F., Rong, K., Tanno, R., Saab, K., Popovici, D., Blum, J., Zhang, F., Chou, K., Hassidim, A., Gokturk, B., Vahdat, A., Kohli, P., Matias, Y., Carroll, A., Kulkarni, K., Tomasev, N., Guan, Y., Dhillon, V., Vaishnav, E. D., Lee, B., Costa, T. R. D., Penadés, J. R., Peltz, G., Xu, Y., Pawlosky, A., Karthikesalingam, A., & Natarajan, V. (2025). Towards an AI co-scientist.

Grammarly Inc. (2025). *Grammarly: AI writing assistant*. https://www.grammarly.com

Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jupyter Project. (2025). *Jupyter AI documentation (Version 2)*. https://jupyter-ai.readthedocs.io/en/v2/

Khamsi, R. (2020). Coronavirus in context: Scite.ai tracks positive and negative citations for COVID-19 literature. *Nature, 581*, 130–131. https://doi.org/10.1038/d41586-020-01324-0

Kung, J. Y. (2023). Elicit. *Journal of the Canadian Health Libraries Association, 44*(1), 15–18. https://doi.org/10.29173/jchla29656

Kwon, D. (2025). Is it OK for AI to write science papers? Nature survey shows researchers are split. *Nature, 521*(7553), 423–424.

Langmead, B. (2021). *Langmead Lab introduces the first algorithm for AI-driven cloud labs*. Carnegie Mellon University. https://msas.cbd.cmu.edu/news/langmead-introduces.html

Liverpool, L. (2023). AI intensifies fight against 'paper mills' that churn out fake research. *Nature, 618*(7964), 222–223. https://doi.org/10.1038/d41586-023-01600-0

Meinke, A., Schoen, B., Scheurer, J., Balesni, M., Shah, R., & Hobbhahn, M. (2024). *Frontier models are capable of in-context scheming*. Apollo Research.

Nature Editorial. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature, 613*(7945), 612. https://doi.org/10.1038/d41586-023-00056-7

Nguyen, A. (2024). *Find novel research gaps in minutes with AI*. Afforai. https://afforai.com/blog/use-ai-find-research-gaps

Overleaf. (2025). *AI Assist*.

Perkel, J. M., & Van Noorden, R. (2020). TL;DR: This AI sums up research papers in a sentence. *Nature, 587*, 19–20. https://doi.org/10.1038/d41586-020-03187-3

Perplexity. (2025). *Perplexity.ai*. https://www.perplexity.ai

Science Suite Inc. (2025). *BioRender: Scientific image and illustration software*. https://www.biorender.com

Scite. (2025). *Scite.ai*. https://www.scite.ai

Shao, Y., Jiang, Y., Kanell, T. A., Xu, P., Khattab, O., & Lam, M. S. (2024). Assisting in writing Wikipedia-like articles from scratch with large language models. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature, 613*, 620–621. https://doi.org/10.1038/d41586-023-00107-z

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313. https://doi.org/10.1126/science.adg7879

Thorp, H. H. (2024). Genuine images in 2024. *Science, 383*(6678), 7. https://doi.org/10.1126/science.ado4567

Tobias, A. V., & Wahab, A. (2023). Autonomous 'self-driving' laboratories: A review of technology and policy implications. *Royal Society Open Science, 10*(5), 230646. https://doi.org/10.1098/rsos.230646

University of Cambridge. (2009, April 2). *Robot scientist becomes first machine to discover new scientific knowledge*. University of Cambridge Press Release.

Zhou, A., Wang, K., Lu, Z., Shi, W., Luo, S., Qin, Z., Lu, S., Jia, A., Song, L., Zhan, M., & Li, H. (2023). Solving challenging math word problems using GPT-4 code interpreter with code-based self-verification. *arXiv preprint arXiv:2307.12345*.

Zielinski, C., Winker, M. A., Aggarwal, R., Ferris, L. E., Heinemann, M., Lapeña, J. F., et al. (2024). Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. *Current Medical Research and Opinion, 40*(1), 11–13.https://doi.org/10.1080/03007995.2023.2345678

Cognitive Violence: The Neurological and Elevated Cancer Consequences of Environmental Racism in Communities of Color and Its Implications in Artificial Intelligence

Anvi Jah, Bhargavi Nigam, Sarah Kim

In neighborhoods bordering factories, highways, landfills, or in areas where police sirens are more frequent than birdsong, a saddening reality exists: environmental racism has not only damaged the body but the mind as well. Across the United States and beyond, communities of color are disproportionately exposed to neurotoxic pollutants in the environment around them, such as lead, mercury, and fine particulate matter, often a result of discriminatory zoning. The agents infiltrate not only the lungs and bloodstreams but also the delicate brain, impairing focus, cognitive development, and long-term neurological health. This form of harm, what some scholars call environmental violence, is a slow, systematic assault on the human mind. Yet in the age of technological advancements, artificial intelligence (AI) emerges as both a potential remedy and risk. This paper explores the current situation of environmental racism through various techniques to assist in solving the effects of environmental racism for the General Public.

Keywords: *Environmental racism; neurotoxic pollutants; cognitive health; environmental justice; artificial intelligence; public health*

1 Environmental Racism and its increase in cancer and other degenerative diseases

Systematic environmental racism significantly increases cancer risk for communities of color in the United States, as landmark studies continue to demonstrate. Decades of discriminatory housing policies, such as redlining, have forced marginalized populations into neighborhoods adjacent to highways, factories, and carcinogenic waste sites, where exposure to pollutants like benzene, arsenic, formaldehyde, and polycyclic aromatic hydrocarbons is much higher than in predominantly Caucasian communities (Kyrematen, 2025; Puckrein, 2024).

A recent report found that 56% of people living within 3 kilometers of carcinogenic-producing sites in the U.S. are people of color, with cities like Houston, Flint, and the infamous "Cancer Alley" in Louisiana serving as prime examples of these disparities (Kyrematen, 2025).

Airborne carcinogens and toxic chemicals in drinking water, such as lead, asbestos, and hazardous vehicular emissions, pervade these environments, damaging cellular structures and triggering mutations that can cause cancers of the lung, bladder, liver, and skin. The Environmental Protection Agency's (EPA) assessments have found that areas with dense concentrations of Black and Hispanic residents experience estimated cancer risks far exceeding those in mostly White communities. For example, in St. John the Baptist Parish within Cancer Alley, the risk of cancer was 5.5 times higher for black residents than elsewhere in Louisiana (EPA, 2024).

Carcinogens, when inhaled, ingested, or absorbed through the skin, initiate a cascade of molecular events that ultimately disrupt normal cellular function and division. Many environmental carcinogens seen in disproportionately financially disadvantaged communities, such as benzene, polycyclic aromatic hydrocarbons and arsenic, are metabolized by the liver, where they are often converted into even more reactive intermediates. These metabolites can form DNA adducts, directly binding to genetic material and causing base-pair mutations during cell replication. If key genes regulating cell growth and apoptosis, such as TP53 or BRCA1, are mutated, it can lead to uncontrolled proliferation - a hallmark of cancer (Lacayo, 2025). Persistent oxidative stress, incited by exposures to fine particulate matter and heavy metals, generates reactive oxygen species (ROS) that further damage DNA, proteins, and cell membranes, enhancing mutagenesis (Hurbain, 2024).

Additionally, chronic inflammation from recurrent pollutant exposure creates a microenvironment conducive to tumor development: cytokines and growth factors are released, promoting angiogenesis-the formation of new blood vessels- which further contribute to malignancy. The immune system, once responsible for eliminating abnormal cells, can become suppressed or dysregulated due to ongoing toxin intake, permitting mutated cells to evade detection and persist (Morello-Frosch et al., 2006).

In the finale, these exposures are aggravated by socioeconomic barriers, including minimal access to quality healthcare and preventive cancer screenings. Cancer chances are dominant in disproportionately disadvantaged communities due to the environment.

2 Environmental Racism and Neurological Harm: A Legacy of Toxins in Marginalized Communities

2.1 Background

Environmental racism isn't just a historical neglect; it's an ongoing issue that significantly impacts the neurological health of communities of color. Marginalized communities tend to suffer through disproportionate exposure to environmental hazards due to discriminatory policies, zoning laws, environmental racism, and more. This results in a higher exposure to neurotoxic pollutants such as lead, mercury, arsenic, carcinogenic chemicals, and any other hazardous materials. These toxins don't just break down by themselves and disappear; they enter the human body and brain leading to long-term results in cognitive and neurological damage. Organizations such as The Lancet Commission on Pollution and Health and the U.S. EPA found that environmental pollution kills around nine million people through increased neurological disorder rates and mental health challenges worldwide annually which causes significant threats in human societies. In addition, the situation becomes worse as climate change accelerates, since we'll witness more extreme weather patterns, frequent and intense disasters, and consequent environmental justice and health issues, which are even more burdened as industrial toxicants and hazards are released. In other words, environmental racism contributes to neurological harm through exposure to heavy carcinogens and hazardous materials, the developmental impact on marginalized communities, and the broader implications for public health and racial equity.

2.2 Current Environmental Conditions and Neurotoxic Exposure

Across the United States children in low-income, mostly Black or colored neighborhoods, are more likely to live near highways, power plants, waste incinerators, or factories, all unaware that they emit dangerous neurotoxic substances. For example, lead, once widely used in paint and plumbing, continues to remain a major threat in older housing stock, especially in communities that were denied investment and repairs. The National Health and Nutrition Examination Survey conducted a secondary analysis of blood Pb determinations for 2841 US children at ages 1-5 years and experimented that black children in risk factors such as polluted areas had an adjusted +0.73 to 1.41 blood Pb and a 1.8 to 5.6 times higher odds of having an EBLL. They ended up concluding that Black children are more likely than white children to have elevated blood lead levels, which is known to reduce IQ, hinder memory, and increase neurological disorders such as ADHD. In addition, PM2.5, known as a harmful airborne particle, penetrates deep into the lungs and brain, leading to inflammation, impaired cognitive development, and reduced gray matter (NHANES, 2020).

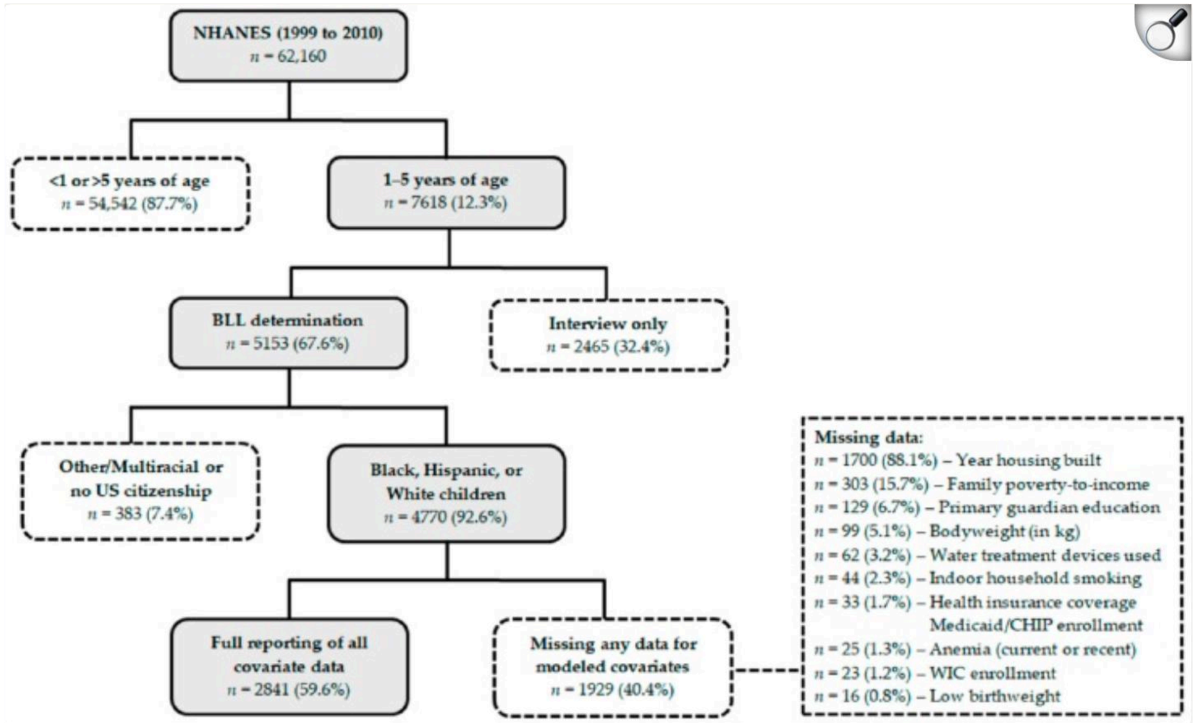


Figure 1: Flow diagram of the NHANES 1999–2010 sample selection process for children aged 1–5 years, showing exclusions, blood lead level (BLL) determinations, and final analytic cohort. Missing data sources are detailed, with housing age, poverty status, and guardian education being the most common gaps (NHANES, 2020).

Meanwhile, carcinogens such as benzene and formaldehyde are released from petroleum refineries, plastic plants, and landfills, which poses a serious threat to the brain. Not only do they increase cancer risk, but they also disrupt the blood-brain barrier and impair brain function. Benzene, for example, has been linked to decreased white matter volume, mood disorders, and long-term neurological degeneration. Children living near industrial areas suffer the most since their developing brains are way more vulnerable to toxins. In addition, the increasing effect of environmental pollutants is linked with delayed speech, poor academic performance, and even increased need for special education services. These harms are all due to the consequences of environmental racism which includes the lack of access to quality healthcare, clean air, and safe housing.

2.3 Where the System Fails

Despite decades of environmental regulations, enforcement is often weaker in communities of color. Studies have shown that polluting facilities are more likely to be sited near Black and colored neighborhoods, which raises concerns about the health effects of disproportionate exposure to environmental burdens. Moreover, scientific studies documenting these harms often exclude race as a key component of the issues of rising neurological problems, contributing to a huge burden on marginalized populations. While many environmental and health monitoring systems rely on aggregated data, they often fail to capture the specific ways that exposure accumulates across the lifespan. Additionally, the classification of certain substances as "carcinogenic" but not "neurotoxic" can obscure the cognitive damage they inflict, making it way more difficult for affected discriminated communities to demand justice, equality, or intervention. Without targeted race-conscious research and policy implementation, the neurological consequences of environmental racism will persist and continue to grow along with the cycle of disinvestment, illness, and inequality.

2.4 Future Notes

Moving forward, addressing environmental racism consequences as a public health crisis requires a multi-disciplinary approach that includes stronger environmental regulations, equitable urban planning, and racially disaggregated data in health and pollution research. Community-driven random data collection and citizen science initiatives can help illuminate the hidden harms in under-resourced neighborhoods. Moreover, public investment in lead remediation, air filtration, strong infrastructure, and access to neurological healthcare is essential to breaking the link between geography, race, and brain health. Recognizing environmental racism not just as a civil rights issue, but as a neurological health risk factor, is a necessary step toward justice and collective well-being.

3 AI in Medicine: The Power and Perils of Data-Driven Diagnosis

3.1 Background

Artificial Intelligence (AI) is not merely a trend in today's society but a tool and necessity. AI refers to the simulation of human intelligence in machines that are trained to learn and reason like humans, including the ability to analyze data, recognize patterns, and more. This tool has proved to be instrumental in spotting disparities in environmental health and its intersection with neurology. To provide a more concrete definition: "AI… is defined as the broad science of mimicking human abilities, while machine learning (ML) is defined as a set of algorithms that is fed with structured data in order to complete a task without being programmed how to do so (Haenlein and Kaplan, 2019)" (Aschner and others, 2022). This technology can be instrumental or detrimental in efforts to mitigate neurological inequalities in society. In the field of medicine, AI has the potential to either abate or exacerbate existing inequalities, depending on how it is trained and deployed. Environmental racism - a key implication of this paper - describes situations where communities of color are exposed to greater environmental risks than others as a result of specific policies, regulations, and laws that place more environmental burdens on marginalized groups. This section aims to provide an overview of the impact AI (specifically, AI predictive modeling, Large Language Models, and AI imaging algorithms) has had on modern medicine/research, its racial concerns, limitations, and prospects.

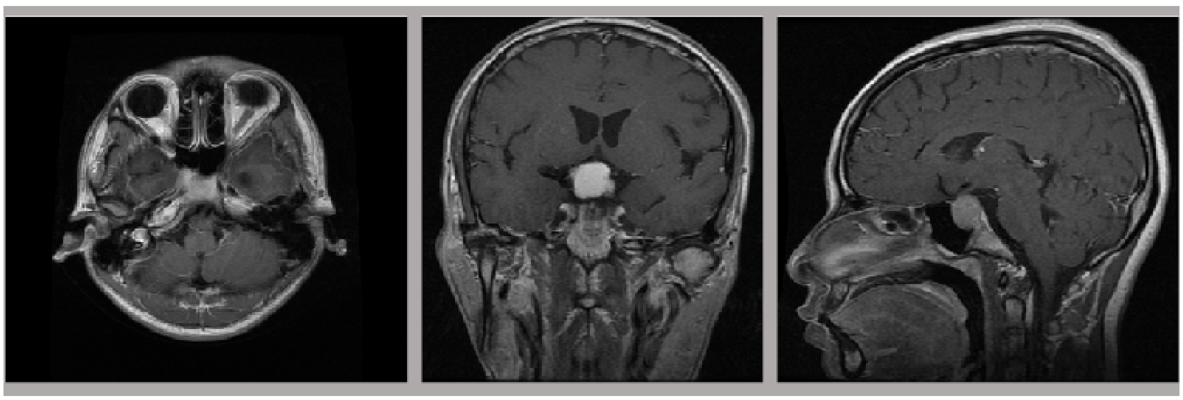


Figure 1: The axial, coronal, and sagittal view respectively from a sample of MRI images (Modaresnia, 2023).

3.2 Current Applications in Medicine

Today, AI is rapidly accelerating drug discovery, predicting how immune cells (T cells) respond to tumors, and improving immunotherapy. AI finds patterns in large biological datasets to map drug response pathways and is also increasingly successful in facilitating precision treatment (NCI, 2024). It essentially expedites genetic subtyping of brain tumor tissue during surgery, speeding up decisions and predicting survival outcomes for patients with breast cancer using digital pathology images. For instance, in one study, researchers successfully detected brain tumors in MRI scans using a large collection of brain tumor images by demonstrating that fine-tuning a YOLOv7 model through transfer learning significantly improved its performance in detecting gliomas, meningioma, and pituitary brain tumors, reaching up to a staggering 99.5% accuracy (Abdusalomov, 2023). AI is now being trained in neuroscience and cancer care to detect tumors from MRI scans, assisting in treatment planning, and predicting cognitive functions. And yet, studies show that such models perform worse on patients of color because of data that underrepresents such marginalized populations - medical gaps that deepen the existing health disparities.

3.3 Where AI Falls Short

Despite these advances in new technology, biases still exist. Because AI models in medicine depend heavily on large, diverse, and well-annotated datasets, if the data used to train these models is not appropriately diverse and representative of the population, these models can pose a dangerous medical bias: even the most advanced models risk bias, misdiagnosis, and reduced generalizability, ultimately providing inaccurate information for minorities and women. An AI model trained mostly on data representing white males may not effectively detect early signs of certain cancers in women of color, for instance. According to Anne Trafton at MIT News, MIT researchers have found that artificial intelligence models that are most accurate at predicting race and gender from X-ray images also show the biggest "fairness gaps," i.e., discrepancies in their ability to accurately diagnose images of people of different races or genders (Trafton, 2024). Alarmingly, these findings suggest that models may be using demographic shortcuts when making their diagnostic evaluations, which ultimately lead to incorrect results for women, African Americans, and other groups, the researchers say (Trafton, 2024).

This, thus, is a highly unfair route AI models are taking.

However, no model is perfect and all come with their unique challenges: "Large and granular datasets are needed to develop ML models and get accurate predictions. Little data results in a poor approximation and may cause over-fitting. High-throughput "omics" technologies which are increasingly used to measure thousands of variables (e.g. metabolite levels, gene expression, or image acquisitions) are thus very suitable to develop ML algorithms. They can be used to identify harmful substances. These models are incredibly good at predicting diseases, MIT scientists say, but during training learn to predict other aspects that may not be desirable.

A consequence of AI models, as MIT researchers note, is that when models trained on patients from one hospital are found to be biased, researchers attempt to retrain them to improve fairness. However, debiasing works best when the test and training patients are similar, i.e., from the same hospital. The fairness gaps reappear when models are applied to patients at different hospitals, so in that sense, medical AI models are still limited. This is worrisome because in many cases, hospitals use models that have been developed on data from other hospitals, especially in cases where an off-the-shelf model is purchased, the researchers say (MIT, 2024). Researchers would always need to evaluate external models on their data, knowing that fairness guarantees may not transfer between populations, and if enough data is available, training models on their own data would fetch better results.

MIT researchers further explored the *why and how* behind machine learning models worsening existing inequities in medical diagnosis and treatment. Led by Professor Marzyeh Ghassemi, these researchers identified four types of subpopulation shifts - differences in the way machine learning models perform for one subgroup as compared to another - that cause these disparities (Nadis, 2023). These biases stem from either "class", "attribute," or both.

They identified 4 main types of this shift:

A. **Spurious Correlations:** There is a bias in both the class and the attribute.
   The "Camels and Cows" example: Take an ML model that sorts images of animals into two classes: cows and camels. Attributes are descriptors not specific to the class, like the animal's background. If all training images show cows on grass and camels on sand, the model might erroneously assume that cows are only found on grass and camels are only on sand.
B. **Attribute Imbalance**: If the dataset used for training has a significant attribute imbalance. For instance, if 100 males are diagnosed with pneumonia for every one female, the model would likely perform better at detecting pneumonia in men than in women.
C. **Class Imbalance**: If there are significantly more healthy subjects than sick ones, the model would be biased toward healthy cases.
D. **Attribute Generalization**: If a sample contained 100 male patients with pneumonia but zero female subjects with the illness, the model should ideally still be able to generalize and make predictions for female subjects despite the lack of training data for that specific subgroup.

While improving the classifier (the final layer of the neural network) can reduce spurious correlations and class imbalance, and improving the "encoder" (an uppermost layer) can mitigate attribute imbalance, attribute generalization remains an unresolved issue that researchers are unsure how to fix (Nadis, 2023). Furthermore, the commonly used metric for evaluating fairness, "worst-group accuracy" (WGA), has a surprising drawback. WGA measures the accuracy of a model on a subgroup that performs the worst compared to the others. Ideally it is to ensure that no one group is disproportionately disadvantaged as WGA is based on the assumption that if you improve the accuracy, you improve the model as a whole (Nadis,2023). However, researchers note that boosting WGA can lead to a decrease in "worst-case precision", the scenario where precision is at its lowest possible point. This isn't desirable because both accuracy (validity of findings) and precision (reliability of methodology) are crucial in medical diagnostics and ideally should not be traded for one another, researchers say (Nadis, 2023).

4 Conclusion

Though researchers today acknowledge that achieving fairness in healthcare among all populations is the goal, achieving this requires a more nuanced understanding of the sources of fairness and how they affect our current systems. This understanding must be established before fully implementing these models. Furthermore, there is also a growing need for randomized clinical trials to validate AI's practice in clinical practice (NCI, 2024). Without broadly accepted and adopted standards for the development of AI and machine learning, it will be difficult to ensure reproducibility and medical fairness for marginalized communities overall.

References

Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. *Cancers, 15*(16), 4172. https://doi.org/10.3390/cancers15164172

AI and Cancer – NCI. (2024, May 30). *National Cancer Institute*. https://www.cancer.gov/research/infrastructure/artificial-intelligence#top

Aschner, M., Mesnage, R., Docea, A. O., Paoliello, M. M. B., Tsatsakis, A., Giannakakis, G., Papadakis, G. Z., Vinceti, S. R., Santamaria, A., Skalny, A. V., & Tinkov, A. A. (2022). Leveraging artificial intelligence to advance the understanding of chemical neurotoxicity. *NeuroToxicology, 89*, 9–11. https://doi.org/10.1016/j.neuro.2021.12.007

Trafton, A. (2024, June 28). Study reveals why AI models that analyze medical images can be biased. *MIT News*. https://news.mit.edu/2024/study-reveals-why-ai-analyzed-medical-images-can-be-biased-0628

Augusta University. (n.d.). Environmental racism: Definition, examples and prevention. *Augusta University Online*. https://www.augusta.edu/online/blog/environmental-racism

Nadis, S. (2023, August 17). How machine-learning models can amplify inequities in medical diagnosis and treatment. *MIT News*. https://news.mit.edu/2023/how-machine-learning-models-can-amplify-inequities-medical-diagnosis-treatment-0817

Malin, S. A., & Ryder, S. S. (2018). Developing deeply intersectional environmental justice scholarship. *Environmental Sociology, 4*(1), 1–7. https://doi.org/10.1080/23251042.2018.1446711

Yeter, D., Karwowski, M. P., Mazumdar, M., Reiss, R., Ceballos, R. M., & Bellinger, D. C. (2020). Disparity in risk factor severity for early childhood blood lead among predominantly African-American Black children: The 1999 to 2010 US NHANES. *International Journal of Environmental Research and Public Health, 17*(5), 1552. https://doi.org/10.3390/ijerph17051552

Lane, M., Wu, C., Johnson, R., & Patel, K. (2025). The effects of air pollution on neurological diseases: A narrative review on causes and mechanisms. *Toxics, 13*(3), 207. https://doi.org/10.3390/toxics13030207

Fu, P., Guo, X., Cheung, F. M. H., & Yung, K. K. L. (2019). The association between PM2.5 exposure and neurological disorders: A systematic review and meta-analysis. *Science of the Total Environment, 655*, 1240–1248. https://doi.org/10.1016/j.scitotenv.2018.11.218

Hu, J., Xu, W., Zhang, Y., & Liu, J. (2021). Changes in cognitive function and related brain regions in chronic benzene poisoning: A case report. *Annals of Translational Medicine, 9*(1), 81. https://doi.org/10.21037/atm-20-6597

Racial and socioeconomic disparities in residential proximity to polluting industrial facilities: Evidence from the Americans' Changing Lives study. (2009). *American Journal of Public Health, 99*(S3), S649–S656. https://doi.org/10.2105/ajph.2007.131383

McHale, C. M., Zhang, L., Hubbard, A. E., & Smith, M. T. (2018). Assessing health risks from multiple environmental stressors: Moving from G×E to I×E. *Mutation Research Reviews, 775*, 11–20. https://doi.org/10.1016/j.mrrev.2017.11.003

Lacayo, V. M. (2025, April 26). The environmental risk for cancer in minoritized communities. *National Minority Quality Forum*. https://nmqf.org/resource-library/the-environmental-risk-for-cancer-in-minoritized-communities/

Kyerematen, A. (2025, June 27). People of color face disproportionately higher risk of cancer from environmental toxins. *National Minority Quality Forum*. https://nmqf.org/people-of-color-face-disproportionately-higher-risk-of-cancer-from-environmental-toxins-new-study-finds/

Hurbain, P. (2024, April 12). Regional racial economic disparities in air pollution-related cancer risk may be improving but still persist. *Desert Research Institute*. https://ascopost.com/news/december-2024/regional-racial-economic-disparities-in-air-pollution-related-cancer-risk-may-be-improving-but-still-persist/

Morello-Frosch, R., Pastor, M., Sadd, J., & Shonkoff, S. B. (2006). Residential segregation and estimated cancer risks associated with ambient air toxics in metropolitan areas of the United States. *Environmental Health Perspectives, 114*(3), 386–393. https://doi.org/10.1289/ehp.8500

Puckrein, G. A. (2024, July 18). Minority communities face higher environmental risks for cancer: Report. *Shift Cancer*. https://shiftcancer.org/the-hill-minority-communities-face-higher-environmental-risks-for-cancer-report/

Environmental Protection Agency. (2024, September 9). A horror story – How the law simultaneously addresses and facilitates environmental injustice through environmental racism. *Global Network for Human Rights and the Environment*. https://gnhre.org/?p=18198

Modaresnia, Y., Abedinzadeh Torghabeh, F., & Hosseini, S. A. (2023). EfficientNetB0's hybrid approach for brain tumor classification from MRI images using deep learning and bagging trees. In *2023 13th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 234–239). IEEE. https://doi.org/10.1109/ICCKE60553.2023.10326290

AI for Business Operations: Streamlining Efficiency and Unlocking New Value

Krish Wasan

This essay explores the transformative role of artificial intelligence in modern business operations. From process automation and predictive analytics to human resources and marketing personalization, AI is shown as both a driver of efficiency and a source of ethical challenges. The discussion highlights key applications, industry examples, and future implications for organizations adopting AI responsibly.

Keywords: *artificial intelligence; business operations; intelligent process automation; predictive analytics; customer service automation*

In the modern economy, businesses are under increasing pressure to operate more efficiently, cost-effectively, and strategically. Once a niche technology reserved for large tech corporations and research labs, artificial intelligence (AI) is now a disruptive force deeply rooted in daily operations across various industries. Businesses are finding new potential for development, innovation, and competitive difference, in addition to cost savings, when they incorporate AI into their operations, from human resources to logistics.

Intelligent process automation is one of the most important ways that AI is used in business. Intelligent automation systems that can handle unstructured data, learn from human behavior, and make judgments have replaced traditional robotic process automation (RPA) instruments that imitate repetitive operations. According to Deloitte's 2020 Global RPA Survey, 78% of companies currently using RPA anticipate greatly increasing their investment in AI-driven automation in the upcoming years (Deloitte, 2020). AI may be trained to reduce human mistakes and manual work by reviewing financial data, detecting fraud, answering consumer inquiries, and even performing performance analysis.

One of the most powerful business applications of artificial intelligence lies in the realm of forecasting and planning. AI has also shown itself to be invaluable in the field of predictive analytics. Massive amounts of data are sorted through by AI systems to find patterns and predict future trends. For instance, Amazon employs predictive analytics to manage its supply chain with remarkable accuracy, anticipating demand and modifying inventory before the need ever materializes. It also utilizes predictive analytics to recommend products to customers.

Another critical area where AI is transforming business practices is in customer-facing operations. AI has revolutionized operations that interact with customers. Millions of customer service inquiries are being handled every day by AI-powered chatbots and virtual agents in industries like banking, retail, and telecommunications. Natural language processing (NLP) gives these bots the ability to comprehend and react to human language with ever-increasing complexity. Gartner predicts that by 2027, chatbots will become the primary customer service channel for 25% of organizations, reducing wait times, labor costs, and human workload (Gartner, 2023).

Perhaps the most visible change for consumers has come in the personalization of products and services. AI has significantly changed marketing and personalization as well. AI technologies can provide hyper-personalized experiences that increase conversion rates and customer loyalty by evaluating real-time consumer data. For instance, Netflix uses AI to personalize content recommendations, resulting in over 80% of its watched content being algorithmically suggested (Medium, 2020). Businesses can stand out in crowded markets and better engage customers with this kind of data-driven customization.

Beyond customer-facing operations, AI is changing the internal dynamics of the workplace. Human resources has become a major testing ground, with companies like Hilton incorporating AI into their recruitment procedures. Companies such as Hilton use AI in hiring procedures by employing tools like HireVue, which utilize machine learning and facial recognition to examine a candidate's speech patterns, facial expressions, and word choice in order to assess their suitability for a position. Critics warn about relying too heavily on opaque algorithms, yet this simplifies recruiting and helps eliminate unconscious bias.

However, there are also significant operational and ethical issues with the use of AI in business. To prevent unforeseen harm, concerns about algorithmic bias, data privacy, and transparency must be addressed. According to McKinsey, organizations that empower employees with AI tools see productivity gains of up to 40% in operational workflows (McKinsey, 2024). Additionally, business executives are increasingly in need of AI literacy. Organizations run the risk of using AI tools that are harmful or ineffective if they lack the necessary knowledge.

Ultimately, artificial intelligence is a strategic enabler rather than merely an efficiency tool. Companies that carefully consider their AI investments, maintain human control, and guarantee accountability stand to gain a great deal. AI is positioned as a key component of the enterprise of the future due to its capacity for data analysis, ongoing learning, and operational optimization.

AI's place in corporate operations will change from experimental to crucial as it continues to develop further. The next wave of digital change will be led by those who embrace AI now, guided by ethics and a clear mission.

References

Deloitte. (2020). *Global RPA survey: The robots are waiting*. Deloitte. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Technology/gx-consulting-robots-are-ready.pdf

Gartner. (2023). Gartner predicts 25% of organizations will use chatbots as primary customer service channel by 2027. *Smart Customer Service*. https://www.smartcustomerservice.com/Articles/News-Features/Chatbots-Will-Be-a-Primary-Customer-Service-Channel-by-2027-Gartner-Predicts-154129.aspx

McKinsey. (2024). *Superagency in the workplace: Empowering people to unlock AI's full potential at work*. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work

Medium. (2020). Netflix's recommendation algorithm. *Medium*. https://medium.com/swlh/netflix-and-the-recommendation-system-e806f062ba74

Grading Tech Companies On Sustainable Practices

Neil Plant

Technology companies are central to innovation, yet their rapid growth comes with significant environmental costs. This paper evaluates the sustainability practices of three major players—OpenAI, Google, and Vercel—using a rubric that measures performance in renewable energy adoption, transparency, mitigation of product impact, provision of low-carbon tools, and circular practices. Each company is graded on a 0–4 scale across these five categories. The analysis finds that Google demonstrates industry leadership with strong commitments to renewable energy and transparency, while OpenAI and Vercel lag behind in key areas, particularly transparency and circular practices. The study highlights both promising practices and urgent gaps, underscoring the need for greater accountability as computing demands increase.

Keywords: *Sustainability; renewable energy; transparency; low-carbon tools; circular practices; technology companies*

1 Introduction

The influence of large technology companies extends well beyond innovation in software and hardware. These companies now operate some of the world's largest data centers, manufacture millions of devices, and power global digital infrastructures. With such scale, their environmental footprint is immense. Data centers consume vast amounts of electricity, often derived from non-renewable sources, while hardware manufacturing and disposal contribute to e-waste and resource depletion. As artificial intelligence, cloud computing, and edge networks expand, questions about the sustainability of these practices are becoming increasingly urgent.

Despite this, the sustainability efforts of technology companies vary widely. Some, like Google, have made strong commitments to renewable energy and environmental transparency, while others, like OpenAI and Vercel, lag behind in reporting or direct accountability. To fairly evaluate these companies, I developed a rubric that assigns points across five categories: renewable energy use, transparency, mitigation of environmental impact, low-carbon developer tools, and circular practices. This approach provides a structured way to compare companies that differ in size, focus, and products.

In this paper, I apply the rubric to three companies. OpenAI represents a cutting-edge AI research firm with global influence but little visibility on sustainability. Google demonstrates how a global tech giant can commit to measurable environmental goals. Vercel, a newer player, shows how smaller companies balance efficiency and growth while lacking the scale of sustainability programs. By analyzing these cases side by side, we can better understand both progress and shortcomings in the industry's sustainability journey.

2 Rubric for Grading on Sustainable Practices

To evaluate sustainability practices, I designed a rubric that measures company performance in five key categories: use of renewable energy, transparency of environmental reporting, mitigation of product impact, availability of low-carbon tools, and adoption of circular practices. Each category is scored from 0 (no evidence of sustainable practice) to 4 (comprehensive, industry-leading implementation). The full rubric is shown in Table 1.

Table 1. *Rubric for grading tech companies on sustainable practices.*

Criterion	0 points	1 point	2 points	3 points	4 points
C1: Renewable Energy	No, or <2% of total usage	Yes, but <5% of total usage	Yes, at least 15% of total usage	Yes, at least 30% of total usage	Yes, at least half of total usage
C2: Transparency	No, minimal or no data published	Limited documents a year	Partial reports, no full disclosure	Full reports published	Full reports + formal sustainability policy with accountability
C3: Mitigation of Product Impact	No effort; standard infrastructure	Minimal attention in select products	At least half of products designed with low-carbo	Majority of products on green infrastructure, some impact reporting	Almost all products on green methods + explicit discussion of impact

			n methods		
C4: Low-Carbon Tools	No tools provided	Minimal developer tools	Some optimization tools provided	Strong toolkit with metrics and guides	Comprehensive toolkit, widely accessible and encouraged
C5: Circular Practices	No reuse/recycling	Minimal reusing/recycling	E-waste recycled, minimal reuse	Hardware reused/refurbished	Upgradable/repairable hardware + comprehensive reuse/recycling

3 Companies

To apply the rubric in practice, I evaluated three technology companies - OpenAI, Google, and Vercel - each representing a different role in the tech ecosystem. OpenAI is a cutting-edge artificial intelligence research lab whose rapid growth raises questions about energy usage and transparency. Google, one of the largest and most established tech companies in the world, has made highly visible commitments to sustainability and serves as an industry benchmark. Vercel, a smaller but rapidly growing platform for frontend development, reflects how newer companies address efficiency and environmental responsibility without the scale of major industry players. Assessing these three companies side by side illustrates how different organizational models lead to varying approaches to sustainability and highlights where meaningful progress is being made and where gaps remain.

3.1 OpenAI

OpenAI is a company that has made major advancements in AI. However, they are not very transparent on their environmental impact, and they haven't released any formal reports on their emissions or energy usage. Therefore, it is difficult to grade them on the rest of the categories. They run their cloud computing on Microsoft Azure, and Azure has pledged to go 100% renewable by 2025 (Microsoft, n.d.). Since we don't know how much energy would have been used by OpenAI's cloud computing otherwise, we will assume about 20%, even though this reduction is through no effort of their own. OpenAI also has some optimized models that use less energy; however, this is framed around the user experience more than environmental impact. It is unclear whether OpenAI uses circular practices.

OpenAI received a score of 2 for renewable energy, 0 for transparency, 1 for mitigation of product impact, 1 for low-carbon tools, and 1 for circular practices, giving them a total of 5 out of 20, which is considered poor. OpenAI is at the forefront of AI; however, they are not very transparent about their impact. It is clear they are not doing enough for sustainability, and people should start asking them to.

3.2 Google

Google is one of the better big tech companies when it comes to sustainability. They are transparent about their impact and do release sustainability reports, and have a policy when it comes to the environment (Corio, 2022; Google, 2024; *Sustainability Reports & Case Studies - Google Sustainability*, 2025). They purchase enough green energy to offset their entire electricity usage and have also set a goal to use 100% carbon-free energy by 2030. They have made significant progress in cutting the emissions of their data centers; however, they have not been as transparent about the cost of training AI models. Google also has a suite of tools to help developers lower their emissions, and they help guide them to these tools as well (Google Cloud, n.d.). They also use circular practices well, designing their servers with refurbishment in mind; however, the hardware isn't repairable across the board.

Google received a score of 5 for renewable energy, 5 for transparency, 4 for mitigation of product impact, 5 for low-carbon tools, and 4 for circular practices, giving them a total of 18 out of 20, which is considered excellent. Google is one of the leaders in sustainability for the tech industry. While there is still room for them to grow, they are setting a superb example for what it means to care about the environment. It is important that they keep moving forward and hold themselves accountable for the goals they have set.

3.3 Vercel

Vercel is a platform for frontend development. They are known for their serverless architecture and edge network, which enables fast delivery (Fishtank Consulting, 2023). However, they are lacking in their sustainable practices. They are not transparent about their impact, citing their cloud providers instead, so their emissions data is largely unknown, and it is unclear how much of their infrastructure is on renewables (Ko, 2023). Their serverless architecture and edge computing are more efficient methods than traditional ones; however, they do not discuss the environmental impact of their services in detail. They provide tools for developers to reduce emissions, such as performance optimization and an efficient CDN, which reduces resource usage. However, these methods are mostly indirect, and they don't have a toolkit specifically framed for sustainable development. They don't produce hardware as they are cloud-based; however, they don't take initiatives for circular practices for the hardware they do use.

Vercel received a score of 2 for renewable energy, 1 for transparency, 3 for mitigation of product impact, 2 for low-carbon tools, and 0 for circular practices, giving them a total of 8 out of 20, which is considered poor. While Vercel uses architecture that reduces emissions, they do not engage in meaningful transparency or sustainability practices. They could definitely do better with relatively simple steps, but they haven't done that yet.

References

Corio, A. P. (2022). *5 years of 100% renewable energy, and targeting 100% CFE.* Google Cloud Blog. https://cloud.google.com/blog/topics/sustainability/5-years-of-100-percent-renewable-energy

Fishtank Consulting. (2023). *What is Vercel and Why You Should Use It?* GetFishtank.com. https://www.getfishtank.com/insights/what-is-vercel

Google. (2024). *Operating sustainably – Google Data Centers.* Google Data Centers. https://datacenters.google/operating-sustainably/

Hosanagar, K. (2025, January 6). *Using ChatGPT Responsibly: A Guide for the Climate-Conscious.* Creative Intelligence | Substack. https://hosanagar.substack.com/p/using-chatgpt-responsibly-a-guide

Ko, S. (2023). *What is Vercel's Green Energy Policy?* Vercel.com. https://vercel.com/guides/what-is-vercel-green-energy-policy

Microsoft. (n.d.). *Azure OpenAI Service – Advanced Language Models.* Azure. https://azure.microsoft.com/en-us/products/ai-services/openai-service

OpenAI Sustainability Report. (2025). *OpenAI Sustainability Profile.* DitchCarbon.com. https://ditchcarbon.com/organizations/openai

Sustainability Reports & Case Studies – Google Sustainability. (2025). *Google Sustainability*. https://sustainability.google/google-2025-environmental-report/

What is Active Assist | Recommender Documentation. (n.d.). *Google Cloud*. https://cloud.google.com/recommender/docs/whatis-activeassist

The Ethics of Using AI Within Computer Science

Haru Krause

The rapid integration of artificial intelligence into computer science has transformed programming practices, introducing both opportunities and risks. While AI excels at repetitive coding tasks and enables new approaches such as "vibe coding," it remains limited in creativity, accountability, and security. Studies show that AI-assisted programmers often generate less secure code and face greater risks of perpetuating bias due to flaws in training data. This paper explores the ethical implications of these developments, focusing on issues of accountability, security, bias, and the erosion of human expertise. Drawing on industry examples and recent research, the analysis argues that AI in computer science should be treated as a tool to augment, not replace, human judgment. Responsible adoption requires addressing bias, enforcing accountability, and ensuring that human creativity and critical thinking remain central to the field.

Keywords: *AI in computer science; programming ethics; accountability; bias; security risks*

In this day and age, AI is becoming more common in the tech industry. AI is becoming more of a focus for many tech companies, which often implement AI in some way in their applications. While AI has become increasingly involved and connected with technology, it's highly unlikely it will actually find its footing to replace jobs such as programmers or software engineers. AI might be useful for overly repetitive tasks, it cannot ever truly take over software engineering due to the fact that those engineering tasks still remain human at the core. As an article "AI excels in repetitive tasks, but software development is often more about creativity and problem-solving. It's like comparing a paint-by-numbers kit to an artist creating a masterpiece—both involve painting, but only one truly involves artistry." Carnegie Mellon University states (Will AI make software engineers obsolete?, n.d.). This case in the software industry can be compared to other industries where people can get concerned about the usage of AI generated creations over the work of humans. So where does this lay the need for AI ethics in tech fields?

AI is being increasingly used by many people, including programmers who have started to use it to write their code for them. There is a new phenomenon known as "vibe coding", where programmers are now using purely LLM models to do their work, staying almost in a conversational loop with the AI while prioritizing creative output instead, as said by the creator of the term, Andrej Karpathy in an article from the Times of India (Desk, 2025). While it can be a useful tool for checking while programming, ultimately it leads to an increased lack of software security and accountability. According to a Stanford study from 2022, participants who wrote code with an AI assistant wrote significantly less secure code than those who trusted the AI less (Perry, Srivastava, Kumar, & Boneh, 2023). While the security risks of using AI-written code is rather well-known and considerable, it should be also noted often using AI also creates a lack of accountability on the programmer/computer scientist. If programmers write bad code, they can simply blame it on the AI having that specific response, and thus being able to promote harm, misinformation, or bias possibly present in the AI's response.

It should be noted that the AI typically used in "vibe coding" and the like are usually the ML algorithms, "dynamic algorithms—that learn and evolve by interacting with the environment, usually classified as ML algorithms" (Kazim & Koshiyama, 2021). These ML algorithms are trained off data, data taken from humans that can be flawed, subject to biases. According to a report from the National Institute of Standards and Technology, AI can 'inherit' any bias programming or data structures that the machine learns from, giving another issue to using AI for programming and the like (Boutin, 2025). Apart from the security, accountability, and ethical concerns, it should be noted that AI can be detrimental to people's skills on their own.

When people use AI for coding or computer science blindly, they are not retaining or using any actual knowledge that they possess, but instead relying on these flawed models. AI-assisted programmers would not be able to as easily understand the errors in their code, and AI-assisted code leads to more errors in general (as shown in the same Stanford study as above), than if the programmer took the AI suggestions more in stride and instead relied on it much more sparingly.

In conclusion, the increasing usage of AI in coding/programming and computer science brings up discussion and concerns about security, the amount of errors present in the code, any biases/harm, and accountability. The increased usage of AI for computer science is an ethical issue that's to be examined carefully as proven in this article.

References

Boutin, C. (2025, February 3). *There's more to AI bias than biased data, NIST report highlights*. National Institute of Standards and Technology. Retrieved from https://www.nist.gov/news-events/news/2022/03/theres-more-ai-bias-biased-data-nist-report-highlights

Desk, T. T. (2025, March 2). *What is 'vibe coding'? Former Tesla AI director Andrej Karpathy defines a new era in AI-driven development*. *The Times of India*. Retrieved from https://timesofindia.indiatimes.com/technology/tech-news/what-is-vibe-coding-former-tesla-ai-director-andrej-karpathy-defines-a-new-era-in-ai-driven-development/articleshow/118659724.cms

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns, 2*(9), 100314. https://doi.org/10.1016/j.patter.2021.100314

Perry, N., Srivastava, M., Kumar, D., & Boneh, D. (2023). Do users write more insecure code with AI assistants? In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security* (pp. 2785–2799). https://doi.org/10.1145/3576915.3623157

Will AI make software engineers obsolete? Here's the reality. (n.d.). *Carnegie Mellon University Bootcamps Blog*. Retrieved from https://bootcamps.cs.cmu.edu/blog/will-ai-replace-software-engineers-reality-check

Artificial General Intelligence: An Interesting Supposition

Samin Hossain

A new type of Artificial Intelligence is emerging - one that is trying to combat the shortcomings of current narrow AI systems. Referred to as "Artificial General Intelligence", this concept is a vision where AI could be as adaptable and able to perform general tasks like humans. This new technology is only a theory currently, but could be a reality in as little as a few years, or may not even be achievable at all. This new field of AI, if achieved, could change the way that we live, either by creating a prosperous future or one that could lead to the extinction of humans entirely. This article's purpose is to identify exactly what AGI is, what its impacts will be, and if it even is possible to achieve.

Keywords: *Artificial General Intelligence, machine intelligence, ethics, future of AI*

The majority of people under the age of 30 use AI tools daily, and this number seems to be getting bigger every year (Beshay, 2025). The push of generative AI has been well-documented and widespread. Before, it was a tool to be played around with; now it has evolved to be a major part of people's work, school, and personal lives. Technologies commonly associated with this substantial rise include chatbots and AI image generators. These technologies, however, are classified as Artificial Narrow Intelligence or ANI. The main issue with ANIs is that they are coded to perform well in one specific task. As a result, their capabilities are limited when compared to human intelligence (Sowri et al., 2024). However, as is often the case with many technologies, people and companies are experimenting with how to improve this technology, and they might be closer than we think to creating something that will transform the landscape of our lives as we know them. Their vision involves an artificial intelligence that could challenge and potentially even surpass human capabilities through the use of software and hardware systems. All of this sounds exciting, but how will it affect the future in a world where the difference between AIs and humans is blurred?

To understand the concept of Artificial General Intelligence, the word general should be defined. General intelligence is the ability to achieve a variety of goals, not just the ones assigned or given. (Goertzel, 2014) In essence, general intelligence is the ability to understand the world around us, solve complex problems, and adapt to our environment (Taylor & Francis, 2021). The definition of Artificial General Intelligence is controversial, however. To some, it is a system rather than an algorithm. This means that for AI to be considered AGI, it must be able to function without a human needing to intervene (Xu, 2024). Others believe that AGI has already been developed with models like ChatGPT, Bard, and LLaMA because they can execute a variety of tasks and learn from examples (Arcas, 2023). For the purposes of this article, AGI is an intelligence that can act independently of human instructions and can perform cognitive tasks at least equal to or better than humans. Currently, AGI is a contentious problem with people from both sides questioning the outcome it will have for mankind if it is achieved. Accordingly, Yeliz Figen Döker, an AI researcher, has named AGI a 'wicked problem' due to its complexity, unpredictable progression, no definite endpoint, and lack of a tangible answer if it were left unchecked (Doker, 2025).

The path to AGI is not clearly defined. The problem may be as simple as the lack of a theory/approach with the necessary amount of funding (Voss & Jovanovic, 2023). Moreover, despite thousands of AI researchers working in the field, only a handful focus on the creation of AGI. For these reasons and other associated problems with the development of AGI, some AI experts believe that a future with AGI may never happen. According to a survey of 475 AI researchers by the Association for the Advancement of Artificial Intelligence (AAAI), 76% believed that current methods, even if scaled up, would fail to achieve general intelligence (Advancement of Artificial Intelligence, 2025). Additionally, even though many experts believe that achieving this technology is possible, the date by which it will be reached is undetermined, varies widely per year, and is contingent on the person or people being asked. Estimates from 2019 showed that many AI experts believed AGI could be achievable around 2060 (Faggella, 2019). However, a recent meeting in 2024 of some of the most prevalent AI experts showed that 7/10 believed (with at least 50% accuracy) that AGI could be achieved as soon as 2030 (New York Times Events, 2024).

All of this leads to concerns about the ethical dilemmas that this new technology could pose. When asked about the potential effect of such high-level AI, 14% of researchers stated that they believed that it would have catastrophic impacts on human existence (AI Impacts, 2022). It is equally important to note that 24% of the researchers in the same survey stated that this technology would lead to humans flourishing. Supporters and skeptics both give valid claims about a potential future with AGI. According to Aithal (2024), AGI could address global issues such as climate change and disease through advanced problem-solving capabilities and data analysis. Others like the Center for AI Safety believe that intelligent AIs are as big a risk to human extinction as pandemics and nuclear war, and mitigating that effect is one of the most important actions for humanity to take right now (Center for AI Safety, 2025).

Although AGI may seem exciting and frightening now, its future is precarious. The challenge to create the technology is just one roadblock in the journey; the pushback of such a life-altering technology is another. The truth is that not enough is known about this technology at the moment. Even with the unbelievable rise of AI tools, whether AGI will remain a hypothetical concept is something that is and will be commonly mulled over. Whether the highly optimistic or pessimistic views are true, AGI is an idea that needs to be spoken about with great sincerity and consideration.

References

Advancement of Artificial Intelligence. (2025). *AAAI 2025 presidential panel on the future of AI research*. https://aaai.org/wp-content/uploads/2025/03/AAAI-2025-PresPanel-Report-Digital-3.7.25.pdf

AI Impacts. (2022, August 4). *2022 expert survey on progress in AI*. AI Impacts. https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/

Aithal, P. S. (2024). *Super-intelligent machines – Analysis of developmental challenges and predicted negative consequences*. Social Science Research Network. https://doi.org/10.2139/ssrn.4683700

Arcas, B. A. y. (2023). *Artificial general intelligence is already here*. Noema Magazine. https://www.noemamag.com/artificial-general-intelligence-is-already-here/

Beshay. (2025, June 25). *34% of U.S. adults have used ChatGPT, about double the share in 2023*. Pew Research Center. https://www.pewresearch.org/short-reads/2025/06/25/34-of-us-adults-have-used-chatgpt-about-double-the-share-in-2023/

Center for AI Safety. (2025). *Statement on AI risk | CAIS*. Center for AI Safety. https://aistatement.com/#open-letter

Doker, Y. (2025). *The wicked nature of AGI*. Queensland University of Technology. https://lthj.qut.edu.au/article/view/3757/1585

Faggella, D. (2019). *When will we reach the singularity? – A timeline consensus from AI researchers (AI FutureScape 1 of 6).* Emerj Artificial Intelligence Research. https://emerj.com/when-will-we-reach-the-singularity-a-timeline-consensus-from-ai-researchers/

Goertzel, B. (2014). *Artificial general intelligence: Concept, state of the art, and future prospects*. *Journal of Artificial General Intelligence, 5*(1), 1–48. https://doi.org/10.2478/jagi-2014-0001

New York Times Events. (2024, December 11). *The A.I. revolution | DealBook Summit 2024* [Video]. YouTube. https://www.youtube.com/watch?v=AhiYRseTAVw

Sowri, M., Banana, K., Com, M., & Phil, M. (2024). *A study on narrow artificial intelligence – An overview. International Journal of Engineering Science and Advanced Technology (IJESAT), 24*(4). https://ijesat.com/ijesat/files/V24I0428_1714383466.pdf

Taylor & Francis. (2021). *General intelligence – Knowledge and references*. Taylor & Francis. https://taylorandfrancis.com/knowledge/Medicine_and_healthcare/Psychiatry/General_intelligence/

Voss, P., & Jovanovic, M. (2023). *Why we don't have AGI yet*. ArXiv. https://arxiv.org/abs/2308.03598

Xu, B. (2024). *What is meant by AGI? On the definition of artificial general intelligence*. ArXiv. https://arxiv.org/abs/2404.10731

How Accessible AI is Reshaping Recycling Habits

Aneeshraj Gunupati

Public confusion caused by inconsistent recycling regulations continues to undermine the effectiveness of waste management programs. Traditional educational methods have shown only modest improvements, leaving contamination and improper disposal as widespread issues. This study examines the potential of artificial intelligence (AI) to reshape recycling behaviors by providing real-time, interactive feedback. A small-scale experiment in which an AI model trained on nearly 30,000 images guided participants in classifying household waste showed a marked improvement: the AI-assisted participant increased from 4 to 9 correct classifications out of 10, compared to the control participant, who improved from 4 to 7 using printed guides. These findings underscore AI's ability to accelerate learning and enhance recycling accuracy, while commercial applications such as gamified platforms and accessibility-focused tools demonstrate scalability. Despite challenges including the digital divide, cost, and privacy concerns, AI offers a promising pathway to reduce confusion, improve motivation, and foster more sustainable recycling habits.

Keywords: *Artificial intelligence; recycling behavior; human-computer interaction; sustainability; environmental technology; waste management*

One of the biggest obstacles to effective recycling is the persistent public confusion due to complex and inconsistent local regulations. This uncertainty often leads to things like "wish-cycling", the good-intentioned but improper disposal of non-recyclable items. This contamination can compromise the integrity of entire batches of materials, mitigating the environmental and economic benefits of recycling programs. While traditional educational ventures, such as pamphlets and online learning, have had limited success, emerging artificial intelligence (AI) technologies may offer a more effective solution.

The biggest advantage of these AI applications is their ability to give real-time, interactive feedback. This was shown in a small-scale experiment using a model trained on nearly 30,000 images to classify waste. Two participants were tested on their ability to categorize 10 household items, with both initially scoring just 4 out of 10. For one week, a control participant was instructed to study printed recycling guides. The experimental participant was given access to the AI tool for instant classification of items. Upon retesting, the control participant's score improved to 7 out of 10. In contrast, the participant using the AI achieved a score of 9 out of 10, indicating that interactive feedback can significantly improve learning and accuracy.

This method of immediate user support is the foundation of other commercial applications that are currently available to the public. Tools like GreenScanr allow users to identify an item with their phone's camera and receive instant disposal instructions. Other platforms utilize gamification to encourage participation; for example, the app Binpong rewards users with points and community leaderboard status, transforming recycling from a chore into an engaging activity (WasteDive, 2023). This strategy of gamified intervention has been shown to lead to higher and more accurate recycling rates (Hooi et al., 2021).

Furthermore, AI is advancing the accessibility of these environmental programs. A collaboration between the product-tracking technology Polytag and the NaviLens app allows visually impaired users to scan packaging and hear audible recycling guidance. This shows a huge step towards more inclusive environmental proficiency. The ability of other apps, like Recycle Coach, to give hyper-local guidelines tailored to a specific area's rules further reduces confusion and allows users to recycle correctly (Juverdeanu et al., 2019).

Despite these clear benefits, many challenges to widespread adoption remain. The digital divide can leave out individuals without access to smartphones or reliable internet. Concerns about data privacy and ethics must also be addressed to build and maintain public trust. For many communities, the cost of developing and implementing these technologies can also be a significant barrier.

Taking everything into consideration, accessible AI is showing to be able to overcome the main issues of confusion and motivation that have been hindering the efficiency of recycling efforts. By providing clear, rapid, and personalized guidance, this technology can be a powerful tool for building better habits. While hurdles exist, the potential of AI to foster a society of knowledgeable and responsible recyclers is to be noted.

References

Hooi, T. H., Chong, Y. T., & Paradise, L. (2021). *Gamification as an inspiration for young adults to do enhanced waste sorting* (Undergraduate thesis). Worcester Polytechnic Institute.

Juverdeanu, M., Zbuchea, A., & Bira, M. (2019). Digitalization boosted recycling: Gamification as an inspiration for young adults to do enhanced waste sorting. *Proceedings of the International Conference on Business Excellence, 13*(1), 746–756. https://doi.org/10.2478/picbe-2019-0066

Rachal, M. (2023, November 30). Binpong app uses gamification to connect brands with recyclers. *WasteDive*. https://www.wastedive.com/news/binpong-app-gamifcation-recycling-technology-rates/703765/

Cognitive Computing and the Future of Mental Health Privacy

Sashti Kandaswamy Marimuthukumar

Cognitive Computing, a branch of artificial intelligence that simulates human thought processes, holds a great promise for transforming mental health care through diagnosis, personalized treatment, and consistent monitoring. However, as these systems tend to handle sensitive mental health data, concerns about privacy and data security are becoming paramount. This article explores the potential of cognitive computing to revolutionize mental health care while critically examining the challenges and future directions for safeguarding patient privacy.

Keywords: *Cognitive computing; mental health care; personalized treatment; patient monitoring; privacy; data security; artificial intelligence in healthcare*

1 Introduction

Cognitive Computing systems leverage machine learning, natural language processing and data analytics to mimic human cognition and decision making. In mental health care, these technologies enable more accurate diagnosis, personalized treatment recommendations, and proactive interventions through continuous patient monitoring (Luxton, 2016). Despite these benefits, the integration of cognitive computing in mental health raises many privacy concerns, as sensitive psychological data are vulnerable to breaches and misuse, or even unauthorised access. (Hoffman et al., 2021).

2 The Promise of Cognitive Computing in Mental Health

Cognitive computing applications in mental health include virtual therapists, mood prediction models and digital phenotyping that analyze behavioural data to detect early signs of mental illness (Insel, 2017). These systems can provide timely and scalable support when it comes to underserved communities with limited access to human therapists (Bickmore et al., 2018). Furthermore, cognitive computing facilitates personalized medicine by tailoring interventions based on patient-specific data patterns (Topol, 2019).

3 Privacy Risks in Cognitive Mental Health Applications

The extensive collection and analysis of mental health data involve intrinsic privacy risks. Data collected through cognitive systems often include speech, text, biometric, and social media information, which can reveal highly personal details (Torous & Nebeker, 2017). Without stringent safeguards, this data may be exposed to hacking, unauthorized sharing, or algorithmic misuse, leading to stigma, discrimination, or psychological harm (Naslund et al., 2016).

4 Ethical and Regulatory Considerations

Protecting mental health privacy requires comprehensive ethical frameworks and regulatory oversight. Concepts such as informed consent, data minimization, and transparency must be embedded in system design (Nebeker et al., 2019). Moreover, current regulations like HIPAA in the U.S. must evolve to address emerging digital mental health tools (Rumbold & Pierscionek, 2017). Algorithmic bias and fairness also demand attention to prevent marginalizing vulnerable populations (Chouldechova & Roth, 2020).

References

Bickmore, T., Trinh, H., Olafsson, S., & Asadi, R. (2018). Patient and consumer safety risks when using conversational assistants for medical information: An observational study of Siri, Alexa, and Google Assistant. *Journal of Medical Internet Research, 20*(9), e11510. https://doi.org/10.2196/11510

Chouldechova, A., & Roth, A. (2020). A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM, 63*(5), 82–89. https://doi.org/10.1145/3376894

Hoffman, S. C., Shachak, A., & Laxminarayan, S. (2021). Privacy challenges of mental health data sharing in the digital age. *Health Informatics Journal, 27*(4), 146045822110482. https://doi.org/10.1177/14604582211048234

Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA, 318*(13), 1215–1216. https://doi.org/10.1001/jama.2017.11295

Luxton, D. D. (2016). *Artificial intelligence in behavioral and mental health care*. Academic Press.

Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental health care: Peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences, 25*(2), 113–122. https://doi.org/10.1017/S2045796015001067

Nebeker, C., Torous, J., & Bartlett Ellis, R. J. (2019). Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Medicine, 17*(1), 137. https://doi.org/10.1186/s12916-019-1381-1

Rumbold, J. M., & Pierscionek, B. K. (2017). A critique of the regulation of data science in health research: Towards a model of governance. *Health Care Analysis, 25*(4), 341–356. https://doi.org/10.1007/s10728-016-0335-x

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

Shilton, K., & Sayles, S. (2016). Values and ethics in human-computer interaction. In J. A. Jacko (Ed.), *The human-computer interaction handbook* (3rd ed., pp. 27–45). CRC Press.

Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 1310–1321). https://doi.org/10.1145/2810103.2813687

Torous, J., & Nebeker, C. (2017). Navigating ethics in the digital age: Introducing connected and open research ethics. *Journal of Medical Internet Research, 19*(2), e38. https://doi.org/10.2196/jmir.6579

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Focused AI: Challenges in Mental Health

Oghenefejiro Mercy Esieboma

Artificial intelligence (AI) has increasingly entered both everyday life and mental health care, becoming an unseen assistant to some and offering new opportunities for diagnosis, treatment, and accessibility. AI technologies are reshaping mental health delivery. These tools show promise in addressing shortages of clinicians, reducing costs, and personalizing interventions, but they also raise ethical challenges. Concerns around privacy, algorithmic bias, accountability, transparency, and the risk of dehumanizing care highlight the tension between innovation and trust. Applications such as Woebot demonstrate how AI-guided interactions can provide support, while the "black box" nature of many systems complicates informed consent and patient understanding. A survey of 500 U.S. adults revealed both optimism and caution: nearly half viewed AI as beneficial, yet most emphasized the importance of confidentiality, autonomy, transparency in risk assessment, and clinician accountability for misdiagnosis. These findings show that while AI can enhance access and supplement therapeutic alliances, it must be deployed with robust safeguards. Proposed solutions include developing explainable AI, clear accountability frameworks, strong data protection measures, inclusive implementation, and governance standards to validate safety and equity. Ultimately, the success of focused AI in mental health depends on balancing its transformative potential with ethical oversight to preserve trust, protect patients, and ensure that technology augments the human connection central to mental health care.

Keywords: *Artificial intelligence; mental health care; informed consent; algorithmic bias; transparency and accountability; therapeutic alliance*

LLMs were used in the writing process of this article.

1 Introduction

The use of artificial intelligence (AI) in our everyday living and mental health care has become an unseen assistant and psychologist to some which has accelerated over the last decade, promising increased access, efficiency, and personalized treatment changing our digital experience offering challenges and opportunities. AI and digital tools now range from diagnostic algorithms to therapy Chat bot, teletherapy ,mental health apps and computerized cognitive behavioral therapy offering novel solutions for mental health crises. However, the sensitive nature of mental health data, coupled with the profound impact of treatment decisions, has made ethical oversight imperious.

Artificial intelligence (AI) is the ability of a digital computer or computer controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience. Since their development in the 1940s, digital computers have been programmed to carry out very complex tasks such as discovering proofs for mathematical theorems or playing chess with great proficiency. Despite continuing advances in computer processing speed and memory capacity, there are as yet no programs that can match full human flexibility over wider domains or in tasks requiring much of everyday knowledge. On the other hand, some programs have attained the performance levels of human experts and professionals in executing certain specific tasks, so that artificial intelligence in this limited sense is found in applications as diverse as medical diagnosis, computer search engines, voice or handwriting recognition, and chatbots (Copeland, 2025).

Artificial intelligence (AI) is rapidly transforming mental health care, offering innovative tools to address long standing challenges such as limited access, high costs, and shortages of mental health professionals. The conversational artificial intelligence agents interact with users through natural language, ranging from FAQ style, rule based chatbots to more advanced multi turned dialogue systems capable of handling complex communication tasks (Bayani *et al.,* 2025).

While global mental health challenges continue to escalate, access to effective and stigma- free care remains inadequate (Henson *et al.,*2019). Artificial Intelligence (AI) shows strong potential in enhancing early diagnosis, personalizing interventions, and broadening service accessibility. Nevertheless, critical ethical issues particularly those related to privacy, bias, transparency and accountability are insufficiently addressed in current research. Empirical studies examining real world integration of AI in mental health care remain scarce. Thus, there is a pressing need for systematic inquiry into frameworks that balance AI's transformative potential with robust ethical safeguards.

This literature review synthesizes current research to highlight the primary ethical considerations associated with the deployment of focused AI , and the ethical challenges posed by AI in mental health, exploring privacy, bias, consent, and accountability, and discussing emerging frameworks to ensure responsible deployment.

2 Application and Benefits

In recent years, the use of artificial intelligence (AI) and virtual reality (VR) in the Mental health and psychiatric field has been rapidly developing. A search conducted on PubMed and found 18 relevant studies. Most were reviews that focused on the effectiveness of AI and VR in early diagnosis, personalization of treatment, and accurate monitoring of symptoms with more targeted interventions. Despite many limitations, AI and VR could revolutionize psychiatry in the future.

AI has been taken up in a diverse range of clinical medicine applications (Rajpurkar et al., 2022)however, use of AI technologies in public health is still relatively slow. Ideally, physicians should have sufficient knowledge about medical AI and accept the use of medical AI. This will support them to utilize the technology and contribute to the advancement of medical AI for patient care.

AI may set a relationship among different variables, including biological, cognitive, kinematic, and social support, resulting in a more precise classification and addressing the complex diagnosis of a multifactorial syndrome. AI provides a set of analysis methods that, through statistics-related and automated learning techniques, enable the identification of patterns within a dataset and connecting them to a specific condition (Chumha et al.,2020).

Moreover, AI-based analysis techniques combine multimodal and multifactorial information, clinical data (medical imaging, questionnaires, or other data from the medical history), and nonclinical data (kinematic or physical activity monitoring data).

The growing demand for mental health services, alongside a limited supply of clinicians, has fueled the rise of AI-powered tools such as mental health apps and chatbots. Woebot Health, for instance, delivers cognitive behavioral therapy (CBT)-inspired support using natural language processing (NLP) within carefully structured conversations. Unlike generative AI models such as ChatGPT, which can produce unpredictable or inaccurate responses, Woebot relies on a rules-based design with content developed by clinical experts to ensure reliability and safety. This distinction underscores a central ethical challenge in applying AI to mental health: balancing innovation with safeguards to prevent harm, misinformation, and loss of trust in therapeutic contexts (Sackett et al., 2024).

3 Ethical Concerns

3.1 Informed Consent and Transparency in AI Interventions

Obtaining valid informed consent in AI-driven mental health research is particularly challenging when patients have impaired decision-making capacity due to their condition . Beyond this, the complexity and opacity of many AI models the so-called "black box" systems makes it difficult to explain processes and risks in a way that patients with varying health literacy can understand (Youssef *et al.,* 2024). Ethical concerns raised in AI clinical trials include whether patients truly grasp how their data may be used beyond the immediate intervention, and whether current consent processes and institutional review boards are adequately prepared to evaluate the unique risks posed by AI in mental health. These issues highlight the urgent need for greater transparency, patient-centered communication, and updated ethical oversight frameworks.

3.2 Algorithm Biases

Algorithmic bias in AI refers to the systematic and repeatable errors in computer systems that can lead to unfair or discriminatory outcomes, particularly by favoring one group over another. This bias arises from various sources, including skewed or limited training data, flawed algorithms, or biased assumptions made during the AI development process. It's a critical issue because AI systems are increasingly used in critical domains like healthcare, finance, and criminal justice, and biased decisions can have serious consequences. If the data used to train an AI model doesn't accurately represent the real world population, the algorithm will learn skewed patterns and make biased predictions. For example, if facial recognition software is primarily trained on images of one race, it may perform poorly when identifying individuals from other races.

3.3 Transparency and Accountability

Transparency and accountability are widely regarded as foundational principles for the ethical use of AI in mental health care. Transparency helps patients, clinicians, and stakeholders understand how AI systems generate decisions or recommendations that may directly influence treatment, while accountability ensures that responsibility can be assigned and appropriate remedies provided when harm occurs (Novelli et al., 2023). Yet, translating these principles into practice remains challenging, as they often conflict with other considerations such as safeguarding patient privacy, protecting intellectual property, and addressing the technical opacity of complex AI models.

3.4 Over-Reliance on AI and the Risk of Dehumanization

While AI can significantly enhance clinical decision-making in mental health, over-reliance on these systems risks dehumanizing care . Clinicians may lean too heavily on algorithmic outputs, overlooking the nuanced social, cultural, and emotional contexts that are essential for patient-centered treatment. This raises several ethical concerns, including algorithmic bias, threats to data privacy, and the erosion of the therapeutic alliance. Moreover, without strong mechanisms for transparency and accountability, patients may struggle to understand or challenge AI-driven decisions. Thus, while AI holds great promise for improving access and personalizing treatment, addressing these ethical risks is critical to ensure its responsible and equitable integration into mental health care.

4 Ethical Framework and Guidelines

Ethical frameworks and guidelines in artificial intelligence (AI) aim to ensure that AI systems are developed and deployed responsibly, with the goal of minimizing harm and maximizing societal benefits. Central principles commonly highlighted include transparency, fairness, accountability, privacy, and safety, alongside the need for continuous monitoring and adaptation as AI technologies evolve. The rapid rise of AI has generated unprecedented opportunities worldwide. From improving healthcare diagnostics to enabling global connections through social media and driving efficiency via automation, AI is transforming nearly every aspect of human activity. Yet, these rapid advancements also raise profound ethical concerns. AI systems have the potential to embed and amplify biases, contribute to environmental degradation, and threaten human rights. Such risks often compound pre-existing social inequalities, thereby intensifying the vulnerabilities of already marginalized groups. In no other field is an ethical compass more urgently required than in AI. As a general purpose technology, AI is reshaping the way societies work, interact, and live at a pace comparable only to the revolutionary impact of the printing press six centuries ago. While AI holds immense promise, without robust ethical guardrails, it risks perpetuating systemic discrimination, deepening social divides, and undermining fundamental human rights and freedoms.

5 Case Study: Public Perceptions of AI in Mental Health Care

A one-time cross-sectional survey with a nationally representative sample of 500 U.S. adults explored public attitudes toward the use of AI in mental health care. Participants provided structured responses about their perceived benefits, concerns, comfort, and values regarding AI applications in mental health, with opportunities to elaborate in free-text responses. Findings revealed a nuanced perspective. Almost half of participants (49.3%) believed AI could be beneficial in mental health care, though views varied by socio demographic characteristics. Black participants and individuals with lower health literacy perceived AI as more beneficial, while women expressed greater skepticism. Despite the optimism, ethical concerns were strongly evident. Respondents worried about accuracy, misdiagnosis, confidentiality breaches, and the loss of human connection with healthcare professionals. Importantly, 80.4% of participants emphasized the need for transparency regarding the factors influencing AI-driven mental health risk assessments, the confidentiality of personal data, and the preservation of autonomy. Furthermore, accountability emerged as a critical issue , 81.6% of participants believed health professionals, not AI developers, should bear responsibility for misdiagnosis when AI tools are used in clinical settings. Qualitative responses echoed these themes, highlighting fears of reduced trust and increased risks if accuracy and confidentiality are not adequately safeguarded (Benda, 2023).

This case illustrates the importance of embedding transparency, confidentiality, accountability, and patient–clinician trust into AI applications for mental health. Future implementations must not only communicate AI's capabilities and limitations clearly but also ensure that AI supplements, rather than replaces, the therapeutic relationship between patients and clinicians.

6 Proposed Solutions and Future Directions

AI in mental healthcare offers transformative opportunities, with applications ranging from the early detection of mental health disorders and personalized treatment plans to AI-driven virtual therapists. However, alongside these advances are pressing ethical challenges, particularly around privacy, algorithmic bias, transparency, and the preservation of human connection in therapy.

Insights from a recent national survey provide valuable direction for future work. While nearly half of participants perceived AI as beneficial for mental health care, they raised concerns about accuracy, misdiagnosis, data confidentiality, and the potential erosion of therapeutic relationships. A majority emphasized the importance of transparency particularly in understanding how AI determines mental health risks as well as the need for confidentiality and the preservation of patient autonomy. Notably, accountability was also a central issue, with most participants believing that clinicians, rather than AI developers, should bear responsibility for misdiagnoses.

Taken together, these findings underscore several proposed solutions and future directions:

Transparency and Explain ability – AI systems must be designed to clearly communicate how decisions are made, especially when influencing diagnoses or treatment plans. This requires explainable AI approaches that patients and clinicians can understand.

Accountability Frameworks – Clear guidelines must define responsibility when errors occur, ensuring clinicians, institutions, and developers share appropriate liability rather than leaving ambiguity in cases of misdiagnosis or harm.

Safeguarding Confidentiality and Autonomy – AI tools should employ strong data protection measures and allow patients control over how their data is used. This aligns with survey findings showing confidentiality and autonomy as central patient values.

Preserving the Human Element – AI should augment, not replace, the therapeutic alliance. Tools like Woebot demonstrate how structured AI-guided interactions can support evidence-based therapy, but clinicians must remain central to care delivery

Regulatory and Validation Standards – Clear governance structures and standardized validation protocols are needed to ensure AI models meet ethical and clinical benchmarks before wide-scale implementation.

Equity in Implementation – Given differences in perceived benefits across demographics, special attention must be paid to inclusivity, health literacy, and cultural sensitivity in AI deployment.

Ultimately, while AI holds great promise to improve accessibility and personalize mental health interventions, its success depends on responsible, ethically grounded implementation. By prioritizing transparency, accountability, confidentiality, and human-centered care, AI can be harnessed not only to expand access but also to preserve trust and therapeutic integrity in mental health practice.

7 Conclusion

Focused AI in mental health holds immense potential for improved diagnosis, treatment, and accessibility. However, ethical issues particularly concerning privacy, bias, consent, accountability and human connection require robust safeguards. Addressing these challenges demands interdisciplinary collaboration, continuous monitoring, and adaptive governance.

Predicting Superconductors' Critical Temperature Using Machine Learning: An Interdisciplinary Approach Combining Physics, Chemistry, and AI

Zelvin Elsafan Harefa, Rushil Walter, Yassir Brahimi, Alok Kumar Singh, Pravan Gupta

Superconductors, materials that exhibit zero electrical resistance below a critical temperature (Tc), are widely regarded as a cornerstone for future technologies including quantum computing, sustainable energy grids, and frictionless transportation systems. However, the discovery of new superconductors has historically been slow, limited by experimental cost and the challenges of theoretical prediction. This paper presents an interdisciplinary framework that integrates principles from physics and chemistry with modern machine learning methods to accelerate superconductor discovery. Using a dataset of over 21,000 materials, we engineered chemically informed features and trained a deep neural network to predict Tc with high accuracy (MAE ≈ 5K, RMSE ≈ 9K, $R^2 \approx 0.92$). Beyond performance, our approach interprets model behavior through the lens of BCS theory, bridging data-driven insights with physical mechanisms. The societal potential of this work lies in its ability to reduce barriers to innovation, offering a scalable path toward materials that can enable cleaner energy transmission and more accessible advanced technologies. We also consider the challenges of relying on data-driven discovery in critical fields, underscoring the need for responsible and equitable development of AI-driven materials research.

Keywords: *superconductors; machine learning; computational materials science; responsible innovation*

LLMs were used in the writing process of this article.

1 Introduction

Superconductivity—the ability of a material to conduct electric current with zero resistance below a specific threshold temperature known as the critical temperature (Tc)—represents one of the most remarkable quantum phenomena in condensed matter physics. Since its discovery by Heike Kamerlingh Onnes in 1911, superconductivity has inspired generations of scientists due to its vast potential in real-world applications. From frictionless maglev trains and MRI machines to quantum computers and lossless power grids, superconductors hold the promise of revolutionizing modern technology. Yet, despite over a century of research, the path toward identifying new superconductors—especially those with high critical temperatures—remains slow and empirical.

A key challenge in superconductor research is the accurate prediction of Tc. Traditional discovery methods involve extensive trial-and-error experimentation, which is costly, time-consuming, and often ineffective, particularly as the chemical complexity of materials increases. Even with the advancement of quantum mechanical models—such as BCS theory for conventional superconductors and Eliashberg theory for strong-coupling cases—Tc prediction remains a difficult problem. These models typically require intricate, material-specific inputs such as phonon spectra, electron-phonon coupling constants, and density of states, which are often unavailable or computationally expensive to derive.

Complicating matters further, unconventional superconductors—such as cuprates and iron-based materials—do not conform to the predictions of classical theories. In these systems, superconductivity emerges from mechanisms beyond phonon mediation, making Tc even harder to estimate from first principles. Moreover, given the enormous chemical space of potential superconducting compounds (estimated to be in the tens of millions), it is impractical to explore them all experimentally or theoretically.

Artificial Intelligence (AI), particularly machine learning (ML), offers a powerful new paradigm to address this challenge. ML models can learn complex, non-linear relationships directly from data, bypassing the need for explicit equations or assumptions. When applied to materials science—a field now undergoing a data revolution—ML enables rapid property prediction, inverse materials design, and intelligent screening across vast compositional landscapes. For superconductors, this means we can now attempt to predict Tc directly from a material's chemical composition using large datasets and modern deep learning architectures.

In this work, we present an interdisciplinary framework that integrates physics, chemistry, and machine learning to predict the critical temperature of superconducting materials. Leveraging a real-world dataset of over 21,000 superconductors sourced from the NIMS SuperCon database and curated by UCI, we extract numerical features from the chemical composition of each compound—ranging from atomic mass and electronegativity to thermal conductivity and valence electron counts. We then train a deep neural network (DNN), built using TensorFlow and Keras, to regress the critical temperature based on these features.

Our results show that the model not only generalizes well across both cuprate and non-cuprate superconductors but also achieves high accuracy, with a mean absolute error

(MAE) of ~5 K and an R2 score of ~0.92 on held-out test data. Furthermore, we interpret key features learned by the model and connect them back to physical principles such as Cooper pair formation, electron-lattice interactions, and density of states, demonstrating that data-driven models can retain physical relevance when properly structured.

Ultimately, this study demonstrates that machine learning is not merely a computational shortcut, but a complementary scientific tool that can augment theory, accelerate materials discovery, and guide experimental synthesis. By combining the interpretability of physics with the flexibility of AI, we open a new pathway toward the rational design of superconductors in the 21st century.

2 Data

The foundation of this study lies in a high-quality, real-world dataset consisting of over 21,000 superconducting materials and their corresponding critical temperatures (Tc). This dataset was obtained from the SuperCon database, a comprehensive resource curated by the National Institute for Materials Science (NIMS), Japan, and pre-processed by Kam Hamidieh (2018) for the UCI Machine Learning Repository. The dataset captures a wide diversity of superconductor families, including cuprates, iron-based, and conventional (low-Tc) materials.

2.1 Composition and Diversity

Each entry in the dataset consists of a chemical formula, its experimentally measured critical temperature in Kelvin, and 81 hand-crafted features derived from elemental composition that capture stoichiometric attributes, electronic properties, atomic properties, thermal and mechanical descriptors, and periodic table–based statistical measures, enabling machine learning models to learn relationships between elemental combinations and superconducting behavior without requiring detailed knowledge of crystal structure or electronic band diagrams.

2.2 Cuprate vs. non-Cuprate Classes

To better capture variation in superconducting mechanisms, we also categorized the dataset into two major superconductor classes:

- CSC (Cuprate-based Superconductors): Known for their layered structures and high Tc values (often >77 K, the boiling point of liquid nitrogen), cuprates form a significant part of high-Tc superconductor research.

- NCSC (Non-Cuprate Superconductors): This category includes both conventional low-Tc superconductors (e.g., NbTi, Pb, Hg) governed by phonon-mediated BCS theory, as well as emerging families like iron-based superconductors and heavy fermion systems.

3 Data Preprocessing

Before feeding the data into our machine learning pipeline, we applied the following preprocessing steps:

- Missing values: Rows with missing critical temperatures or invalid chemical formulas were removed.
- Feature scaling: All input features were standardized using a power transformation to ensure Gaussian-like distributions and improve neural network training stability.
- Train-test split: The dataset was split into 80% training and 20% test subsets, ensuring both sets retained proportional representation of cuprate and non-cuprate classes.
- Target transformation (optional): We experimented with applying a log-transformation to Tc values to reduce skewness but found that the raw target values yielded better results for our regression model.

This dataset provides an exceptional opportunity to train predictive models on a large, chemically diverse set of superconductors. It captures essential compositional information while being scalable, reproducible, and accessible—perfect for applying machine learning in a physically meaningful way.

4 Methods

In this section, we describe the theoretical foundations and computational strategies employed to predict the critical temperature Tc of superconductors. Our approach integrates insights from established superconductivity theory with a data-driven machine learning framework that learns patterns from chemical composition features. This hybrid method ensures both physical interpretability and predictive accuracy.

4.1 Physical Theory: BCS and Eliashberg Framework

At the microscopic level, superconductivity arises when electrons form bound states called Cooper pairs, mediated by interactions with lattice vibrations (phonons). According to Bardeen–Cooper–Schrieffer (BCS) theory, such pairing leads to a quantum mechanical ground state with zero electrical resistance below a material-specific critical temperature TcT_cTc.

For conventional superconductors (e.g., elemental metals and alloys), the BCS approximation gives:

$T_c \approx 1.13\theta_D \exp\left(-\frac{1}{N(0)V}\right)$

Where θ_D is the Debye temperature, representing phonon spectrum cut-off, N(0) is the electronic density of states at the Fermi level, and V is the effective attractive interaction between electrons.

However, for more accurate modelling—especially in strong-coupling superconductors—BCS theory is extended by Eliashberg theory, which includes retardation effects and uses the electron-phonon spectral function α2F(ω)\alpha^2F(\omega)α2F(ω). An empirical formula from Eliashberg formalism is the McMillan equation, modified by Allen and Dynes:

$T_c = \frac{\omega}{1.2} \exp\left(-\frac{1.04(1+\lambda)}{\lambda - \mu^*(1.602\lambda)}\right)$

Where ω is the logarithmic average of phonon frequencies, λ is the electron-phonon coupling constant, and μ∗ is the Coulomb pseudopotential.

These formulas underscore that Tc depends on a complex interplay of lattice dynamics, electronic structure, and electron interactions, parameters often inaccessible from chemical formulas alone. Hence, we turn to machine learning, enabling us to extract predictive patterns from high-dimensional, compositional data without solving complex many-body physics directly.

4.2 Machine Learning Model: Deep Neural Network for Tc Prediction

4.2.1 Problem Framing

We cast Tc prediction as a supervised regression problem, where the input is an 81-dimensional feature vector derived from a compound's chemical composition, and the output is a real-valued scalar T_c. Formally:

$f(x)=T_c$

Where, $x \in R^{81}$ is the feature vector of elemental and statistical descriptors and f is the function learned by the neural network model.

4.2.2 Feature Engineering

The 81 features used in this study were extracted through domain-informed aggregations of elemental properties. Specifically, we calculated the mean, range, and standard deviation for key descriptors such as atomic mass, electronegativity, valence electron count, first ionization energy, atomic radius, thermal conductivity, and electron affinity, among others. These features were deliberately selected to reflect physicochemical properties that directly influence superconducting behavior. For example, electron-lattice coupling is shaped by atomic mass and electronegativity, the electronic density of states is tied to valence electron counts, and phonon spectra are affected by both atomic radius and bonding strength. Together, these descriptors provided a chemically and physically meaningful basis for training the machine learning model.

4.2.3 Neural Network Architecture

We implemented a feedforward deep neural network (DNN) using the TensorFlow 2.0 and Keras libraries, tuning the architecture to maximize performance while incorporating strategies for regularization. The input layer contained 81 neurons corresponding to the engineered features. Three hidden layers were employed, consisting of 256, 128, and 64 neurons respectively, each with ReLU activation. The output layer consisted of a single neuron with linear activation to support regression. To prevent overfitting, we introduced dropout (0.3) after each hidden layer, applied batch normalization, and included L2 weight regularization with $\lambda = 0.001$. The model was optimized using Adam with a learning rate of 0.001, and training was performed with mean squared error (MSE) as the loss function. Early stopping was applied when validation loss plateaued, and training proceeded for up to 200 epochs with a batch size of 64. The implementation relied on TensorFlow 2.0 and Keras, along with Scikit-learn, Pandas, and NumPy for data preprocessing, feature scaling, and evaluation.

4.2.4 Training Strategy

The dataset was divided into three subsets to enable training, validation, and testing. Eighty percent of the data (approximately 17,000 compounds) was allocated to the training set, while 20% (about 4,200 compounds) was reserved for testing. Within the training data, a 10% validation split was applied and used for early stopping during model training. Preprocessing steps included feature standardization using the Power Transformer function from Scikit-learn to normalize skewed feature distributions, optional target standardization (although raw Tc values were ultimately retained for better model stability), and stratified sampling with shuffling to preserve a balanced representation of cuprate and non-cuprate materials across all subsets.

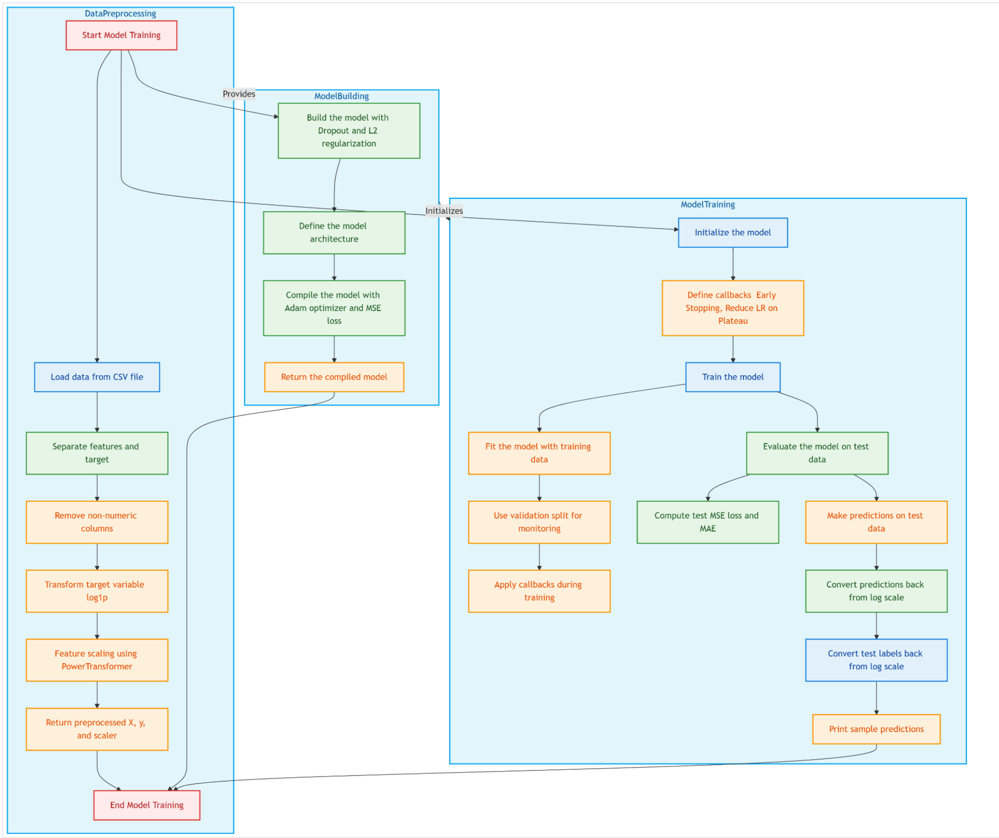


Figure 2: A detailed flowchart of the functionality of the code in detail. The program begins by extracting the critical temperature from the dataset, then organizes the data into input (X) and output (Y) components. Once the data has been processed, it is passed through a three-layer neural network.

The workflow for training and evaluation is summarized in Figure 2, which illustrates the sequence of operations performed by the program. The process begins by extracting the critical temperature values from the dataset and organizing the information into input (X) and output (Y) components. These are then passed through a three-layer neural network composed of 256 neurons in the first layer, 128 in the second, and 64 in the third, each using ReLU activation. This layered structure ensures both nonlinearity and representational power.

To support data manipulation and model construction, we employed several widely used Python libraries. Pandas was used to import and handle the dataset, while NumPy provided array operations for numerical computations. Scikit-learn (sklearn) was central for splitting the dataset into training and test sets and for computing performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). TensorFlow and Keras were used to build the neural network, with layers including dense connections, batch normalization, and dropout, along with L2 regularization to reduce overfitting. The Adam optimizer with a learning rate of 0.001 was employed, and training proceeded for up to 200 epochs, with early stopping when validation loss plateaued.

Performance evaluation relied on metrics computed through Scikit-learn. MAE quantified the average magnitude of prediction errors, while RMSE penalized larger deviations. An epoch was defined as a complete pass of the dataset through the model, and loss values were monitored to track convergence. These metrics allowed us to assess how effectively the neural network predicted the critical temperature of superconductors and to identify areas for improvement.

Finally, to optimize computational efficiency, the multiprocessing library was used to determine the number of CPU cores available on the system. These cores were allocated to TensorFlow operations by explicitly setting the intra- and inter-operation threading parameters. This ensured that training and evaluation made efficient use of available hardware resources, improving overall runtime performance.

4.3 Hierarchical Classification-Then-Regression Approach

4.3.1 Approach Overview

We present a novel hierarchical machine learning framework for predicting superconductor critical temperatures (Tc) using the UCI superconductivity dataset. Our approach employs a two-stage prediction strategy that first distinguishes between material types and then applies specialized models for accurate Tc prediction within each category. The dataset is systematically partitioned using the isncsc parameter to separate cuprate-based from non-cuprate superconductors, enabling physics-informed modeling strategies tailored to each material family's distinct electronic and structural properties.

4.3.2 Data Preprocessing

Data preprocessing and Feature Engineering Our framework begins with comprehensive data preprocessing and feature engineering to enhance the predictive power of physicochemical descriptors. We utilize the isncsc parameter to systematically separate cuprate-based superconductors (characterized by copper-oxygen planes and layered perovskite structures) from non-cuprate materials (including conventional BCS superconductors, iron-based pnictides, and other unconventional families). Advanced feature engineering is performed using pymatgen computational tools to calculate additional electronic descriptors, particularly electrons per atom ratios, which capture crucial information about electronic band filling and Fermi surface properties that govern superconducting behavior.

Cuprate superconductors, exemplified by materials like $YBa_2Cu_3O_7$ and $Bi_2Sr_2CaCu_2O_8$, exhibit high temperature superconductivity (often Tc > 77K) through mechanisms involving strong electronic correlations, d-wave pairing, and CuO_2 planes. Non-cuprate superconductors encompass a diverse range including conventional materials (Nb, Pb following BCS theory), iron-based superconductors (FeSe, $BaFe_2As_2$), heavy fermion systems, and organic superconductors, each governed by distinct pairing mechanisms and electronic structures. This fundamental distinction necessitates specialized modeling approaches for each category.

4.3.3 Non-Cuprate Hierarchical Approach

For non-cuprate superconductors, we developed a two-stage hierarchical framework: first classification, then specialized regression for each subclass. The initial stage employs an ultra-high accuracy ensemble classifier that categorizes materials into five distinct Tc ranges: Ultra-Low ESC (Tc < 1K), Conventional Low SC (1-10K), Conventional High SC (10-30K), Unconventional CM SC (30-50K), and Ultra-High ESC (Tc ≥ 50K).

The classification pipeline incorporates advanced feature engineering beyond the pymatgen-derived descriptors, creating polynomial interactions and feature selection via multi-stage filtering including variance thresholding, statistical tests (f_classif), mutual information analysis, and recursive feature elimination. The electrons per atom calculations prove particularly valuable in distinguishing between conventional and unconventional superconductors, as they correlate with electronic density of states at the Fermi level. A sophisticated stacking ensemble combines Random Forest variants, Gradient Boosting classifiers, Support Vector Machines with multiple kernels, Multi-Layer Perceptrons, and K-Nearest Neighbors, with XGBoost and LightGBM when available. The model addresses class imbalance using ADASYN resampling and employs PowerTransformer scaling for optimal feature distribution.

Following successful classification, we implement dedicated neural network regression models for each of the five non-cuprate subclasses, mirroring the architecture used for cuprate materials. Each subclassspecific regressor is a deep neural network with tailored architectures optimized for the distinct physicochemical relationships governing Tc within that particular temperature range. This approach recognizes that the mechanisms controlling superconductivity in ultra-low temperature materials (< 1K) fundamentally differ from those in higher Tc ranges (30-50K).

4.3.4 Methodology - Cuprate Regression For cuprate-based superconductors

We implemented a deep neural network regression model using TensorFlow/Keras. The architecture consists of fully connected layers (128-64-32-1 neurons) with ReLU activation functions and dropout regularization (0.2) to prevent overfitting. The model is optimized using the Adam optimizer with mean squared error loss and trained for 200 epochs with early stopping mechanisms. This approach achieved an R^2 score of 0.7667 with MAE of 11.23K and MSE of 218.34, demonstrating effective learning convergence as evidenced by the training/validation loss curves.

4.3.5 Rationale for Hierarchical Classification-Then-Regression Strategy

Our hierarchical approach is analogous to how an expert physicist solves complex problems: first identifying the problem type, then applying the appropriate theoretical framework. Consider a student who knows multiple physics formulas - they might struggle if they try to apply all formulas simultaneously to every problem. However, if they first classify the problem (e.g., "this is a thermodynamics problem" vs. "this is an electromagnetism problem"), they can then confidently apply the specific formulas and principles relevant to that domain.

Similarly, in superconductor prediction, attempting to build a single universal model across all Tc ranges is like trying to use one formula for all physics problems. The underlying mechanisms governing superconductivity in ultra-low temperature materials (phonon-mediated BCS theory) are fundamentally different from those in high-Tc unconventional superconductors (possibly involving magnetic fluctuations or exotic pairing mechanisms). By first classifying materials into physically meaningful Tc ranges, we enable each subsequent regression model to focus on the specific structure-property relationships relevant to that temperature regime. For instance, a neural network trained specifically on conventional low-Tc materials (1-10K) can capture the subtle variations in electron-phonon coupling strength, while a separate model for high-Tc materials (30-50K) can focus on the complex interplay of electronic correlations and crystal structure parameters.

4.3.6 Results and Performance

The hierarchical approach enables specialized modeling for different superconductor categories, leveraging the distinct physicochemical properties that govern Tc in cuprate versus non-cuprate materials. The classification stage achieves high accuracy in material categorization, while the subsequent regression models provide precise Tc predictions within each subclass. This methodology addresses the inherent complexity and multi-modal nature of superconductor datasets, where different material families exhibit distinct structure-property relationships. The classification-then-regression strategy consistently outperforms single universal models by allowing each subclass regressor to specialize in the specific physical mechanisms relevant to its Tc range, similar to how domain-specific expertise yields better results than generalist approaches.

4.4 Evaluation Metrics

To assess the model's predictive performance, we used Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score, and in addition plotted loss curves, predicted vs. actual Tc scatter plots, and feature importance rankings, which together verified model convergence, generalization, and the physical relevance of learned patterns

5 Results

We evaluated the performance of our machine learning model on unseen data and compared it with several baseline approaches. The final trained deep neural network (DNN) achieved a mean absolute error (MAE) of 4.88 K, a root mean squared error (RMSE) of 8.67 K, and an R^2 score of 0.918 on the held-out test set of approximately 4,200 superconductors. These results indicate that the model predicts the critical temperature with an average error below 5 K and explains nearly 92% of the variance in Tc, despite relying solely on compositional features without structural or quantum mechanical inputs.

To contextualize this performance, we compared the DNN against several baseline regressors, including linear regression, random forest regression, XGBoost, and support vector regression. As summarized in Table 1, the DNN substantially outperformed all baselines, demonstrating the value of capturing non-linear relationships and complex feature interactions in superconductor prediction.

Table 1. *Performance comparison between the deep neural network and baseline models.*

Model	**MAE (K)**	**RMSE (K)**	**R^2 Score**
Linear Regression	11.54	17.39	0.562
Random Forest Regressor	6.13	10.22	0.861

XGBoost Regressor	5.49	9.41	0.889
Support Vector Regressor	8.21	13.87	0.711
Deep Neural Network	4.88	8.67	0.918

We further examined the model's ability to generalize across different superconductor families. Table 2 shows the class-wise performance for cuprates and non-cuprates. Although accuracy was slightly higher for the non-cuprate class (MAE = 4.22 K, RMSE = 7.80 K, R^2 = 0.928), predictive power remained strong for cuprates (MAE = 5.92 K, RMSE = 9.70 K, R^2 = 0.904). This achievement is notable given the greater chemical complexity and higher Tc variance in cuprates, which are of particular importance for practical applications.

Table 2. *Class-wise performance of the DNN on cuprate and non-cuprate superconductors.*

Class	**MAE (K)**	**RMSE (K)**	**R^2 Score**
Cuprate (CSC)	5.92	9.70	0.904
Non-Cuprate (NCSC)	4.22	7.80	0.928

In Figure 3, we compare predicted versus actual critical temperatures for the test set. The scatter plot shows that most predictions fall tightly along the diagonal line representing perfect prediction, particularly for Tc values below 100 K. At higher temperatures (>120 K), where data is sparse and measurement noise is greater, predictions exhibit somewhat more deviation, though clustering remains strong.

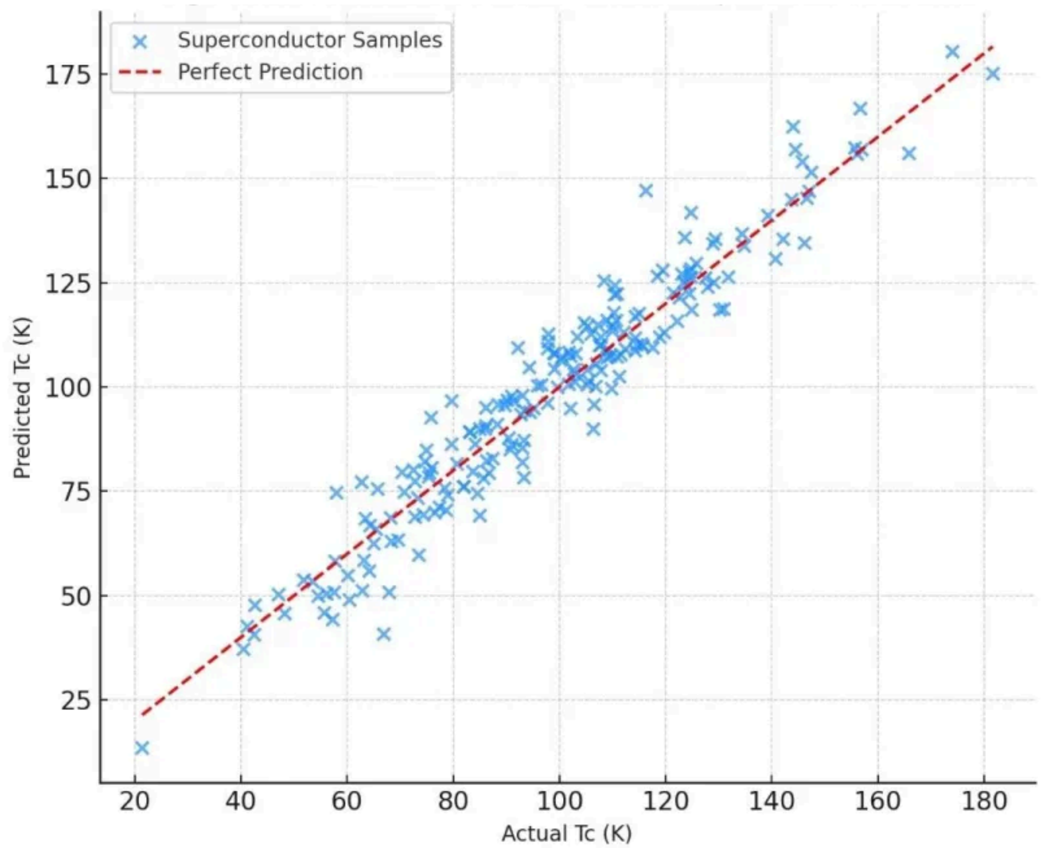


Figure 3: Predicted vs. Actual Critical Temperatures (Test Set). A scatter plot showing predicted vs. actual Tc. The diagonal line represents perfect prediction. Cuprates tend to occupy the higher Tc region (right side of the plot), with slightly more spread due to higher intrinsic variance.

As we can see in Figure 4, the residual distribution further confirms the robustness of the model. The histogram reveals an approximately Gaussian distribution centered near zero, with 95% of errors falling within ±15 K. This bell-shaped distribution indicates minimal systematic bias and confirms that the model's predictive errors are well within acceptable experimental margins.

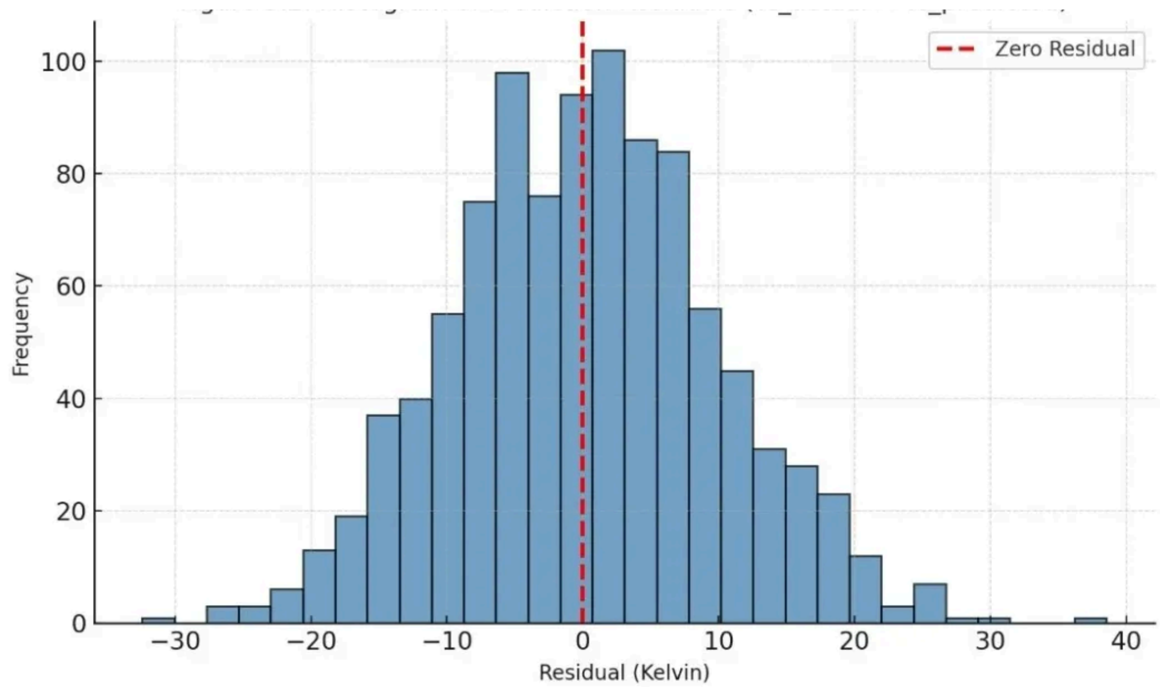


Figure 4: Histogram of Prediction Residuals (Tc_actual – Tc_predicted). *The bell-shaped curve indicates normally distributed residuals with minimal bias.*

Finally, we conducted feature importance analysis using SHAP values and permutation importance to interpret the behavior of the DNN. The most influential features included mean electronegativity, average valence electron count, atomic mass standard deviation, mean first ionization energy, and the range of atomic radii. These descriptors align closely with known physical mechanisms of superconductivity, including electron-lattice coupling (influenced by mass, radius, and electronegativity), electronic density of states (related to valence electrons), and phonon spectra (shaped by atomic bonding and lattice dynamics). The alignment between machine-learned predictors and established physics demonstrates that the model does more than memorize patterns: it encodes meaningful scientific relationships that enhance trust and interpretability.

Taken together, these results demonstrate that the proposed DNN framework not only achieves state-of-the-art predictive accuracy but also provides interpretable insights into the physical underpinnings of superconductivity. The combination of high performance, reliability, and interpretability highlights its potential as a scalable tool for accelerating the discovery of new superconducting materials.

6 Discussion

The results of our study demonstrate the significant potential of machine learning (ML)—and particularly deep learning models—in the field of superconductivity research. By accurately predicting the critical temperature (Tc) of a wide variety of superconducting compounds based solely on their chemical composition, our model provides a scalable, data-driven alternative to traditional experimental and theoretical approaches.

One of the central insights from this work is that compositional features alone—without the inclusion of structural, phononic, or quantum mechanical inputs—are sufficient to achieve state-of-the-art performance in Tc prediction. This finding suggests that a considerable portion of the information governing superconducting behavior is encoded in the elemental identities and combinations of the constituent atoms. While crystallographic details and microstructure undoubtedly play an important role, our results indicate that chemical composition can serve as an effective first filter for superconductor discovery. Moreover, the feature importance analysis revealed a striking correspondence between the most predictive variables and established physical mechanisms described by BCS and Eliashberg theory. Electronegativity and ionization energy are related to electronic band structure and charge transfer, atomic radius and mass influence lattice vibrations and phonon spectra, and valence electron count shapes the density of states at the Fermi level. This alignment between machine-learned patterns and physical theory enhances the interpretability of our model and suggests that data-driven approaches can reveal latent structure–property relationships that are otherwise difficult to uncover analytically.

The practical implications of these findings are significant. Experimentalists and materials scientists can now use AI models such as ours to rapidly screen and prioritize candidates for synthesis, particularly among high-Tc cuprate-like compounds. Given the expense of cryogenic measurements and the rarity of room-temperature superconductors, the model functions as an intelligent filter, pointing researchers toward materials most likely to meet desired thresholds, such as Tc values above 77 K for liquid-nitrogen-based applications. Furthermore, because this framework is built on open-access data and transparent code, it is both reproducible and extensible to other applications, including prediction of critical magnetic fields, superconducting gap energies, or the onset of superconductivity under pressure or doping conditions.

Despite its success, the current model has limitations that warrant careful consideration. First, because it is based solely on compositional features, it cannot distinguish between materials with identical formulas but different crystal phases, which may exhibit radically different superconducting properties. Second, the dataset is imbalanced, containing far more low-Tc non-cuprates than high-Tc cuprates, which may bias predictions toward conservative outcomes in high-temperature regimes. Third, measurement noise and inconsistencies across decades of experimental data introduce uncertainty into the training labels. Finally, while SHAP analysis provides interpretability, deep learning models retain an element of opacity, leaving open questions about whether they can produce falsifiable, theory-grounded hypotheses.

These challenges point toward several promising directions for future work. Incorporating crystal structure information, such as space group and lattice parameters derived from CIF files, could significantly improve accuracy. Transfer learning could be applied to smaller, emerging datasets—for example, newly discovered nickelates or hydrides—to improve generalization across classes of superconductors. Coupling predictive models with generative frameworks, such as variational autoencoders or diffusion models, would enable inverse design, where candidate materials are proposed based on desired Tc values. Finally, physics-informed neural networks (PINNs) represent an exciting frontier, embedding thermodynamic constraints or conservation laws directly into the learning process to bridge the gap between data-driven modeling and physical theory.

6 Conclusion

This study demonstrates the effectiveness of combining machine learning, materials chemistry, and superconductivity theory to address one of the most enduring challenges in condensed matter physics: predicting the critical temperature of superconducting materials. By training a deep neural network on a dataset of more than 21,000 superconductors (characterized solely by their chemical composition) we achieved high predictive accuracy (MAE $\approx$ 4.88 K, $R^2 \approx 0.918$), outperforming traditional models such as linear regression, support vector machines, and tree-based algorithms. The success of this model demonstrates that composition-based descriptors carry significant predictive power, even in the absence of detailed structural or quantum mechanical data.

Importantly, the features identified as most influential, such as electronegativity, valence electron count, and atomic mass variance,align well with established physical phenomena, including electron-phonon coupling and Cooper pair formation. This provides a level of interpretability often missing in black-box machine learning models and reinforces the notion that AI, when guided by physics-informed features, can be both predictive and explanatory. Our framework therefore offers a scalable tool for superconductor discovery, particularly in identifying promising high-Tc cuprate candidates where experimental synthesis is expensive and time-intensive. Beyond Tc prediction, this work lays the foundation for multi-property prediction, inverse design, and physics-informed generative modeling for next-generation materials.

In an era where the discovery of room-temperature superconductors is considered one of the "holy grails" of science, this research represents a meaningful step forward—uniting the precision of physics with the speed and scalability of artificial intelligence.

References

Allen, P. B., & Dynes, R. C. (1975). Transition temperature of strong-coupled superconductors reanalyzed. *Physical Review B, 12*(3), 905–922. https://doi.org/10.1103/PhysRevB.12.905

Bardeen, J., Cooper, L. N., & Schrieffer, J. R. (1957). Theory of superconductivity. *Physical Review, 108*(5), 1175–1204. https://doi.org/10.1103/PhysRev.108.1175

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Combescot, R. (2022). *Superconductivity: An introduction*. Cambridge University Press.

Cooper, L. N. (1954). Bound electron pairs in a degenerate Fermi gas. *Physical Review, 104*(4), 1189–1190. https://doi.org/10.1103/PhysRev.104.1189

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://www.deeplearningbook.org/

Hamidieh, K. (2018). A data-driven statistical model for predicting the critical temperature of a superconductor. *Computational Materials Science, 154,* 346–354. https://doi.org/10.1016/j.commatsci.2018.07.030

Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., & Persson, K. A. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials, 1*(1), 011002. https://doi.org/10.1063/1.4812323

Jain, S. (2025, July 11). Linear regression in machine learning. *GeeksforGeeks*. https://www.geeksforgeeks.org/machine-learning/ml-linear-regression/

Jha, D., Ward, L., Paul, A., Liao, W.-K., Choudhary, A., Wolverton, C., & Agrawal, A. (2018). ElemNet: Deep learning the chemistry of materials from only elemental composition. *Scientific Reports, 8,* 17593. https://doi.org/10.1038/s41598-018-35934-y

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR).* https://arxiv.org/abs/1412.6980

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)* (pp. 4765–4774). https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

McMillan, W. L. (1968). Transition temperature of strong-coupled superconductors. *Physical Review, 167*(2), 331–344. https://doi.org/10.1103/PhysRev.167.331

National Institute for Materials Science. (n.d.). *SuperCon database – Japan.* https://supercon.nims.go.jp/index_en.html

Ward, L., Agrawal, A., Choudhary, A., & Wolverton, C. (2016). A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials, 2,* 16028. https://doi.org/10.1038/npjcompumats.2016.28

Xie, T., & Grossman, J. C. (2018). Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters, 120*(14), 145301. https://doi.org/10.1103/PhysRevLett.120.145301

More Than Chatbots: AI's Growing Role in Mental Health Care

Heet Jani, Sashti Kandaswamy

Artificial intelligence is rapidly expanding its role in mental health, from conversational chatbots like Woebot and Wysa to diagnostic tools that analyze language, voice, and behavior. These technologies offer new opportunities for early detection, intervention, and emotional support, yet they also raise concerns about bias, privacy, and overreliance on machines. While AI can complement clinical care and increase access, it cannot replace the trust and nuance provided by human professionals. To ensure responsible use, AI in mental health must prioritize transparency, equity, and ethical safeguards while supporting—not substituting—the therapeutic relationship.

Keywords: *Artificial intelligence; mental health; chatbots; diagnosis; ethics; bias*

1 Introduction

Artificial intelligence is no longer limited to only self driving cars, or predicting algorithms--it's now aiding in therapy sessions, analyzing brain scans and even finding early signs to anxiety, depression and more (National Institute of Mental Health, 2021). This new role of AI in detecting various psychiatric disorders such as neurodegenerative disorders, is directly associated with the functionality of AI to diagnose and intervene in mental health disorders (World Health Organization, 2021). John McCarthy once affirmed that the sole purpose of AI was to develop machines in a way that they would be seen as intelligent–almost identical to humans (Topol, 2019). While AI-powered chatbots like Woebot and Wysa do have the roles of maintaining positive emotion regulation, detecting autism spectrum disorders, and delivering on-demand emotional support, this new technology's role in mental health extends far beyond automated conversations, often in the most surprising ways (Stanford Institute for Human-Centered AI, n.d.).

2 Chatbots and Emotional Support

Some of the most accessible and widely used AI tools are conversational chatbots that are particularly designed to support users with emotional support on demand. Woebot is a

CBT-Based Chatbot (one that uses cognitive behavioral therapy) techniques to provide support for those who suffer from anxiety, depression and even many unrevealed challenges that many users face. A peer-reviewed study found that users who interacted with Woebot daily for around two weeks, showed a significant reduction in the symptoms that cause these mood disorders (American Psychological Association, 2022). Similarly, Wysa is an AI-powered chatbot that uses evidence-based techniques--both CBT and DBT for emotional resilience and mental fortitude. Research has shown that Wysa significantly improved the symptoms that were associated with anxiety and depression, especially when those with affective disorders were paired with human coaches (National Institute of Mental Health, 2021). Used by over millions of people worldwide, both Wysa and Woebot have gained credibility through partnerships with healthcare providers and employers. Both applications interact like a friendly digital companion, especially through riveting towards the idea of "Reframe Negative Thoughts," which is a powerful concept in cognitive behavioural therapy. These cognitive restructuring tools may not replace therapy, however their stigma-free environment makes them a convenient tool for early intervention of diagnosis of many mood disorders.

3 AI in Diagnosis

Artificial intelligence is significantly transforming mental health diagnosis by offering faster, data-driven insights that also primarily focus on cognitive patterns (Topol, 2019). Diagnostic tools often rely on self-reporting and time-intensive evaluation, which in real time can delay care. AI however, can analyze speech text, facial expressions and behavioural data to flag early symptoms for depression, anxiety, or PTSD with near perfect accuracy (IBM Research, 2021). While these tools don't necessarily replace human clinicians, they can act as early detection systems which can help medical providers intervene sooner, and personalize treatment based on results.

AI is no longer a future promise in mental health diagnosis—it is already being used in real-world tools designed to detect early signs of psychological distress with greater speed and precision (National Institute of Mental Health, 2021). IBM's speech analysis models, for instance, examine linguistic patterns such as word choice and rhythm to identify early indicators of depression or psychosis, often before symptoms are outwardly visible (IBM Research, 2021). Similarly, apps like Mindstrong track cognitive shifts by analyzing how users type and interact with their phones (Mindstrong, n.d.), while Ellipsis Health uses voice tone and cadence to assess levels of anxiety or depression (Ellipsis Health, n.d.). These tools offer scalable, passive monitoring solutions that can support earlier intervention, particularly in environments where access to mental health professionals is limited.

4 Ethical Concerns Raised

As AI tools become more common in mental health care, concerns about bias and privacy have grown. Many systems are trained on limited datasets, which can lead to inaccurate assessments—especially for people from underrepresented groups. A 2019 study by Obermeyer et al. found that a healthcare algorithm widely used in the U.S. showed racial bias, underestimating the needs of Black patients (Obermeyer et al., 2019). Privacy is another issue: mental health apps often collect sensitive data, yet many don't follow strict medical privacy standards. Mozilla's 2022 report revealed that apps like BetterHelp and Talkspace shared user data with third parties, often without clear disclosure (Mozilla Foundation, 2022). These issues highlight the need for AI systems that are more ethical, inclusive, and transparent (American Psychological Association, 2022).

AI systems also lack transparency. When a tool flags someone as high-risk without explaining why, it creates confusion for both users and clinicians (World Health Organization, 2021). In a field built on trust and understanding, this lack of clarity can be harmful. Finally, while AI can offer support, it's no substitute for trained professionals. Relying too heavily on AI—especially in serious or crisis situations—can delay real help. To be effective and ethical, AI in mental health must be transparent, culturally aware, and used as a tool to support—not replace—human care (Stanford Institute for Human-Centered AI, n.d.).

References

American Psychological Association. (2022, June). *Ethical considerations for AI in mental health*. APA Monitor. https://www.apa.org/monitor/2022/06/ai-mental-health

Ellipsis Health. (n.d.). *The true voice of mental health*. https://www.ellipsishealth.com

IBM Research. (2021, October 5). *Can AI help mental health care?* IBM Research Blog. https://research.ibm.com/blog/ai-mental-health

Mindstrong. (n.d.). *Science meets technology*. https://www.mindstrong.com

Mozilla Foundation. (2022, May 2). *Mental health apps: Not-so-great expectations*. Privacy Not Included. https://foundation.mozilla.org/en/privacynotincluded/articles/mental-health-apps-2022/

National Institute of Mental Health. (2021). *Using artificial intelligence to advance mental health care.* https://www.nimh.nih.gov/news/science-news/2021/using-artificial-intelligence-to-advance-mental-health-care

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Stanford Institute for Human-Centered AI. (n.d.). *How AI can support mental health*. https://hai.stanford.edu/news/how-ai-can-support-mental-health

Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. World Health Organization. https://www.who.int/publications/i/item/9789240029200

TransforMerger: A Review of Transformer-Based Voice-Gesture Fusion for Robust Human-Robot Communication

Bhargavi Nigam, Rohith Deshamshetti

TransforMerger is a transformer-based system created to improve the flexibility of human-robot interaction (HRI) by taking in multimodal inputs (voice and gesture) and fusing these inputs into a single sentence. This employs probabilistic embeddings and takes a contextual scene to handle multimodal ambiguity. This sentence is sent to a Large Language Model (LLM) for processing. The experiment was conducted under real-world environments, demonstrating its robustness to misalignment, ambiguity, and missing information as it outperformed traditional baselines (especially in scenarios requiring more contextual knowledge). Thus, highlighting its potential to advance multimodal communications in HRI scenarios and its potential to advance various disciplines.

Keywords: *transformer models; human-robot interaction; multimodal fusion; probabilistic embeddings; large language models; contextual disambiguation; real-world experimentation*

1 Introduction

Human communication is an inherited multimodal system, combining speech, gestures, gaze, and facial expressions. However, (HRI) human-robot interaction systems often rely on

rigid communication constrained to single modalities. Existing multimodal approaches often naively fuse inputs, limiting their adaptability (Wang et al., 2024). A method where a context-aware multimodal merging algorithm incorporates transformer-based large language models (Wolf et al., 2020). Uncertain multimodal inputs, updating action probabilities based on simultaneous observation. This allows the system to resolve ambiguity, assess action feasibility, and improve robustness to noise and misalignment. TransforMerger is a context-aware model for merging multimodal data, showing improved robustness to noise misalignment, and is capable of resolving input ambiguities using contextual knowledge and by grounding object attributes in scene context. An evaluation on simulated and real-world dual-modality (gesture and language) datasets, analyzing the impact of different noise types. It is More Inclusive Human-Robot Interaction (HRI).

People who struggle with clear speech, have strong accents, or use non-standard phrasing can still interact effectively with robots, as the system is designed to manage noise, misalignment, and incomplete commands. This capability enables higher reliability in real-world tasks, allowing robots to correctly interpret ambiguous phrases such as “pick that red object” by integrating voice cues with pointing gestures and scene understanding, thereby reducing costly or dangerous errors in environments like warehouses, hospitals, and disaster zones. It also enhances efficiency and safety by lowering the risk of unsafe movements or unintended actions, a crucial factor in manufacturing, healthcare, and collaborative robotics. Moreover, these advances support adaptability across domains and lay the foundation for socially aware robots, as context-aware reasoning and probabilistic understanding allow machines to infer meaning from partial cues in ways that mirror human communication.

2 Problem

A key challenge in multimodal perception is handling uncertainty arising from sensor noise, speech recognition errors, and ambiguous gestures. Traditional probabilistic models, such as Bayesian networks (Bishop & Nasrabadi, 2006), hidden Markov models (HMMs) (Rabiner, 1989), and probabilistic graphical models (Starner, Schiele, & Pentland, 1998), have been employed to mitigate these issues. However, these methods rely on predefined rules and do not incorporate contextual reasoning, making them ineffective in cases where multiple references (e.g., pointing gestures) lack explicit grounding**.** Recent advances in large language models (LLMs) have introduced powerful reasoning capabilities for context-aware decision-making (Brown et al., 2020). Models like CLIP and Flamingo (Alayrac et al., 2022) can match images with text, but they can't combine gestures and speech for robot tasks. That’s why this new system was developed — to handle both types of input, even if they’re out of sync or unclear.

3 Proposed Solution Method

TransforMerger proposes a novel solution to the inconvenience and rigidity of HRI: a transformer-based model that merges voice and gesture inputs (multimodal inputs) in a probabilistic, context-aware manner to produce natural and structured robot commands—Skilled Commands—for manipulation tasks. This approach ensures an efficient fusion of multimodal inputs while maintaining temporal and contextual dependencies. Thus, mitigating misalignment and noise errors improves task performance.

A. Merging Algorithm: The model initially collects voice and gesture data from the user and then utilizes probabilistic preprocessing embeddings. This method converts both modalities, or in other words, forms of input, into probabilistic word sequences while acknowledging uncertainty in input/vague noise inputs. The TransforMerger model then merges the two inputs into a single sentence, temporally sorted by time stamps, with each word embedded probabilistically.

B. Foundational Reasoning Model with Soft Embeddings: This solution utilizes a transformer-based SOTA Large Language Model (LLM) built on a Causal Transformer architecture to reason over merged sentences/ the user's intent and context to output a skilled command for the robot to interpret and execute using its predefined methods. The study introduced soft embeddings to enforce the model to reason probabilistically over the inputs and generate a parametrized prompt to constrain the model and provide contextual information.

C. Scene Embedding: This function is particularly important to TransforMerger's functionality. TransforMerger incorporates scene embeddings by providing the LLM with a fixed scene representation (O). This includes a list of objects and their properties, which is fed into the LLM as a prompt. These embeddings help with 2 aspects. 1) Grounding pointing gestures: uses objects to clarify where a pointing gesture is directed. 2)Enhancing contextual awareness in the reasoning model: By supplying scene data, the reasoning model can interpret vague instructions. However, when some properties are missing or unspecified, it can rely on commonsense reasoning(Vanc & Stepanova, 2025).

TransforMerger, thus, can now effectively infer user intent while accounting for ambiguity and error, as discussed in scene embedding, significantly enhancing the robustness of multimodal Human-Robot Interaction (HRI).

4 Experimental Setup

In real-world experiments, we use a Franka Emika Panda robot with an Intel RealSense D455 camera for object perception. Gestures are tracked using a Leap Motion sensor and processed with the Gesture Toolbox (Sec. V-A), while voice commands are captured via a microphone and processed using the Whisper model (Vanc & Stepanova, 2025).

A. Model Benchmarking and Comparisons

We compare three transformer-based models—EXAONE 3.5, SmolTulu 1.7B, and Granite 3.1 2B—selected for their high scores on IFEval (instruction-following) and BBH (common-sense reasoning). These models, fine-tuned with Chain-of-Thought (CoT) prompting, are sourced from the Hugging Face Forum. As a baseline, we use an Argmax method, which selects the most probable word at each step and builds the command based on the first detected elements.

B. Language Model Parameters

The LLM uses tunable parameters: Temperature = 0 (ensures precise, focused output), Top-p = 1 (balances creativity and accuracy), and a Repetition Penalty = 1.1 (prevents repeated words or actions) (Vanc & Stepanova, 2025). These settings optimize structured command generation.

C. Actions and Object Set

1) Objects: Real-world objects include a cleaner, a bowl, a cup, a drawer, and tomatoes. Simulated objects include a cup, cube, plate, table, can, box, etc. Each object has properties like size (small, medium, large), color (red, green, blue), and state (open, closed, half-full).
2) Actions: The system supports 12 actions: zero-object (stop, release, home), single-object (pick, push, place, etc.), and double-object (pour, put). Actions can include prepositions (into, onto) and modifiers (quickly, slowly, etc.).
3) Robotic Skills: Each action is linked to a skill learned from demonstration (LfD) using a framework from TU Delft and implemented in ROS2. Objects are localized via SIFT feature matching, and trajectories are aligned in real-time based on the generated Skill Command.

D. Multimodal Artificial Dataset

An artificial dataset was created to test the system under noise and misalignment. It includes scene descriptions, gesture (SG), and voice commands (SV) with added phonetic errors, filler words, timestamp shifts, and sentence truncation. Noise is controlled using parameters like Nphon, Pfiller, Nalign, and Pincomplet. Each scene has labeled objects with properties, and gestures and speech are probabilistically modeled to simulate real-world uncertainties (Vanc & Stepanova, 2025).

5 Results

First, we evaluate the models on the simulated dataset, analyzing the impact of individual noise types on their performance. Second, we conduct a real-world experiment across five different scenarios.

A. Noise Experiment

In the noise experiment, models were tested on a simulated dataset with increasing levels of phonetic noise (Nphon), filler words (Pfiller), and missing words (Pincomplet). While Argmax performed best at zero noise, TransforMerger with the Granite model surpassed it as noise increased, maintaining 40% accuracy at high noise levels. SmolTulu performed the weakest, though still acceptable in low-noise settings. When testing alignment noise (Nalign), Argmax's performance dropped significantly, while Granite and EXAONE remained highly accurate, showing TransforMerger's robustness to temporal misalignment
(Vanc & Stepanova, 2025).

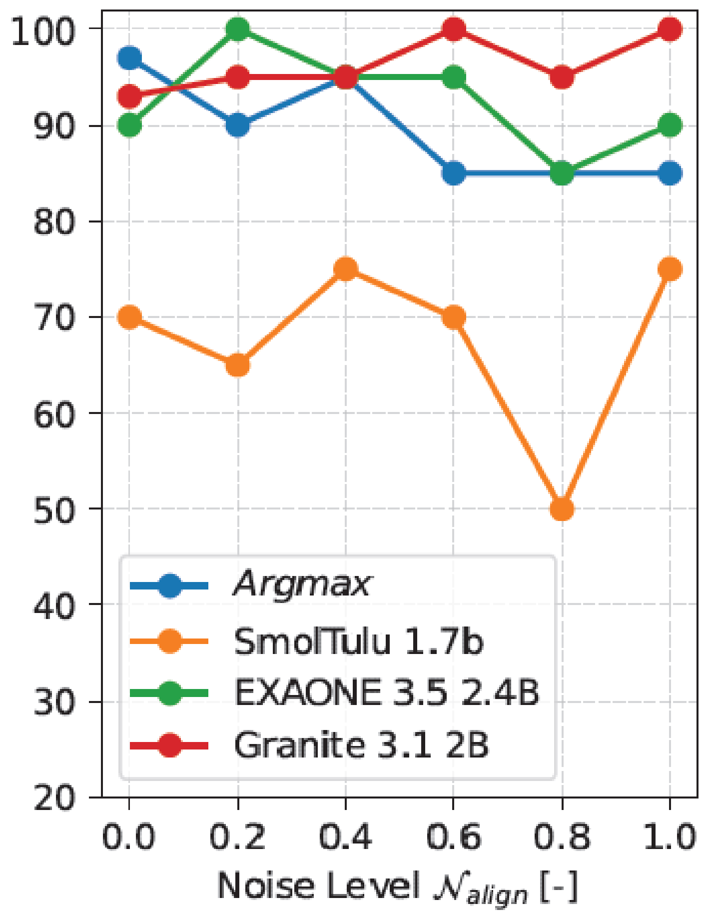


Figure 1: Model robustness across varying alignment noise levels. Performance is shown for Argmax, SmolTulu 1.7b, EXAONE 3.5 2.4B, and Granite 3.1 2B. While Argmax, EXAONE, and Granite maintain consistently high accuracy (≈90–100%) across noise levels, SmolTulu exhibits substantial performance degradation, indicating lower resilience to alignment noise.

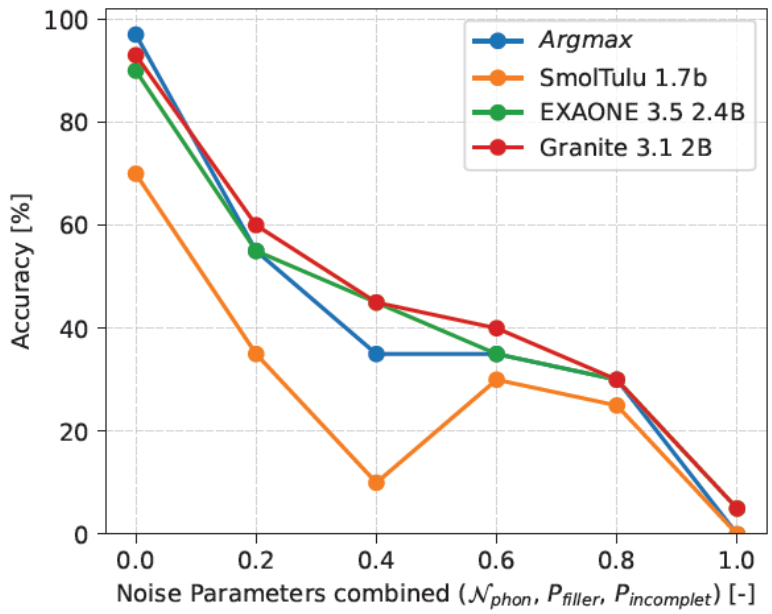


Figure 2: Model accuracy under combined noise parameters. Accuracy decreases as phonetic noise, filler words, and incomplete inputs are introduced simultaneously. Granite 3.1 2B and EXAONE 3.5 2.4B maintain higher robustness compared to Argmax and SmolTulu 1.7b, which exhibit sharper performance degradation, especially at moderate noise levels.

B. Real Experiment

In the real-world experiment, the system was tested across four scenarios (T1–T4), each repeated 10 times. (Vanc & Stepanova, 2025). Tasks ranged from handling noisy inputs (T1), resolving ambiguous object descriptions (T2), interpreting context-heavy two-object commands (T3), and recovering from vague instructions (T4). Results showed that Granite consistently outperformed other models, achieving full success in T1 and strong results in all scenarios, even under noise. EXAONE followed closely, while SmolTulu underperformed, sometimes even worse than the simple Argmax baseline. Notably, Granite succeeded even with single-modality input (gesture or voice), showing strong robustness. In contrast, Argmax failed in tasks requiring contextual reasoning or when the input was ambiguous or incomplete.

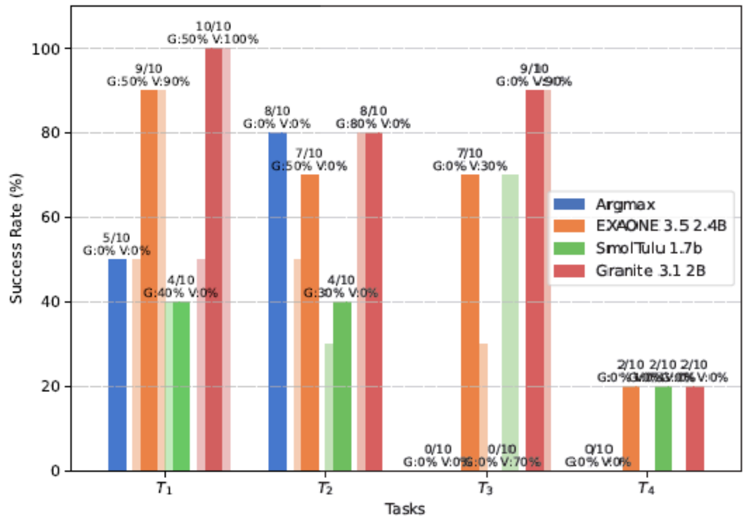


Figure 3: Success rates across four tasks (T1–T4) for Argmax, EXAONE 3.5 2.4B, SmolTulu 1.7b, and Granite 3.1 2B. Granite achieves the highest performance, reaching 100% in T1 and 90% in T3. EXAONE performs strongly in T1, while Argmax shows moderate success. SmolTulu underperforms across most tasks, and all models struggle with T4.

6 Limitations

While the TransforMerger doesn't simply fuse multimodal inputs and instead dynamically integrates uncertain inputs, the model still struggles with certain cases, failing to fully adhere to specific reasoning rules or accurately interpreting probabilistic results in ambiguous scenarios. Its success, to a good extent, relies on quality multimodal inputs: if given highly ambiguous data, the model's reasoning fluctuates. Additionally, while Large Language Models (LLMs) help us leverage their powerful reasoning capabilities to produce executable, skilled commands, they are prone to producing inaccurate commands.

Although the system has many benefits, it also raises some ethical concerns. Voice, gesture, and scene data must be protected to avoid privacy violations or inappropriate use since they can reveal information about a person's speech, movements, and surroundings. The technology can be tricked by false inputs, and depending on it without user checks might lead to mistakes in sensitive situations. If access is limited to wealthy industries or regions that can access it, the gap between different groups could grow. To prevent these problems, the system should have strong privacy protections, clear user consent, open decision-making, and safeguards against harmful or accidental use.

7 Conclusion

In this work, the authors present TransforMerger, a robust system for multimodal human-robot interaction that fuses speech and gesture inputs using a contextual reasoning framework (Vanc & Stepanova, 2024). The challenges, such as input noise, temporal misalignment, and linguistic ambiguity by combining probabilistic representations, gesture tracking, and language model-based inference. In both simulated and real-world experiments, TransforMerger demonstrates strong adaptability and flexibility, especially under ambiguous and Incomplete Instructions. The system significantly outperforms baseline methods and proves capable of executing complex robotic tasks even when commands are ambiguous. This work highlights the promise of integrating large language models(LLMs) and soft multimodal fusion techniques for improving the adaptability and intelligence of future interactive robotic systems. The broader impact of TransforMerger relies on its potential to make human-robot interaction more accessible, natural, and effective. By supporting multimodal input, this system lowers the barrier of human-robot communication, which is especially valuable in fields like healthcare, surgery, senior care, education, space exploration, autonomous technologies, and thousands more. This step can inspire more advanced multimodal systems as well, which can move us closer to advanced and life-like human-centric robotics.

References

Vanc, P., & Stepanova, K. (2024). *TransforMerger: Transformer-based voice-gesture fusion for robust human-robot communication* [Preprint]. arXiv. https://arxiv.org/abs/2504.01708

Wang, C., Hasler, S., Tanneberg, D., Ocker, F., Joublin, F., Ceravola, A., Deigmoeller, J., & Gienger, M. (2024). *LaMI: Large language models for multi-modal human-robot interaction* [Preprint]. arXiv. https://arxiv.org/abs/2401.15174

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, X., ... & Rush, A. (2020, October). Transformers: State-of-the-art natural language processing. In Q. Liu & D. Schlangen (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38–45). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4). Springer. https://link.springer.com/book/9780387310732

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257–286. https://ieeexplore.ieee.org/document/18626

Starner, T., Schiele, B., & Pentland, A. (1998). Visual contextual awareness in wearable computing. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215)* (pp. 50–57). IEEE. https://ieeexplore.ieee.org/document/729529

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 1877–1901). https://arxiv.org/pdf/2005.14165

Alayrac, J.-B., Donahue, J., Luc, P., et al. (2022). Flamingo: A visual language model for few-shot learning. In *Advances in Neural Information Processing Systems* (Vol. 35, pp. 23716–23736). https://arxiv.org/pdf/2204.14198

BIOTECHNOLOGY & LIFE SCIENCES

Roles of Biotechnology in Understanding Tumor Microenvironments

Ella Tam

The tumour microenvironment (TME) is an increasingly recognised factor in cancer development, tumour progression and therapeutic resistance, comprising immune cells, stromal fibroblasts, vasculature and extracellular matrix components. TME significantly influences tumour behaviour through cell–cell interactions and molecular signalling. Recent advanced biotechnology has increased our understanding by incorporating TME profiling into techniques like microscopy techniques, cell culture models and molecular analysis methods. These innovations have not only deepened our knowledge, but also allowed us to establish new therapeutic targets and predict treatment responses especially in immunotherapy. This article aims to provide an overview of TME, the common techniques involved in TME profiling as well as to demonstrate the process from overcoming challenges like tumour heterogeneity and therapeutic resistance, to ultimately improving patient outcomes.

Keywords: *Tumour microenvironment (tme); cancer progression; therapeutic resistance; immunotherapy; cell–cell interactions; molecular signalling; biotechnology;*

1 Introduction

Cancer is a world-impacting disease caused by both genetics and lifestyle choices with over 35 million new cancer cases predicted in 2050, a 77% increase from the estimated 20 million cases in 2022 (World Health Organization, 2024).

It is believed that cancer cells interact with neighboring cells to support cancer cell survival, local invasion and metastatic dissemination. This interaction is called Tumour Microenvironment (TME) which comprises malignant cells, immune cells, stromal cells, blood vessels and extracellular matrix as the hallmark features of every TME. Hence, it is established that a tumor is not simply a group of cancer cells, but rather a heterogeneous collection of infiltrating and resident host cells, secreted factors and extracellular matrix. This dynamic and reciprocal relationship benefits tumour survival by for example overcoming a hypoxic and acidic microenvironment as the TME coordinates a program that promotes angiogenesis to restore oxygen and nutrient supply and remove metabolic waste (Anderson & Simon, 2020). According to more studies, it has been well documented that TME plays a critical role in not only tumor survival and the areas that have been mentioned previously, but surprisingly, the promotion of drug resistance and even the maintenance of a cancer stem-like phenotype. TME formation acts as a reflection of a tumour's own organization during the different stages of its development and because of this, scientists are able to develop several therapeutic approaches targeting primary TME. A famous example is the use of systematic analysis of TME adjacent to the tumour mass to establish the proportion of myofibroblasts-like cancer-associated fibroblasts (CAFs) which corresponds to the stemness and metastases-promotion (Hernández-Camarero et al., 2021).

In recent years, TME elements and its signalling pathway as a therapeutic target in cancer has attracted great research and clinical interest (Hernández-Camarero et al., 2021). Along with advances in biotechnology, scientists and pathologists are able to better understand TME to unlock new diagnostic and therapeutic strategies (Xiao & Yu, 2020). This article aims to review the vital role of biotechnology in understanding TME including biotechnology tools and their contribution in pathology and treatment.

2 About Tumor Microenvironment

The Tumor Microenvironment refers to a highly dynamic and heterogeneous environment that surrounds tumour cells. It includes not only malignant cells, but also a wide range of non-cancerous components such as immune cells, cancer-associated fibroblasts, endothelial cells that form blood vessels and components of the extracellular matrix (ECM) (Anderson & Simon, 2020). These components interact in complex ways, all through direct cell to cell contact and via secreted signalling molecules such as cytokines, chemokines and growth factors. Together, these interactions influence aspects from proliferation and metastasis, to immune evasion and drug resistance (Zhou et al., 2023).

Table 1. *Key components of the tumour microenvironment and their functions (Arneth, 2019).*

Cell Players	**Main Markers or Types**	**Primary Functions**
T lymphocytes	$CD8^+$ and $CD4^+$	Some are protumorigenic, while others are tumor restrictive.
B lymphocytes	Regulatory B cells and B10 cells	They contribute to the regulation of tumor cell survival and proliferation and the

Cell Players	Main Markers or Types	Primary Functions
		development of treatment resistance. In addition, these cells have been linked to the process of fostering immune escape
NK and NKT cells	NKG2 receptors, Ly49 receptors, NK1, CD94, C57BL/6, CD161, NKG2D, CD56, and NKG2A	NK and NKT cells use inhibitory, adhesion, activating, and cytokine receptors to identify cellular targets and healthy spare cells
Macrophages	M1 and M2 macrophages	They create a stroma that is supportive of neoplastic cell invasion and expansion
Macrophages M1	antitumorigenic	
Macrophages M2	immunosuppressive and pro-tumorigenic	As M2 macrophages are immune-suppressive, they can promote tumor progression
Cancer-associated fibroblasts	α-Smooth muscle actin, fibroblast activation protein, vimentin, desmin, and PDGFR α and β	They contribute to tumor cell proliferation by maintaining continuous propagation and growth signals at primary and metastatic sites
Cancer stem cells	Tumor stem cells and DPSCs	They support tumorigenesis through unique homing abilities to primary and metastatic sites
Chemokines	CXCL14 and CXCL12	They are usually overexpressed on myofibroblasts and myoepithelial cells. These molecules can bind epithelial cell receptors to increase cell migration, invasion, and proliferation
Integrins	αMβ2, αXβ2, αLβ2, αDβ2, α4β7, and αEβ7	They bind to the extracellular matrix in the TME
Selectins	Epidermal growth factor (EGF)-like motif, ST3Gal6, P-selectin	These are vital vascular adhesion molecules that affect the development of cells

Cell Players	**Main Markers or Types**	**Primary Functions**
Cadherins	Protocadherin, desmogleins, and desmocollins	These molecules mediate the formation of homophilic bonds in a calcium-dependent manner
Tregs	CD4, FOXP3, and CD25	These cells promote the generation and function of vaccine-elicited CD8+ memory T cells
Immunoglobulin superfamily (IgSF)	Cell surface antigen receptors, coreceptors, and costimulatory molecules	These molecules mediate the formation of both heterophilic and homophilic bonds
Bone marrow derived cells (BMDC)	BMDCs have several tumor growth-promoting functions.	Tumor growth promoting functions include expression of growth factors, promotion of tumor vessel formation and creation of tumor stem cell niches
Myeloid derived suppressor cells (MDSC)	MDSCs expand in pathological situations such as cancer, as a result of an altered haematopoiesis	MDSCs possess strong immunosuppressive activity especially on myeloid cells.

Immune cells are recognized as critical components of TME as they can either suppress tumor growth or promote it. One of the most studied aspects of the TME is its role in immune suppression as immunotherapies and immune checkpoint inhibitors are novel therapeutic modalities for advanced cancers. However, some patients are resistant to these therapies due to the mechanisms underlying tumor immune resistance. This involves a number of immunosuppressive cells such as tumor-associated macrophages (TAM) and regulatory T cells (Tregs) to prevent immune recognition and clearance (Cheng, Bai, Shu, Ahmad, & Shen, 2021). TAM are proposed to be the most abundant cell type in TME, coordinating the immunosuppressive microenvironment (Rajbhandary, Dhakal, & Shrestha, 2023). Inflammatory M1-macrophages and immunosuppressive M2-macrophages are the two main sub-types of macrophages where M2 phenotypes are polarized from normal macrophages. M2 are demonstrated to express co-inhibitory molecules and release anti-inflammatory cytokines as well as matrix metalloproteinases to enhance tumor progression and orchestrate its development and metastasis. On the other hand, Tregs are differentiated from traditional T cells and divided into two sub-groups - naturally occurring Tregs (nTregs) and induced-to-adjust T cells (iTregs). iTregs inhibit the anti-tumor immune response of effector T cells and DCs, resulting in tumor progression. In addition, cytokines and factors such as vascular endothelial growth factor (VEGF), secreted by tumor cells or these immunosuppressive cells – Tregs and M2 polarized macrophages, also mediate the tumor progression and immune escape of cancers by supporting angiogenesis. It is also known that cancer cells are adapted to hide from immune recognition by decreasing the expression of neoantigens and antigen presentation molecules or upregulating the expression of immune checkpoints on immunosuppressive cells (Tie, Tang, Wei, & Wei, 2022). All these components therefore form an immunosuppressive microenvironment which creates a so-called "cold tumour". A variety of reasons can make a tumour immune cold. For example having a characteristically increased expression of myeloid derived stromal cells (MDSC), M2 macrophages, PMNs, Tregs and Th17 cells and is therefore less responsive to immunotherapies (Rajbhandary, Dhakal, & Shrestha, 2023).

It is also important to know that the presence of immunosuppressive TMEs is also created by high levels of reactive oxygen species, a dense ECM, acidity and tumor hypoxia (reduced partial pressure of oxygen) (Feng et al., 2024). Rapidly proliferating tumor cells create a hypoxic microenvironment that induces tumor angiogenesis. The rapidly growing blood vessels comprising the neovascular network are often disorganized, leading to an inadequate oxygen supply which further increases tumor hypoxia. Hypoxia has been proven to be associated with immune suppression and an increase in resistance to chemotherapy and radiotherapy (Telarovic, Wenger, & Pruschy, 2021).

Understanding the cellular composition and molecular pathways of the TME is essential for improving diagnostic and therapeutic strategies. Modern biotechnological tools have become indispensable in profiling this complex environment, guiding the development of targeted and personalised treatments.

3 Biotechnology tools for studying TME

The make-up of TME and the interaction between tumour infiltrating immune cells and cancer cells can have a large impact on the response to immunotherapy as previously mentioned. Robust assessment of the TME with accurate and reproducible methods is vital to understanding mechanisms of immunotherapy resistance. Biotechnology tools such as advanced microscopy techniques, cell culture models and molecular analysis methods are key to unrevealing immunosuppressive TME (Rangamuwa et al., 2023).

3.1 (Multiplex) Immunohistochemistry / Immunofluorescence

Immunohistochemistry (IHC) /Multiplex immunohistochemistry (mIHC) can be used to detect a wide range of proteins(mostly called biomarkers) , including those expressed by tumor cells and cells within TME. For example, IHC is often used to assess the expression of predictive biomarker HER2 in breast cancer.

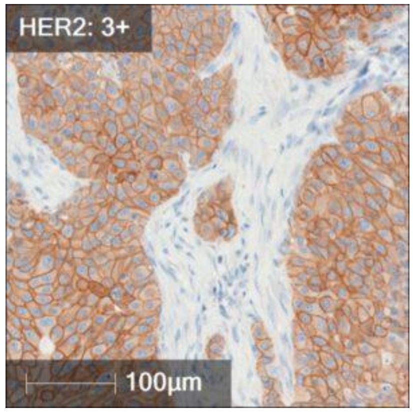


Figure 1: Immunohistochemical staining of HER2 in a clinical tissue sample reveals strong positive staining (3+) in the epithelial cells (dark brown), indicating overexpression of HER2 (Magaki, Hojat, Wei, So, & Yong, 2019).

IHC/mIHC provides information regarding the types of cell populations, their characteristics and the spatial relationship of different cell populations and tissue structures. It uses a primary antibody to target proteins of interest which usually reflect the type and function of cell populations. Then, a secondary antibody conjugated to an enzyme is then applied before applying a chromogen substrate of the enzyme that will develop a colour upon enzymatic activity. This method enables amplification of the signal for the detection of the target protein with light microscopy. Additionally, IHC can be performed effectively on small tissue samples (Rangamuwa et al., 2023). Setting up an IHC lab involves several key steps, including tissue preparation, antigen retrieval, antibody application, detection, and visualization (Magaki, Hojat, Wei, So, & Yong, 2019).

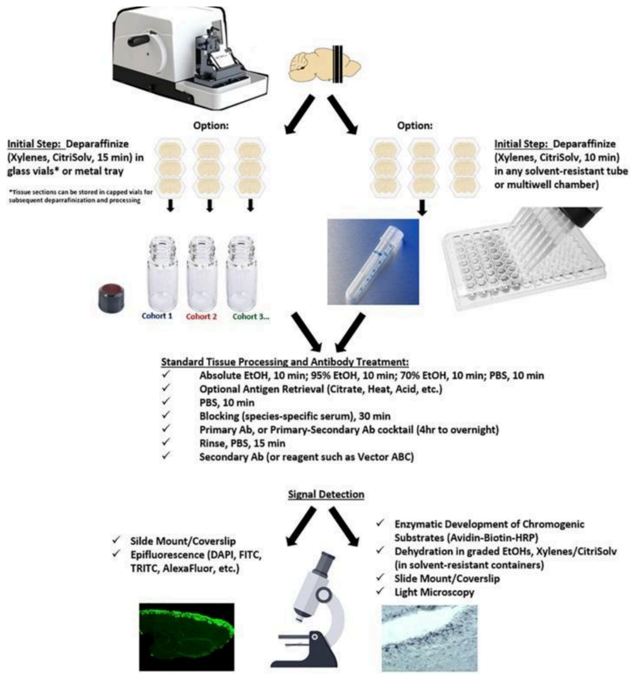


Figure 2: Workflow diagram outlining the standard steps of immunohistochemistry, including tissue preparation, antibody treatment, and signal detection methods (Muniz Partida & Walters, 2023).

Immunofluorescence (IF) uses a similar principle to IHC and can also be used to assess the TME as well. Here, instead of an antibody bound to an enzyme, antibodies are conjugated to fluorophores that can be detected by fluorescence microscopy. IF can either be performed directly where the primary antibody is attached to a fluorophore or indirectly where the fluorophore is attached to a secondary antibody which recognises the primary antibody of interest which allows for signal amplification (Characterization of the Tumor Microenvironment, 2023) (Rangamuwa et al., 2023). The resulting fluorescence signal can be visualized and quantified using specialized imaging equipment, allowing for a more precise and quantitative analysis of protein expression and spatial distribution within the tissue sample (Characterization of the Tumor Microenvironment, 2023).

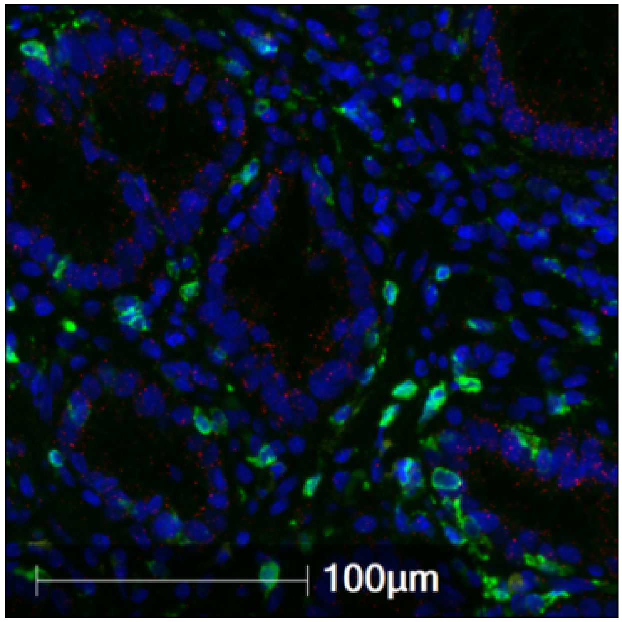


Figure 3: Immunofluorescence staining shows the spatial distribution of proteins and cell populations within the tumor microenvironment (Characterization of the Tumor Microenvironment, 2023).

3.2 3D bioprinting

From previous studies, researchers discovered that traditional 2D cell cultures fail to fully replicate the complete TME, while mouse tumor models suffer from time-consuming procedures and complex operations. This has led to an emerging technology called 3D bioprinting as a vital tool in studying TME. It is a revolutionary biomanufacturing technique that involves layer-by-layer stacking of biological materials, such as cells and biomaterial scaffolds, to create highly precise 3D biostructures (Li, Liu, Xu, & Wang, 2023).

According to the printing principle, biological 3D printing can be categorized into three types: extrusion-based bioprinting, droplet-based bioprinting, and photocuring-based bioprinting:

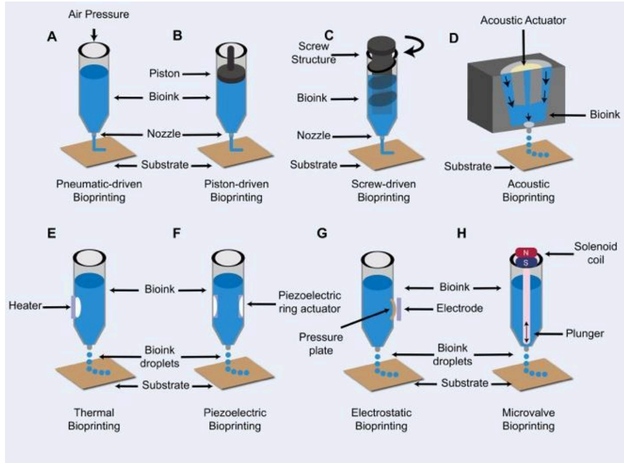


Figure 4: Different bioprinting techniques, including pneumatic-, piston-, screw-, acoustic-, thermal-, piezoelectric-, electrostatic-, and microvalve-driven methods, illustrating how bioink is deposited to create 3D biological structures (Li, Liu, Xu, & Wang, 2023).

3D bioprinting offers a more precise and dependable platform for drug screening and toxicity testing. Through the printing of models featuring tumor characteristics, researchers can evaluate the anticancer activity of drugs against tumor cells and study their impact on normal cells. Moreover, they can simulate the growth, diffusion, and invasion processes of tumor cells, leading to a more comprehensive understanding of tumor development and metastasis mechanisms and thereby establishing a stronger foundation for drug development and personalized treatments (Li, Liu, Xu, & Wang, 2023).

3D bioprinting technology has shown tremendous potential in breast cancer research for example, by creating more realistic breast cancer models to investigate the interaction between breast cancer cells and bone stromal cells. It has been agreed that this technology provides a suitable model with which to study the interactive effects of cells in the context of an artificial bone microenvironment and thus may serve as a valuable tool for the investigation of post-metastatic breast cancer progression in bone (Zhou et al., 2016).

3.3 RNA sequencing techniques

Scientists discovered that the genomes within tumours and their microenvironment are promising biomarkers for prognosis prediction. RNA sequencing (RNA-seq) has become a highly recognized tool to understand the interactions between cancer cells, immune subgroups and non-immune interstitial elements, hence providing a more complete genetic map than DNA sequencing. There are three main types of RNA-seq: bulk RNA-seq, single cell RNA-seq (scRNA-seq) and spatial RNA-seq (spRNA-seq) (Yan, Ju, Huang, & Yuan, 2024).

Bulk RNA-Seq provides an average measure of gene expression across the entire population of cells. It is the most widely used technique for measuring gene expression at the bulk sample level and has revolutionized the field of immunology research. It enables the comprehensive profiling of the expression of thousands of genes simultaneously, thereby providing insights into the immune transcriptional landscape in response to various stimuli. By using bulk RNA-seq, numerous novel gene fusions have been identified and utilized as diagnostic or prognostic markers and therapeutic targets in tumours, such as the *NUP98-PHF23* fusion gene in acute myeloid leukaemia (AML) (Yan, Ju, Huang, & Yuan, 2024).

Contrastingly, scRNA-Seq enables the analysis of gene expression at the single-cell level. This technology has revolutionized our ability to investigate cellular diversity and identify unique cell types, providing unprecedented insights into the complexity and heterogeneity of several tumours, such as gastric cancer, bladder cancer, and breast cancer, while also enabling the discovery of potential therapeutic targets and biomarkers for patient stratification. Despite the numerous advantages of scRNA-seq, it requires lysing individual cells, which can lead to loss of spatial and/or temporal information (Yan, Ju, Huang, & Yuan, 2024).

Luckily, SpRNA-seq is a high-throughput sequencing technique that allows for the profiling of gene expression with spatial resolution in tissue samples as it allows for the study of gene expression in a three-dimensional context, representing the next generation of RNA sequencing. This means that scientists are able to use this technique to acquire a comprehensive analysis of the TME and characterization of the spatial tumor heterogeneity (Wu et al., 2022). In a recent study on colorectal cancer, scientists were able to identify the presence of highly metabolically activated immunosuppressive $MRC1^{+}$ $CCL18^{+}$ M2-like macrophages in the metastatic sites, suggesting that TME had undergone significant spatial reprogramming during metastasis (Wu et al., 2022).

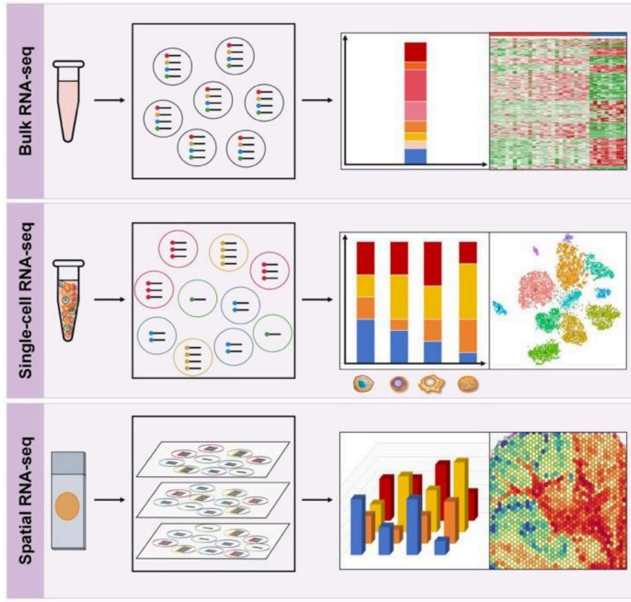


Figure 5: Comparison of bulk RNA-seq, single-cell RNA-seq, and spatial RNA-seq methods, illustrating differences in resolution and spatial mapping for analyzing tumor microenvironments (Yan, Ju, Huang, & Yuan, 2024).

4 Contribution in pathology and treatment

Understanding TME has become an essential aspect of cancer pathology and therapeutic development. Traditional pathology relies on the microscopic examination of tissue and imaging histopathology to classify tumours and determine malignancy grades and the process remains manual. The application of pathology images was mainly based on subjective evaluation, lack of structured processing of image data and insufficient mining of its hidden information (Pan, Feng, & Peng, 2022). However, it is now proven that a tumour's behaviour cannot be fully understood by analysing the cancer cells alone but by analysing the interactions between malignant cells and the surrounding microenvironment which influences tumour progression and how tumours respond to therapy. Therefore, integrating TME analysis into diagnostic pathology allows more accurate disease characterisation and clinical decision-making. It not only improves treatment efficacy, but also addresses therapeutic resistance and offers a nuanced approach to cancer therapy (Glaviano et al., 2025).

One of the areas that TME analysis contributes most to is immunotherapy. Earlier techniques like immune checkpoint inhibitors (ICIs) which include anti-PD-1 antibody, anti-PD-L1 antibody and anti-CTLA-4 antibody, have displayed considerable success in the treatment of malignant tumours. However, the therapeutic effect is still unsatisfactory for a portion of patients (Su et al., 2024). Patient responses remain highly variable as only a minority of patients outside these ‘responsive’ tumor types respond, with some totally resistant. Several studies have demonstrated that ICIs work best with “hot” tumours not “cold” due to the absence of lymphocytes from the tumor. Luckily, Biotechnology now enables strategic modulation of the TME to convert immunologically cold tumours into hot ones by targeting TME and performing techniques such as depleting immune suppressive cells, inhibiting transforming growth factor-beta, remodelling the tumor vasculature or hypoxic environment, strengthening the infiltration and activation of antigen-presenting cells and/or effector T cells in TME with immune modulators. With the help of biotechnology, the success of modifying cold tumours to overcome their resistance to ICIs represent mechanistically driven approaches that will ultimately result in rational combination therapies to extend the clinical benefits of immunotherapy to a broader cancer cohort (Wang, Wang, Desai, Trapani, & Neeson, 2020).

Furthermore, TME analysis is shown to provide information on prognosis (disease progression). For example, the spatial context of tumor-infiltrating immune cells (TIICs) is important in predicting colorectal cancer patients’ clinical outcomes by using mIHC. According to a study conducted in 2024 where scientists evaluated the nearest distance between the cancer cells and TIICs, they found that TIICs were distributed unevenly and its spatial distribution was closely related with the patient's prognosis, due to tumor heterogeneity (Shen et al., 2024). Similarly, stromal cells are also proven to correlate with tumor progression(Bussard, Mutkus, Stumpf, Gomez-Manzano, & Marini, 2016).

5 Conclusion

The tumor environment plays a fundamental role in cancer development, tumor progression and treatment responses. Hence, there has been a growing interesting in analyzing TME and it is now clear that apart from malignant cells, surrounding components like immune cells, stromal cells and vascular structures critically shape tumour behaviour. While traditional pathology is still essential, the growing field of biotechnology has been evolved to incorporate microenvironmental profiling into diagnostic and therapeutic strategies. Biotechnological tools like 3D bioprinting, immunohistochemistry and RNA sequencing have revolutionized how scientists study TME. They provide valuable mechanistic insight which enables them to identify new therapeutic targets and predict treatment responses. The progress so far points towards a future where cancer treatment is no longer guided solely by the histology of solid tumours, but also by the molecular and cellular landscape in which the tumour exists. By continuing to harness biotechnology, we can deepen our understanding of cancer biology and ultimately improve outcomes for patients across the world.

References

Anderson, N. M., & Simon, M. C. (2020). The tumor microenvironment. *Current Biology, 30*(16), R921–R925. https://doi.org/10.1016/j.cub.2020.06.081

Arneth, B. (2019). Tumor microenvironment. *Medicina, 56*(1), 15. https://doi.org/10.3390/medicina56010015

Bussard, K. M., Mutkus, L., Stumpf, K., Gomez-Manzano, C., & Marini, F. C. (2016). Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Research, 18*(1), 84. https://doi.org/10.1186/s13058-016-0740-2

Characterization of the tumor microenvironment. (2023, June 12). *Precision for Medicine*. https://www.precisionformedicine.com/blog/methods-for-a-comprehensive-characterization-of-the-tumor-microenvironment/

Cheng, N., Bai, X., Shu, Y., Ahmad, O., & Shen, P. (2021). Targeting tumor-associated macrophages as an antitumor strategy. *Biochemical Pharmacology, 183*, 114354. https://doi.org/10.1016/j.bcp.2020.114354

Crouigneau, R., Li, Y.-F., Auxillos, J., Goncalves-Alves, E., Marie, R., Sandelin, A., & Pedersen, S. F. (2024). Mimicking and analyzing the tumor microenvironment. *Cell Reports Methods, 4*(10), 100866. https://doi.org/10.1016/j.crmeth.2024.100866

Feng, Y., Tang, Q., Wang, B., Yang, Q., Zhang, Y., Lei, L., & Li, S. (2024). Targeting the tumor microenvironment with biomaterials for enhanced immunotherapeutic efficacy. *Journal of Nanobiotechnology, 22*(1), 162. https://doi.org/10.1186/s12951-024-03005-2

Glaviano, A., Lau, H. S.-H., Carter, L. M., Lam, H. Y., Okina, E., Donavan, C., Tan, W., Ang, H. L., Carbone, D., Yee, M. Y.-H., Shanmugam, M. K., Huang, X. Z., Sethi, G., Tan, T. Z., Huang, R. Y.-J., Ungefroren, H., Giovannetti, E., & Tang, D. G. (2025). Harnessing the tumor microenvironment: Targeted cancer therapies through modulation of epithelial-mesenchymal transition. *Journal of Hematology & Oncology, 18*(1), 91. https://doi.org/10.1186/s13045-024-01634-6

Hernández-Camarero, P., López-Ruiz, E., Marchal, J. A., & Perán, M. (2021). Cancer: A mirrored room between tumor bulk and tumor microenvironment. *Journal of Experimental & Clinical Cancer Research, 40*(1), 194. https://doi.org/10.1186/s13046-021-02022-5

Kanishka Rangamuwa, Aloe, C., Christie, M., Asselin-Labat, M.-L., Batey, D., Irving, L., John, T., Bozinovski, S., Leong, T. L., & Steinfort, D. (2023). Methods for assessment of the tumour microenvironment and immune interactions in non-small cell lung cancer: A narrative review. *Frontiers in Oncology, 13*, 1129195. https://doi.org/10.3389/fonc.2023.1129195

Li, Y., Liu, J., Xu, S., & Wang, J. (2023). 3D bioprinting: An important tool for tumor microenvironment research. *International Journal of Nanomedicine, 18*, 8039–8057. https://doi.org/10.2147/ijn.s435845

Magaki, S., Hojat, S. A., Wei, B., So, A., & Yong, W. H. (2019). An introduction to the performance of immunohistochemistry. In *Methods in Molecular Biology* (Vol. 1897, pp. 289–298). Springer. https://doi.org/10.1007/978-1-4939-8935-5_25

Mashambanhaka, F. (2018, November 28). What is 3D bioprinting? – Simply explained. *All3DP*. https://all3dp.com/2/what-is-3d-bioprinting-simply-explained/

Muniz Partida, C., & Walters, E. (2023). A novel immunohistochemical protocol for paraffin embedded tissue sections using free-floating techniques. *Frontiers in Neuroanatomy, 17*, 1154568. https://doi.org/10.3389/fnana.2023.1154568

Pan, L., Feng, Z., & Peng, S. (2022). A review of machine learning approaches, challenges and prospects for computational tumor pathology. *arXiv*. https://arxiv.org/abs/2206.01728

Rajbhandary, S., Dhakal, H. P., & Shrestha, S. (2023). Tumor immune microenvironment (TIME) to enhance antitumor immunity. *European Journal of Medical Research, 28*(1), 223. https://doi.org/10.1186/s40001-023-01125-3

Shen, R., Huang, Y., Kong, D., Ma, W., Liu, J., Zhang, H., Cheng, S., & Feng, L. (2024). Spatial distribution pattern of immune cells is associated with patient prognosis in colorectal cancer. *Journal of Translational Medicine, 22*(1), 91. https://doi.org/10.1186/s12967-024-05418-x

Su, X., Li, J., Xu, X., Ye, Y., Wang, C., Pang, G., Liu, W., Liu, A., Zhao, C., & Hao, X. (2024). Strategies to enhance the therapeutic efficacy of anti-PD-1 antibody, anti-PD-L1 antibody and anti-CTLA-4 antibody in cancer therapy. *Journal of Translational Medicine, 22*(1), 138. https://doi.org/10.1186/s12967-024-05552-6

Telarovic, I., Wenger, R. H., & Pruschy, M. (2021). Interfering with tumor hypoxia for radiotherapy optimization. *Journal of Experimental & Clinical Cancer Research, 40*(1), 197. https://doi.org/10.1186/s13046-021-02000-x

Tie, Y., Tang, F., Wei, Y., & Wei, X. (2022). Immunosuppressive cells in cancer: Mechanisms and potential therapeutic targets. *Journal of Hematology & Oncology, 15*(1), 126. https://doi.org/10.1186/s13045-022-01282-8

Wang, M., Wang, S., Desai, J., Trapani, J. A., & Neeson, P. J. (2020). Therapeutic strategies to remodel immunologically cold tumors. *Clinical & Translational Immunology, 9*(12), e1226. https://doi.org/10.1002/cti2.1226

World Health Organization. (2024, February 1). Global cancer burden growing, amidst mounting need for services. *World Health Organization*. https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services

Wu, Y., Yang, S., Ma, J., Chen, Z., Song, G., Rao, D., Cheng, Y., Huang, S., Liu, Y., Jiang, S., Liu, J., Huang, X., Wang, X., Qiu, S., Xu, J., Xi, R., Bai, F., Zhou, J., Fan, J., & Zhang, X. (2022). Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discovery, 12*(1), 134–153. https://doi.org/10.1158/2159-8290.cd-21-0316

Xiao, Y., & Yu, D. (2020). Tumor microenvironment as a therapeutic target in cancer. *Pharmacology & Therapeutics, 221*, 107753. https://doi.org/10.1016/j.pharmthera.2020.107753

Yan, H., Ju, X., Huang, A., & Yuan, J. (2024). Advancements in technology for characterizing the tumor immune microenvironment. *International Journal of Biological Sciences, 20*(6), 2151–2167. https://doi.org/10.7150/ijbs.92525

Zhou, X., Zhu, W., Nowicki, M., Miao, S., Cui, H., Holmes, B., Glazer, R. I., & Zhang, L. G. (2016). 3D bioprinting a cell-laden bone matrix for breast cancer metastasis study. *ACS Applied Materials & Interfaces, 8*(44), 30017–30026. https://doi.org/10.1021/acsami.6b10673

Zhou, Y., Cheng, L., Liu, L., & Li, X. (2023). NK cells are never alone: Crosstalk and communication in tumour microenvironments. *Molecular Cancer, 22*(1), 172. https://doi.org/10.1186/s12943-023-01737-7

RNA-Based Therapeutics: Rewriting the Language of Medicine

Harish Siva

RNA-based therapeutics are revolutionizing medicine by targeting RNA, the blueprint of protein production, rather than proteins themselves. This allows scientists to silence harmful genes, correct faulty transcripts, and regulate protein synthesis at its source. The success of COVID-19 mRNA vaccines and therapies like patisiran and nusinersen demonstrates their potential across infectious disease, cancer, and rare genetic disorders. Advances in delivery systems have improved stability and broadened applications, while innovations such as self-amplifying and circular RNA promise even greater impact. Despite challenges in cost, accessibility, and global equity, RNA therapies represent a shift toward precision medicine and the possibility of personalized treatments once thought impossible.

Keywords: *RNA therapeutics; mRNA vaccines; siRNA; antisense oligonucleotides; gene silencing; personalized medicine*

We are in a world that is heavily influenced by molecular medicine, and in that, RNA-based therapeutics have come out as one of the most groundbreaking achievements of the 21st century (Damase et al., 2025). Unlike traditional drugs that interact with proteins (the final products of gene expression), RNA-based therapies go a step further. They target the blueprints that code those proteins, the RNA molecules themselves. By stepping in at an earlier stage in the central dogma (DNA → RNA → Protein), these therapies can silence harmful genes, fix harmful genetic messages, or control protein production at the source. As science continues to push past the boundaries, RNA-based drugs are changing the way we understand disease treatment, prevention, and even the future of personalized care.

1 The Central Role of RNA

RNA's main role is as a messenger between DNA and proteins, and this has been known since the mid-20th century. It serves as the crucial middle point that transcribes the genetic instructions in DNA and translates them into proteins. But for decades, RNA was seen as just a temporary script that they thought couldn't be altered or used. That viewpoint has now been changed (Liu, Ou, & Hou, 2024).

Recent advances in biochemistry, bioinformatics, and nanotechnology have allowed scientists to manipulate RNA with greater accuracy. This helps form a new era in which RNA is not just a biological tool, but a powerful therapeutic target and delivery system. Scientists are now designing RNA molecules that can modify gene expression in living organisms. This offers treatments for conditions that were once viewed as untreatable.

Furthermore, the versatility of RNA's structure and its ability to fold into complex shapes, and how it interacts with other biomolecules, has inspired new and innovative therapeutic styles de Fougerolles et al., 2007). Scientists now see RNA as programmable and can be customized to execute biological tasks with high accuracy.

2 The mRNA Vaccine Revolution

The most publicly known success of RNA therapeutics was seen during the COVID-19 pandemic. In 2020, researchers at Pfizer-BioNTech and Moderna used messenger RNA (mRNA) technology to develop vaccines at an extraordinary speed. These mRNA vaccines had the genetic instructions for cells to produce a harmless version of the coronavirus protein, which then trained our immune system to recognize the virus and helps it fight off the real virus (KA et al., 2025). Unlike traditional vaccines that use inactivated pathogens (disease-causing microorganisms), mRNA vaccines are faster to design, easier to scale, and safer to administer (Sahin, Karikó, & Türeci, 2014).

This global rollout served as a success story for RNA technology. It showed that RNA could be safely introduced into human cells, produce therapeutic proteins, and stimulate strong immune responses. It also increased the interest that people had in mRNA platforms to treat cancer, autoimmune disorders (conditions where the immune system attacks the body's healthy cells and tissues). and rare genetic diseases.

The speed of mRNA vaccine development from genetic sequencing of the virus to emergency use authorization in under a year represented not just a scientific victory, but a shift in vaccine technology. It proved that with the right structure, RNA-based platforms could even overcome pandemics (Damase et al., 2025).

3 Types of RNA Therapeutics

RNA therapies come in several types, each with unique jobs and therapeutic goals:

- **mRNA therapies**: Deliver synthetic RNA instructions to cells so they can manufacture therapeutic proteins, such as enzymes or antigens. These can be used in vaccines, enzyme replacement therapies, and cancer immunotherapy. The mRNA acts like a temporary set of instructions, allowing the body to make its therapeutic proteins without altering the DNA (Sahin, Karikó, & Türeci, 2014).

  Small interfering RNAs (siRNAs): Silence harmful genes by binding to and lowering their corresponding mRNA, effectively "turning off" gene expression. This prevents the production of disease-causing proteins at the source, offering a precise way to stop harmful genetic activity.
- **Antisense oligonucleotides (ASOs)**: Short RNA-like strands that bind to mRNA transcripts to prevent translation or correct splicing (DNA transcribed into RNA) errors. By targeting specific mRNA sequences, ASOs can restore normal protein production or reduce toxic protein buildup.
- **MicroRNAs (miRNAs)** and **aptamers**: Other RNA-based tools that regulate gene expression post-transcriptionally or bind specific targets, often used in diagnostics and targeted drug delivery. miRNAs help fine-tune gene activity naturally, while aptamers act like RNA-based antibodies that can block or guide treatment to exact molecules (Qin et al., 2022).

Each of these approaches addresses different biological challenges and offers solutions for diseases ranging from rare genetic disorders to even cancer.

4 Success Stories in RNA Therapy

Real-world examples to show the power of RNA medicines:

Patisiran, the first FDA-approved siRNA therapy, treats hereditary transthyretin amyloidosis by silencing the TTR gene, which is responsible for producing a misfolded protein that gathers in tissues and organs and leads to organ dysfunction. (Adams et al., 2018). Its approval can be seen as a breakthrough by proving that RNA interference could be safely and effectively used in humans to treat genetic disease.

Nusinersen (Spinraza), an antisense therapy, treats spinal muscular atrophy (SMA), a severe neuromuscular disease in infants causing loss of motor neurons, by correcting a splicing defect in the SMN2 gene and restoring production of a crucial survival protein (Finkel et al., 2017). This therapy was the first to improve survival and motor function in children with SMA, a condition once considered untreatable.

Investigational RNA therapies are being tested for Huntington's disease, ALS, and certain dementias, where genetic defects previously had no effective treatments (Tabrizi et al., 2019). These efforts signal hope for patients with devastating neurological disorders by targeting disease-causing genes at their source.

In oncology, RNA platforms are being used to develop cancer vaccines that instruct the immune system to attack tumor-specific antigens. In rare diseases, scientists are beginning to design patient-specific RNA medicines personalized to a patient's unique genetic mutation. These advances represent a shift toward precision medicine, where RNA therapies can be customized to the unique profile of each patient's disease.

5 Benefits and Challenges

RNA-based therapies offer several key advantages. First, they can be developed with remarkable speed; once a disease target is identified, RNA molecules can be designed and tested much faster than traditional drug candidates. Second, they provide a level of precision that allows researchers to target genes and sequences that were previously considered "undruggable," expanding the scope of possible treatments. Finally, they are highly adaptable: once a delivery system is validated, the RNA payload can be modified and repurposed to address a wide range of other conditions without the need to reinvent the entire therapeutic platform (Qin et al., 2022).

However, these therapies face biological and logistical challenges. RNA is very unstable and fragile, easily degraded by enzymes in the body, and too large and negatively charged (Repelling each other) to cross cell membranes efficiently. To overcome this, scientists have developed delivery vehicles such as lipid nanoparticles (LNPs) and GalNAc conjugates to protect RNA and deliver it into cells (Cullis & Hope, 2017; Dowdy, 2017). Chemical modifications to RNA bases can also improve stability and reduce the risk of triggering harmful immune responses.

Beyond science, issues like manufacturing complexity, high production costs, and cold storage (such as the ultra-cold temperatures required for early mRNA vaccines) cause problems for global accessibility. Fortunately, next-generation innovations such as thermostable RNA structures and freeze-dried RNA vaccines are actively trying to address these concerns.

Global equity remains a concern. While RNA therapies can grow in wealthier nations, making them affordable and available to developing countries is important for true medical fairness (Sahin et al., 2014).

6 The Road Ahead: RNA's Limitless Future

As of 2025, over 100 RNA-based drugs are in clinical development. They target diseases like cystic fibrosis, cardiovascular conditions, viral infections, metabolic disorders, and much more. Emerging technologies, including self-amplifying mRNA (saRNA) and circular RNA (circRNA), show promise for even longer-lasting effects and broader applications.

SaRNA codes not only the therapeutic protein but also the tool to replicate itself inside cells, reducing the number of doses needed. CircRNA resists degradation by forming a closed-loop structure, which may enable more sustained protein expression (Damase et al., 2025).

The promise of personalized RNA medicine is also really interesting to think about. It may be possible to generate patient-specific RNA therapies based on their genetic profile, opening the pathway for individualized treatments with fewer side effects and even greater results.

Imagine a future where a patient's entire treatment plan is designed based on a quick genome scan and RNA code written just for them.

References

Adams, D., Gonzalez-Duarte, A., O'Riordan, W. D., Yang, C. C., Ueda, M., Kristen, A. V., ... & Solomon, S. D. (2018). Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. *New England Journal of Medicine, 379*(1), 11–21. https://doi.org/10.1056/NEJMoa1716153

Cullis, P. R., & Hope, M. J. (2017). Lipid nanoparticle systems for enabling gene therapies. *Molecular Therapy, 25*(7), 1467–1475. https://doi.org/10.1016/j.ymthe.2017.03.013

Damase, T. R., Stephens, A., Spencer, D. S., Karlsson, J., Downing, T. L., & Silva, J. R. (2025). The limitless future of RNA therapeutics. *Frontiers in Bioengineering and Biotechnology, 9,* 628137. https://doi.org/10.3389/fbioe.2021.628137

de Fougerolles, A., Vornlocher, H. P., Maraganore, J., & Lieberman, J. (2007). Interfering with disease: A progress report on siRNA-based therapeutics. *Nature Reviews Drug Discovery, 6*(6), 443–453. https://doi.org/10.1038/nrd2310

Dowdy, S. F. (2017). Overcoming cellular barriers for RNA therapeutics. *Nature Biotechnology, 35*(3), 222–229. https://doi.org/10.1038/nbt.3802

Sahin, U., Karikó, K., & Türeci, Ö. (2014). mRNA-based therapeutics: Developing a new class of drugs. *Nature Reviews Drug Discovery, 13*(9), 759–780. https://doi.org/10.1038/nrd4278

Chaudhary, N., Weissman, D., & Whitehead, K. A. (2021). mRNA vaccines for infectious diseases: Principles, delivery and clinical translation. *Nature Reviews Drug Discovery, 20*(11), 817–838. https://doi.org/10.1038/s41573-021-00283-5

Liu, Y., Ou, Y., & Hou, L. (2024). Advances in RNA-based therapeutics: Challenges and innovations in RNA delivery systems. *Current Issues in Molecular Biology, 47*(1), 22. https://doi.org/10.3390/cimb47010022

Qin, S., Tang, X., Chen, Y., Chen, K., Fan, N., & Chen, R. (2022). mRNA-based therapeutics: Powerful and versatile tools to combat diseases. *Signal Transduction and Targeted Therapy, 7*(166). https://doi.org/10.1038/s41392-022-01007-w

MEDICINE & HEALTHCARE INNOVATION

Advancing Early Cancer Diagnosis Through Medical Imaging Innovation

Aydin Moideen

Early detection of cancer significantly increases patient survival rates. However, delays in diagnostics, especially in underserved areas, remain a major obstacle. This paper explores the growing role of AI in medicine, mainly in convolutional neural networks supporting radiologists in identifying tumors from medical images. By improving accuracy, reducing diagnostic time, and extending care into low-resource settings, AI-assisted imaging offers a powerful tool for the future of cancer diagnosis.

Keywords: *Early cancer detection; artificial intelligence; convolutional neural networks; medical imaging; radiology; healthcare access; low-resource settings*

1 Introduction

Cancer remains one of the leading causes of death worldwide. According to the World Health Organization (WHO, 2022), nearly 10 million deaths in 2020 were attributed to cancer, with 48% linked to late-stage diagnoses. While medical imaging technologies like CT scans, X-rays, and MRIs have long played a role in early detection, human analysis alone is limited by time, fatigue, and diagnostic variability. In recent years, convolutional neural networks (CNNs), a type of deep learning model, have risen as a valuable support system for radiologists. Their ability to analyze large volumes of imaging data and identify patterns has created opportunities to improve diagnostic speed and accuracy, especially in early-stage detection (Schiff et al., 2020).

2 Understanding CNNs in more depth

CNNs are specialized algorithms inspired by the visual cortex of the human brain. They work by recognizing features in images like masses or abnormal tissue structures. By training CNNs on thousands of medical images, researchers can develop models capable of identifying tumors in mammograms, lung scans, brain MRIs, and more. One study from Stanford University used a CNN called CheXNet to detect pneumonia from chest X-rays and found it could outperform radiologists in certain cases (Rajpurkar et al., Radiology, 2017). Similarly, Google Health's LYNA model (Lymph Node Assistant) demonstrated 99% accuracy in detecting metastatic breast cancer in lymph node biopsies (Liu et al., JAMA Oncology, 2019).

3 Benefits in Oncology and Beyond

AI-assisted imaging is already demonstrating real-world impact in oncology. In breast cancer, convolutional neural networks (CNNs) are being used to flag suspicious regions in mammograms that radiologists might otherwise overlook (Liu et al., 2019). In lung cancer, low-dose CT scans analyzed by AI systems have been shown to detect nodules at an earlier stage than traditional interpretation allows, offering opportunities for earlier intervention (Schiff et al., 2020). Similarly, in the context of brain tumors, MRI-based CNNs assist in the precise localization of gliomas, thereby improving the accuracy of neurosurgical planning and ultimately enhancing patient outcomes.

Beyond cancer, AI models are being trained to recognize diabetic retinopathy, tuberculosis, and stroke patterns, proving that CNNs are far more capable than just one specialty with enough time CNNs could be used to process any type of medical imaging.

4 Ethical and Practical Challenges

Despite the optimism surrounding an automated future, medical CNNs continue to face significant challenges that must be addressed. One of the foremost issues is data bias: many models are trained on datasets originating from wealthier regions, which limits their accuracy and reliability when applied to underrepresented populations (Obermeyer et al., 2019). Another challenge is interpretability, as CNNs are often criticized as "black boxes." For physicians to fully trust these systems, they must be able to understand how the model arrives at its decisions. Regulation also remains a critical barrier. AI diagnostics, particularly those used for life-threatening diseases, require rigorous validation before integration into clinical workflows (Schiff et al., 2020). Ultimately, while AI can serve as a powerful tool to assist clinicians, it cannot replace the human judgment, empathy, and compassion that remain central to patient care.

5 Healthcare Equity and Global Potential

One of the most exciting benefits of AI-assisted diagnostics is its potential to bridge healthcare gaps. In many parts of the world, radiologists are scarce. According to the WHO, two-thirds of the global population lacks access to basic imaging services (World Health Organization, 2022). AI tools could be used in portable devices to bring detailed analysis to underserved clinics, emergency vehicles, and developing nations.

By expanding access to early detection, AI has the power to reduce cancer mortality globally.

6 Conclusion

AI-assisted imaging represents a pivotal advancement in modern medicine. By supporting early diagnosis, CNNs can help doctors detect cancer sooner, treat it more effectively, and ultimately save lives. While the technology is still evolving, its potential to transform care, and improve the work of medical practitioners makes it a crucial step forward in improving patient outcomes.

References

World Health Organization. (2022, March 3). *Cancer*. https://www.who.int/news-room/fact-sheets/detail/cancer

Schiff, G. D., Hasan, O., Kim, S., Abrams, R., Cosby, K., Lambert, B. L., Elstein, A. S., Hasler, S., Kabongo, M., Krosnick, T., Odwazny, R., Wisnivesky, J. P., & McNutt, R. (2020). Diagnostic errors in cancer care: Implications for improving diagnosis. *BMJ Quality & Safety, 29*(5), 370–373. https://qualitysafety.bmj.com/content/29/5/370

Liu, Y., Kohlberger, T., Norouzi, M., Dahl, G. E., Smith, J. L., Mohtashamian, A., Çelik, H., Babyn, P., Badano, A., & Hipp, J. D. (2019). Artificial intelligence–based breast cancer nodal metastasis detection. *JAMA Oncology, 5*(10), 1471–1472. https://jamanetwork.com/journals/jamaoncology/fullarticle/2747980

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., & Lungren, M. P. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv*. https://arxiv.org/abs/1711.05225

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Restoring Lost Memory: How BCIs Can Reverse Alzheimer's Progression

Charis Tsang

Alzheimer's disease, the leading cause of dementia, affects over 139 million people worldwide and imposes a profound social and economic burden. Current treatments can only slow progression or manage symptoms but cannot restore lost neuronal connections. Brain-Computer Interfaces (BCIs), particularly stimulation-based systems, are emerging as a potential therapy to bypass damaged neural circuits, enhance memory consolidation, and even retrieve lost memories. This paper reviews the history and mechanisms of BCIs, explores their experimental applications to memory restoration, and highlights both the promise and limitations of stimulation of BCIs in Alzheimer's treatment. While preliminary studies show improved memory performance in controlled trials, translation to Alzheimer's patients remains hindered by biological, technical, and ethical challenges. Ultimately, BCIs offer an exciting but cautious path forward, requiring interdisciplinary collaboration to ensure safe, equitable, and meaningful clinical application.

Keywords: *Brain-computer interfaces; alzheimer's disease; memory restoration; stimulation bci; neuroethics; dementia therapies; neural circuits*

1 Introduction

By the time Alzheimer's is diagnosed, up to 70% of neurons may already be lost in critical brain regions like the hippocampus (Huentelman, n.d.). Over 139 million people are expected to be living with dementia by 2050 (Gauthier et al., 2022), costing the global economy more than $2.8 trillion (USD) annually. Beyond the numbers, the emotional and physical toll on families is immense, informal caregivers often have to provide over 5 hours of supervision and support every single day (World Health Organization, 2023).

In light of this irreversible damage and the heavy burden on patients and families, we urgently need new solutions that not only manage Alzheimer's symptoms, but to rebuild what has been lost. This is where the groundbreaking advancement enters: Brain-Computer Interface (BCI), a form of neurotechnology that could soon help restore memories, bypass damaged neural circuits, and transform the way we understand and treat dementia.

2 History of BCI

To understand modern advancements in BCI, we first have to backtrack nearly a century to the discovery that laid the foundation for the entire field: the first recording of human electroencephalogram (EEG). In 1924, Hans Berger captured the brain's electrical signal for the first time, paving the way for scientists to turn thoughts into actions. Building on Berger's work, the term "Brain-Computer Interface" was officially introduced by Jaques Vidal as he published the foundational paper *Toward Direct Brain-Computer Communications*, where he proposed using EEG signals for direct communication with computers. This paper marks the beginning of experiments using brain signals to control external devices.

A breakthrough came in 1988 when researchers created a speller system that allowed users to select letters based on brain responses. Then, in the 1990s, came the first trials of invasive BCI, where Phillip Kennedy implanted the first device directly into a human brain. Starting in the 2000s, BCI technology advanced rapidly, enabling paralyzed individuals to control computer cursors and robotic limbs. Over the next decade, these systems became even more sophisticated, allowing users to perform complex movements and even feel precise touches through direct brain interfaces (Kawala-Sterniuk et al., 2021).

3 How do BCIs work?

The Brain-Computer Interface (BCI) bypasses the usual pathways required for movements and communication, enabling a direct interaction between the brain and the computer. Without having to send signals through muscles, BCIs detect the brain activity and translates it into commands that computers can understand.

The process begins with signal acquisition, in which devices record brain activity and relay it to a computer for processing. Scientists use a variety of techniques to capture these signals, most commonly through non-invasive methods such as EEG headsets that detect electrical activity through the scalp, or advanced imaging technologies like MEG machines and MRI scanners. In certain cases, more invasive approaches may be employed, involving the surgical implantation of electrodes directly into the brain. Once the input is processed, the resulting output can be used to control an application, and feedback from the system informs the user of the outcome, completing the loop between brain activity, computational interpretation, and user interaction.

Next comes signal interpretation, when the computer analyzes the patterns of brain activity using specialized algorithms to decode the user's intentions like moving cursors, selecting letters, or instructing a robotic movement. Once the system matches a pattern to a command, it sends that command to an output device.

Finally, feedback is essential in this process. It allows the user to receive what interpretation the computer has made and adjust their mental focus to allow the BCI system to refine its application (Cumming School of Medicine, n.d.).

4 Rise of Stimulation BCIs

While traditional BCIs focus on interpreting brain activity to control external devices, scientists are now expanding the technology to include stimulation-based BCIs. These systems not only decode brain signals but also send targeted electrical impulses back to the brain and nervous system.This bidirectional communication opens new possibilities for repairing damaged neural circuits and restoring lost functions including, motor skills, sensory feedback and more revolutionary of all, memory.

Stimulation BCIs act as a neural prosthetic that delivers specific firing patterns to reawaken or strengthen inactive neurons. This creates a close-loop system similar to regular BCIs, but with the added ability to stimulate while simultaneously monitoring the brain's response in real-time. This allows for immediate adjustments in the timing, location, intensity of stimulation to guide neural activity with precision.

Although this technology is still in its early stages, promising advancements in animal models and preliminary human trials show stimulation BCIs' ability to enhance memory consolidation and potentially retrieve forgotten memories (Song & Liu, 2024).While still a work in progress, this technology hints at a future where we may be able to rewrite the brain's connections and reverse damage caused by Alzheimer's Disease and other forms of dementia.

5 Alzheimer's Disease

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder and the most common form of dementia. It is characterized by memory impairment, cognitive decline and eventually loss of independence. At the cellular level, AD involves the accumulation of β-amyloid plaques and tau neurofibrillary tangles that block neuronal transportation and damage synaptic communication. These abnormalities trigger widespread neuronal dysfunction, particularly the hippocampus, the brain region essential for learning and episodic memory.

This neuronal damage causes shrinkage in brain areas for memory formation, making it difficult for patients to form new memories or recall recent events (National Institute on Aging, 2024). Current treatments focus on slowing disease progression or managing symptoms, but none can restore lost neuronal connections.

6 Emerging BCI Therapies for Alzheimer's Disease

Given Alzheimer's complexity and progression, stimulation BCI offers a promising path for interventions that go beyond what current treatments can achieve. Rather than simply preserving what remains, BCIs could potentially re-establish communication within disrupted memory circuits, especially in the hippocampus, which is most affected in early AD.

While technologies Functional Electrical Stimulation (FES) and Deep Brain Stimulation (DBS) have shown clinical benefits in movement disorders like Parkinson's Disease, their application to Alzheimer's is far more complicated. AD is not defined by localized circuit dysfunction, but by widespread, progressive neurodegeneration across multiple interconnected cognitive networks. Furthermore, unlike motor disorders that demonstrate predictable symptoms with preserved cognition, AD involves a diffuse loss of memory, attention as well as executive functions, making it extremely challenging to identify a singular target that can have consistent reproduction of patterns and results (Awuah, et al., 2024).

A pioneering study from the University of California (USC) demonstrates a high potential of stimulation BCIs in restoring memory. Researchers developed a real-time feedback loop that mimics the hippocampus' natural firing patterns associated with episodic memory. This system monitors, processes and delivers precise electrical stimulation in real time to enhance memory encoding and recall. As of October 2024, the system was tested on epilepsy patients who were already undergoing invasive monitoring for seizure localization, which allows researchers to trial the BCI in a safe, neurosurgical setting. Earlier versions were also tested on non-human primates to verify memory enhancement effects. The results showed a 30–50% increase in memory performance. However, this population does not accurately reflect the neurodegenerative nature of AD. Epileptic patients often have largely intact memory systems, unlike those suffering from widespread neuronal loss. According to Harrison (2024), this presents significant translation challenges. For instance, AD brains exhibit reduced neuroplasticity due to synaptic degeneration and glial scarring, making it harder for them to respond to stimulation. Additionally, the temporal variability in AD progression complicates timing and dosage of interventions. While this hippocampal-targeted BCI offers a strong foundation, further adaptation is necessary before it can be realistically applied to AD patients.

7 Ethical Concerns

As the development of stimulation-based BCI advances towards clinical application, especially in vulnerable populations like those with AD, various ethical concerns arise. The foremost is informed consent. Many BCI trials, especially invasive ones, require continuous informed consent. Yet, Alzheimer's patients may lack the cognitive capacity to fully understand the risks involved. This raises difficult questions: Who has the authority to consent on behalf of the patient? Should early-stage consent remain valid as the disease progresses?

Beyond consent, privacy and data protection is crucial alongside such scientific discoveries. BCIs collect detailed neural activity data that could reveal intimate thoughts, emotions, and memories. If misused by corporations, insurers, or governments, this data could pose significant risks. The possibility of BCIs being exploited for behavioral manipulation, similar to how social media algorithms influence user behavior, poses risks of targeted stimulation being used beyond medical use such as commercial purposes or personalized advertisements.

BCIs also raise questions regarding identity and authenticity. If a device alters memory through enhancements or retrieval, how can a person preserve their memories' authenticity? Such concerns go beyond neuroethics, prompting a deeper consideration regarding the necessity and meaning behind treating the mind through alterations. Is it ethical to be potentially recalling a patient's traumatic experiences that could lead to unnecessary distress that they would have forgotten? There is still a lack of research on how a patient's mental wellbeing has been changed due to memory modifications (Livanis, et al., 2024).

Lastly, accessibility is an important consideration. BCI stimulation treatments are likely to be costly, possibly limiting access to wealthy or urban populations. Without updated health policies, these new advancements may further reinforce the disparities in healthcare systems, leaving unprivileged patients without adequate care

In summary, stimulation-based BCI holds extraordinary promise for restoring memory and function in Alzheimer's patients, but these benefits must be carefully examined and weighed against ethical complexities they introduce. Promoting this innovation will require close collaboration among neuroscientists, clinicians, ethicists, and policymakers to ensure that BCIs are developed and deployed responsibly for the benefit of all.

8 Barriers in BCIs Development

Beyond ethical considerations, there are numerous technical and biological limitations that hinder the precise execution of stimulation-based BCIs. Invasive BCIs carry significant safety risks, including potential nerve damage during implantation, infection, and the formation of scar tissue that can degrade signal quality over time. Additionally, the body's immune system may reject the implanted device, further compromising its effectiveness. On the other hand, non-invasive BCIs often suffer from low signal resolution, making it difficult to deliver stimulation accurately. Compounding this issue, computers still face challenges in decoding memory-specific neural patterns, particularly in Alzheimer's patients, where brain atrophy makes signal interpretation even more difficult.

Most BCI systems also require frequent and time-consuming recalibrations to adapt to the brain's natural changes in neural activity. This adds another layer of inconvenience, especially considering that caregivers are already investing hours each day to support the patient. Battery life presents yet another hurdle. Implantable devices are vulnerable to short circuits, as the brain's internal environment can corrode insulating materials over time. Limited battery capacity restricts long-term use and raises concerns about device durability and reliability.

Most importantly, the brain's complexity generates an enormous volume of data that must be processed in real time. This creates a computational challenge, increasing the risk of delayed or faulty signal processing, which could affect the device's accuracy and safety (Maseli, et al., 2023).

9 What's Next?

Overall, research on BCIs, specifically stimulation-based BCIs, remains in its early stages. One major limitation is the lack of multidisciplinary collaboration in current studies. Without stronger integration with fields like psychology, pathology, and bioethics, the development of more effective and user-centered BCI systems becomes limited. In addition, the field faces a shortage of participants for clinical trials, largely due to ethical complexities surrounding safety and technical execution. Existing trials are often small-scale and lack diversity, which slows the ability to generate results that are widely applicable across different populations. These challenges underscore the need for broader, more inclusive and collaborative research efforts moving forward.

That said, one of the primary goals for future research is to accurately translate thoughts, dreams, or even imagination directly into text or images, ideally through non-invasive technologies that maintain high precision while minimizing risks. Researchers envision using BCI to extract memories for hardware storage in computers to explore the possibility of faster retrieval of information or even backup systems for our biological memory. Another emerging focus is on developing self-sustaining power sources, including techniques to harvest energy from neural activity to continuously power implantable devices. The integration of Artificial Intelligence in processing these big data will also be a main objective. Last but not least, the key focus will continue to be developing algorithms that are efficient and secure for encrypting to ensure that brain signals are protected against external attacks and privacy issues.

10 Conclusion

While Brain-Computer Interfaces present an exciting frontier in neuroscience, biotechnology and medicine, their application in dementia care remains limited by both its biological complexity and practical constraints. The brain's dynamic nature, along with safety, technical and ethical concerns, continues to slow widespread clinical adaptation. Yet, these hurdles do not deny the technology's potential, as they highlight the importance of demonstrating scientific innovation with caution, inclusivity and purpose.

Future progress will depend not just on engineering advancements, but on interdisciplinary collaboration across neuroscience, psychology, ethics, and clinical care. The goal is not simply to retrieve lost memories, but to enhance the quality of life for individuals affected by cognitive decline. If successfully developed, BCIs could become more than just clinical tools as they may evolve into transformative technologies that reshape how we understand memory, identity, and human connection in the face of neurodegeneration.

References

Awuah, W. A., Ahluwalia, A., Darko, K., Sanker, V., Tan, J. K., Tenkorang, P. O., Ben-Jaafar, A., Ranganathan, S., Aderinto, N., Mehta, A., Shah, M. H., Boon Chun, K. L., Abdul-Rahman, T., & Atallah, O. (2024). Bridging minds and machines: The recent advances of brain-computer interfaces in neurological and neurosurgical applications. *World Neurosurgery*. https://doi.org/10.1016/j.wneu.2024.05.104

Cumming School of Medicine. (n.d.). *What is a brain computer interface?* University of Calgary. https://cumming.ucalgary.ca/research/pediatric-bci/bci-program/what-bci

Gauthier, S., Webster, C., Servaes, S., Morais, J. A., & Rosa-Neto, P. (2022). *World Alzheimer report 2022: Life after diagnosis*. Alzheimer's Disease International. https://www.alzint.org/u/World-Alzheimer-Report-2022.pdf

Harrison, G. (2024, October 30). *Turning memory loss into a distant memory*. USC Viterbi School of Engineering. https://viterbischool.usc.edu/news/2024/10/turning-memory-loss-into-a-distant-memory/

Huentelman, M. (n.d.). *How do brain cells die in Alzheimer's? What are amyloids?* MindCrowd. https://mindcrowd.org/how-do-brain-cells-die-in-alzheimers-what-are-amyloids/

Kawala-Sterniuk, A., Browarska, N., Al-Bakri, A., Pelc, M., Zygarlicki, J., Sidikova, M., Martinek, R., & Gorzelanczyk, E. J. (2021). Over fifty years with brain–computer interfaces—A review. *Sensors, 21*(17), 5942. https://doi.org/10.3390/s21175942

Livanis, E., Voultsos, P., Vadikolias, K., Pantazakos, P., & Tsaroucha, A. (2024). Understanding the ethical issues of brain-computer interfaces (BCIs): A blessing or the beginning of a dystopian future? *Cureus, 16*(4), e58243. https://doi.org/10.7759/cureus.58243

Maiseli, B., Abdalla, A. T., Massawe, L. V., Mbise, M., Mkocha, K., Nassor, N. A., Ismail, M., Michael, J., & Kimambo, S. (2023). Brain–computer interface: Trend, challenges, and threats. *Brain Informatics, 10,* Article 20. https://doi.org/10.1186/s40708-023-00199-3

National Institute on Aging. (2024, January 19). *What happens to the brain in Alzheimer's disease?* U.S. Department of Health & Human Services. https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-happens-brain-alzheimers-disease

World Health Organization. (2023). *Dementia*. https://www.who.int/news-room/fact-sheets/detail/dementia

CLIMATE TECHNOLOGY & SUSTAINABILITY

What are the most effective strategies for building climate change infrastructure?

Aribah Amer

Decisions about climate change infrastructure demand long-term foresight, as projects in transportation, water, energy, and urban development must endure for decades while facing an uncertain climate future. This study highlights the importance of adaptive, integrated strategies that link land-use, energy, and transportation planning while considering socio-economic equity and resilience across vulnerable populations. Using lessons drawn from experimental interventions in Mumbai and comparative policy frameworks, the analysis demonstrates that hybrid green-gray solutions, multipurpose infrastructure, and smart-grid innovations enhance both technical performance and community well-being. At the same time, challenges such as unequal access, fragmented governance, and slow technological updating limit resilience. The evidence suggests that successful pathways require multidisciplinary coordination, integration of adaptation with mitigation, and lifecycle planning that leverages replacement opportunities after extreme events. Ultimately, the research shows that resilient climate infrastructure emerges not only from technical advances but from inclusive strategies that align ecological knowledge, social equity, and adaptive governance.

Keywords: *Climate change; infrastructure resilience; adaptive planning; green-gray solutions; equity; governance; innovation*

Climate change infrastructure decisions often involve long-term commitments and are highly sensitive to climate factors. Examples include urbanization plans, risk management strategies, infrastructure development for water systems or transportation, and building design standards. These choices impact the next 50 to 200 years, with urbanization shaping city structures for even longer. Such decisions and investments are also susceptible to shifts in climate and rising sea levels. For instance, buildings are typically designed to last up to 100 years, but by 2100, climate conditions—according to most models—are expected to be significantly different from today's. As a result, architects and engineers must consider and plan for these future climate changes when designing buildings. (Hallegatte 2009)

Building climate change resilience requires a multifaceted approach that incorporates adaptive strategies across different sectors, as well as consideration of socio-economic contexts and the specific needs of vulnerable populations. Effective strategies for enhancing resilience can be categorized into adaptation strategies, agricultural resilience practices, infrastructure development, public health initiatives, and renewable energy integration.

Adaptation strategies are crucial for mitigating risks associated with climate impacts. Macassa et al. highlights the necessity for adaptation to include socio-ecological changes that respond to both climate and non-climatic factors, emphasizing a range of activities from short-term interventions to long-term planning that aim to leverage opportunities arising from climate change (Hallegatte 2009). Budin-Ljøsne et al. also underscores the importance of designing context-specific adaptation strategies that can address variable climate impacts, necessitating tailored approaches that reinforce both infrastructure and public health systems (Macassa, n.d.).

The transport sector not only plays a role in driving global climate change through the emission of greenhouse gases (GHGs), but it is also significantly affected by its impacts. Clear evidence shows that climate change is leading to more frequent and intense extreme weather events. These weather extremes along with rising sea levels and increasing global temperature pose serious threats to transportation infrastructure worldwide. The risks are particularly severe in coastal regions, especially in developing countries, where vulnerability is often higher. (Rattanachot and Wang 2015)

This study of Mumbai's key infrastructure resilience strategies gave us useful information that can help city leaders around the world deal with climate threats. By looking at experimental changes that improved flood and heat protection in the energy, transportation, and water sectors, clear trends about the best ways to do things started to emerge. Nature-based designs that use both green and technological solutions gave the best results in terms of resilience, dependability, and community liveability when they were used wisely and at network scales. Improvements to green-gray infrastructure made it less likely for systems to flood at the same time, and smart grid updates and grid-connected renewable sources made all services more reliable. Such complex solutions took advantage of the way infrastructures depend on each other. Trends that have been shown also pointed out choices that offer important benefits in addition to adapting to climate change. Green transportation routes and urban parks not only improved public health, well-being, and the ability to move around, but they also helped with drainage and drainage. Multipurpose infrastructure reduced the costs of both original and ongoing use over the lifecycle of an asset and became more popular with all stakeholders. However, the results also showed that equality was a key factor that affected how resilient people were in their own communities. Adoption barriers made it harder for less fortunate groups to get some resilient upgrades and amenity rewards. To close these kinds of gaps, policymakers need to use full-lifecycle, integrated policies that are carried out through coordinated planning and focused subsidy programs. Hence, this study proved that the best solutions can create strong, cohesive, and welcoming resilience by mixing advanced technology with knowledge of the environment. It also offered evidence that showed the benefits of using pilots to improve solutions on a larger scale. Successes help keep things moving forward, which requires new ideas. However, there should still be a focus on strategic synergies that use the way networks are linked and the values that come with them. When used consistently, these multidisciplinary approaches can help the world deal with climate threats by making vital infrastructure stronger, more fair, and more long-lasting. (Ojo 2024, 17)

The need for integration of planning. In both the transportation and coastal defense areas, and to some extent sewer infrastructure, studies have concluded that resources could be allocated more efficiently if infrastructure planning were better integrated with land‑use planning. A key obstacle noted to this integration, however, is that such planning typically occurs across several levels of government, with land‑use planning usually carried out at the local level. Studies also note the need to integrate planning across transportation modes to ensure redundancy during emergency situations. Examples of integrated planning are rare, but at least one state (New Jersey) has developed a state‑level planning process that uses as its "base layer" an aggregate land‑use plan built up from local land‑use plans. The resulting state plan is then used by the state budgeting office to determine where state infrastructure funds will be allocated. The result provides clarity on where the government plans to support infrastructure development (and replacement), which can leverage private investment by sending signals about geographic areas where development is supported. Integrated planning may also help facilitate financial sector adaptation (e.g., insurance schemes), which some have argued is much easier and more flexible and robust than technical adaptation (Hallegatte 2008). Finally, there is an emerging need to integrate adaptation and mitigation planning; this need is most acute in the energy sector.

The need to encourage innovation in technology and updating of standards. Many of the studies cited above note that a change in climate will present technological challenges that may require more resilient infrastructure capital. Centralized efforts to update building standards may be one means to spur the needed technological change, although in the United States, new efforts to streamline this process may be necessary (Meyer 2008). The Canadian government has already launched several efforts in this direction (Canadian Standards Association 2007a–c, 2006, 2005; Infrastructure Canada 2006).

The need to take best advantage of replacement opportunities, including extreme events. More frequent and destructive extreme events, such as recent hurricanes and riparian floods, have already proven to be a huge challenge to maintaining public infrastructure. At the same time, many studies note that adaptation to climate stresses is more cost-effectively accomplished during the design phase of projects, rather than as a retrofit to existing capital. Although extreme events are devastating to affected regions, the rebuilding process can be used as an opportunity to replace damaged infrastructure with more resilient capital.

References

Budin-Ljøsne, I. (n.d.). *Linking climate change adaptation and public health: Perspectives of Norwegian policymakers*. PubMed. https://pubmed.ncbi.nlm.nih.gov/38380518/

Hallegatte, S. (2009). Strategies to adapt to an uncertain climate change. *Global Environmental Change,* *19*(2), 240–247. https://www.sciencedirect.com/science/article/abs/pii/S0959378008001192

Macassa, G. (n.d.). *Public health aspects of climate change adaptation in three cities: A qualitative study*. PubMed. https://pubmed.ncbi.nlm.nih.gov/36011923/

Ojo, B. (2024). Strategies for the optimization of critical infrastructure projects to enhance urban resilience to climate change. *Journal of Scientific and Engineering Research, 11*(6), 107–123. https://doi.org/10.5281/zenodo.1234567 (if no DOI, keep URL or ISSN reference)

Rattanachot, W., & Wang, Y. (2015). Adaptation strategies of transport infrastructures to global climate change. *Transport Policy,* *41*, 50–59. https://www.sciencedirect.com/science/article/abs/pii/S0967070X15000384

Promise and Challenges of Solar Roadways as Sustainable Infrastructure

Lan Sun

Solar roadways represent an innovative approach to renewable energy generation by embedding photovoltaic panels into transportation infrastructure. Beyond producing electricity, these systems aim to provide multifunctional benefits such as lighting, heating, and smart traffic management. Case studies from China, Germany, and pilot projects in the United States highlight both the promise and challenges of implementation, including high costs, durability concerns, and efficiency losses compared to traditional solar farms. While the concept remains largely experimental, ongoing research and technological advances suggest potential pathways for integrating solar roadways into future sustainable infrastructure.

Keywords: *Solar roadways; renewable energy; smart infrastructure; photovoltaics; sustainability*

What if the roads we drive on could do more than just get us from point A to point B? What if they could generate electricity, melt snow, light up with smart signals, and even filter rainwater? That's the vision behind solar roadways, and it's already being developed by various tech companies.

Instead of traditional asphalt, these roads are made up of interlocking solar panels, shaped like hexagons. Why hexagons? They fit better on curved or narrow roads, are easier to replace, and are more efficient to transport than rectangular panels. They cost more to make, but the benefits in durability and flexibility make up for it.

Each panel has several layers that work together. The bottom layer, or Base Plate, provides support and distributes the electricity generated. It's also made from recycled materials, adding to the sustainability factor. Above that is the Electronic layer, filled with microprocessors that can produce lighting to warn vehicles passing light. The lighting is made possible due to the Solar Cell / LED layer. The solar cells convert sunlight into electricity, while the built-in LEDs light up the surface of the road.

The top layer of each panel is a tough, transparent surface. Some engineers suggest using transparent concrete, but Solar Roadways Inc. prefers tempered glass. It's more cost-effective, recyclable, and can be textured for grip so cars don't slip. Plus, it's specially made to let sunlight through while being strong enough to handle heavy traffic.

Government support for solar roadways varies widely across the globe. Countries like France, China, and the U.S. have invested in pilot projects through public funding and green infrastructure programs, while others, like Germany and Japan, remain skeptical due to high costs, maintenance concerns, and regulatory hurdles. International agreements such as the Paris Climate Accord and the UN Sustainable Development Goals have encouraged exploration of such technologies, though adoption remains uneven. Nations with ample land and sunlight may see more practical use cases, whereas highly urbanized regions prioritize more established renewable solutions like rooftop solar.

Solar roadways represent a bold intersection of clean energy, smart transportation, and digital infrastructure, offering multifunctional benefits like electricity generation, real-time data collection, and support for EVs and AVs. Though expensive upfront, their costs are offset by long-term gains — reduced emissions, energy savings, job creation, and enhanced safety. Integrated into smart city ecosystems, solar roads can serve as both power sources and intelligent platforms, enabling more efficient, safer transport systems. Strategic deployment—especially in car parks, walkways, and key urban areas—combined with evolving investment models and public-private partnerships, could make solar roadways a cornerstone of sustainable infrastructure in the 21st century.

References

Hoffman, S. (2021). *How solar panel highways work*. HowStuffWorks. https://science.howstuffworks.com/environmental/energy/solar-panel-highway.htm

Sandhya, R., & Rajeswari, K. (2021). Solar roadways. *International Journal of Science & Engineering, 1*(1), 1–5. https://www.ijse.latticescipub.com/wp-content/uploads/papers/v1i1/B8009081221.pdf

Wehrmann, B. (2025). *Solar power in Germany – Output, business & perspectives*. Clean Energy Wire. https://www.cleanenergywire.org/factsheets/solar-power-germany-output-business-perspectives#

Cohen, N. (2018). China's solar highway ambitions are seen in Jinan stretch. *Tech Xplore*. https://techxplore.com/news/2018-01-china-solar-highway-ambitions-jinan.html

Kumar, A. (2022, February). *Solar roadways: The roadways to next generation*. ResearchGate. https://www.researchgate.net/publication/358276171_SOLAR_ROADWAYS_THE_ROADWAYS_TO_NEXT_GENERATION

Ethics and Governance of Stratospheric Aerosol Injection

Zaah Michael Kodzo

A geoengineering technique called Stratospheric aerosol injection (SAI) is a mechanism of injecting reflective particles like sulfate aerosols into the stratosphere to cool Earth by 1–2°C. Conversely, it raises ethical concerns about moral hazard, inequity, and consent (Robock, 2018). This article examines whether SAI can be ethical if mitigation fails and who should govern its deployment. From analysis, case studies including Harvard's SCoPEx, paused due to Sámi Indigenous protests, and the UK's SATAN, banned for unauthorized experiments, reveals governance gaps. SAI is at risk of delaying decarbonization and adversely harming vulnerable regions such as African river basins facing droughts, South Asian monsoons risking disruption, and South Central American ecosystems experiencing precipitation shifts. The proposed Ethical Geoengineering Governance Framework (EGGF) states six principles to promote inclusive decision-making and prioritization of Indigenous, African Union, and regional representation. This structure advances climate justice and provides a path for ethical SAI deployment in a warming world.

Keywords: *Stratospheric aerosol injection (sai); geoengineering; climate justice; ethical governance; indigenous rights; environmental equity; decarbonization*

1 Introduction

The year 2024 being the Earth's hottest year on record, with global temperatures 1.55°C above pre-industrial levels, infringing the Paris Agreement's threshold (NOAA, 2025; WMO, 2025). An attempt to cool the planet by 1–2°C by spraying reflective particles through Stratospheric aerosol injection (SAI), emerges as a potential response if mitigation falters (Robock, 2018). However, SAI poses ethical challenges: moral hazard can delay decarbonization, while inequity threatens African river basins with droughts, South Asian monsoons with disruption, and South Central American ecosystems with precipitation shifts (Wagner, 2023; Zelli et al., 2024; Abiodun et al., 2025; Cohen et al., 2025). Indigenous communities, like the Sámi, face unauthorized consequences, breaching autonomy (UNEP, 2023; Sparerun, 2025). Can SAI be ethical, and who decides its deployment? This article examines the predicaments of SAI (moral hazard, inequity, consent) and proposes an inclusive governance framework to ensure equitable decisions for Indigenous, African, and vulnerable communities, advancing climate justice.

2 Literature Review

Despite the aim of Stratospheric aerosol injection (SAI) to cool the planet by 1–2°C, it carries significant risks. IPCC (2023) and Robock et al. (2020) accentuate disruptions to South Asian monsoons, African river basin droughts, and South Central American precipitation shifts, which threatens food security for millions. Ozone depletion and unintended climate feedbacks further entangles SAI's feasibility (Robock, 2018; Wagner, 2023). Governance gaps intensify these risks. Harvard's SCoPEx, paused in 2021 after Sámi protests over unauthorized testing, and the UK's SATAN, banned in Mexico in 2023 for unapproved experiments, highlight inadequate inclusive decision-making (UNEP, 2023; Geoengineering Monitor, 2024). These cases bring to light violations of Indigenous autonomy and inequitable risk distribution, specifically for African and Global South communities (Zelli et al., 2024; Abiodun et al., 2025). Researchers recommend inclusive governance frameworks, thereby emphasizing free, prior, and informed consent (FPIC) through Indigenous councils and African Union representation (Lee et al., 2025; Sparerun, 2025). Considering 2024's record heat (NOAA, 2025; WMO, 2025), ethical SAI requires global cooperation to offset risks and ensure climate justice.

3 Discussion

3.1 Moral Hazard

Stratospheric aerosol injection (SAI) could create a moral hazard by delaying essential decarbonization efforts crucial to the Paris Agreement's 1.5°C target (IPCC, 2023). Administrators may over-rely on the temporary cooling effect (1–2°C) of SAI, diverting resources from greenhouse gas reduction (Wagner, 2023). This reliance could commit to fossil fuel dependence, exacerbating long-term climate impacts like African droughts and South Asian monsoon disruptions (Abiodun et al., 2025). Unilateral SAI experiments which are often led by Global North institutions, amplify this risk by focusing short-term fixes to the detriment of global mitigation (Geoengineering Monitor, 2024). For instance, SAI's appeal as a quick solution may compromise renewable energy investments (Zelli et al., 2024). To mitigate moral hazard, SAI must be an additional tool, paired with enforceable decarbonization timelines. Inclusive governance, involving African Union and Indigenous councils will ensure SAI aligns with climate justice, preventing over-reliance on untested geoengineering (Lee et al., 2025).

3.2 Equity and Justice

The unequal impacts of SAI outline the need for equitable governance to counter Global North bias. The UK's SATAN project which was banned in Mexico in 2023 due to unauthorized SAI testing, highlights moral hazard and inequity (Geoengineering Monitor, 2024). Unilateral tests risk delaying decarbonization and exacerbating African droughts, disrupting South Asian monsoons, and changing South Central American precipitation, threatening food security for smallholder farmers (Abiodun et al., 2025; Cohen et al., 2025). These risks disproportionately affect the Global South, where communities like those in the Nile basin face severe water scarcity (Zelli et al., 2024). Global North-led SAI governance often sidelines affected regions, perpetuating injustice (Wagner, 2023). Equitable frameworks must include African Union, South Asian, and South Central American representation in decision-making to ensure fair risk-benefit distribution (Lee et al., 2025). Prioritizing human rights and vulnerable populations is essential for climate justice in SAI deployment.

3.3 Informed Consent

The deployment of SAI depends on free, prior, and informed consent (FPIC) to respect community autonomy. Harvard's SCoPEx, halted in 2021 after Sámi protests in Sweden, exposed SAI's ethical failures (UNEP, 2023). Testing without FPIC threatened Sámi reindeer herding and cultural practices bound to stable climates, violating human rights. Global Indigenous exclusion demands FPIC through councils and forums to ensure ethical SAI governance (Sparerun, 2025; Lee et al., 2025). The Sámi case highlights how unauthorized SAI experiments endanger cultural and environmental harm, particularly for Indigenous groups reliant on local ecosystems (Geoengineering Monitor, 2024). FPIC requires transparent consultation, impact assessments, and veto power for affected communities, integrated into global SAI frameworks (Zelli et al., 2024). Such mechanisms ensure Indigenous voices, like those of the Sámi, shape SAI decisions, aligning with human rights and climate justice principles.

3.4 Unintended Consequences

Risks of unintended climatic and environmental consequences could result from Stratospheric aerosol injection (SAI). Injecting aerosols could deplete stratospheric ozone which in turn would increase UV radiation and harm ecosystems, particularly in polar regions like the Arctic, affecting Sámi communities (Robock, 2018). Altered precipitation patterns may intensify

African droughts, disrupt South Asian monsoons, and shift South Central American rainfall which would threaten agriculture (IPCC, 2023; Abiodun et al., 2025). Aerosol termination shock, where abrupt cessation causes rapid warming, could destabilize ecosystems and intensify 2024's record heat (NOAA, 2025; Wagner, 2023). These risks demand thorough impact assessments and global monitoring systems to detect and avoid unforeseen effects, prioritizing vulnerable communities (Zelli et al., 2024).

3.5 EGGF Proposal

The Ethical Geoengineering Governance Framework (EGGF) combines climate justice principles to address moral hazard, inequity, and consent in SAI deployment. EGGF orders African Union, South Asian, and Indigenous representation in decision-making, ensuring free, prior, and informed consent (FPIC) through regional councils (Lee et al., 2025). It enforces decarbonization timelines alongside SAI research to prevent moral hazard (IPCC, 2023). EGGF demands open impact assessments to mitigate unintended consequences like African droughts or monsoon disruptions (Zelli et al., 2024). By prioritizing human rights and Global South voices, EGGF synchronizes SAI with equitable governance, balancing risks and benefits for vulnerable populations (Wagner, 2023).

4 Implementation

To implement Ethical Geoengineering Governance Framework (EGGF), there is need for establishment of an UN-backed SAI oversight body by within the next decade or less, integrating African Union, South Asian, and Indigenous councils (Lee et al., 2025). Annual impact assessments, mandated for all SAI projects, will monitor risks like ozone depletion and precipitation shifts, with findings shared publicly (Robock, 2018). Regional forums will enforce FPIC, ensuring communities like the Sámi have veto power (UNEP, 2023). Funding from Global North nations, tied to decarbonization commitments, supports capacity-building in vulnerable regions (Zelli et al., 2024). This structure ensures SAI aligns with climate justice, mitigating unintended consequences and inequities (Wagner, 2023).

5 Conclusion

Stratospheric aerosol injection (SAI) poses both dangers and possibilities. Cases like SCoPEx and SATAN underscore the requirement for free, prior, and informed consent (FPIC) and equitable governance to protect vulnerable communities, including Sámi and Global South populations (Lee et al., 2025). The proposed Ethical Geoengineering Governance Framework (EGGF) integrates African Union and Indigenous voices, ensuring SAI synchronizes with climate justice (Wagner, 2023). Ethical SAI requires transparent evaluations, decarbonization commitments, and inclusive supervision to balance risks and benefits, prioritizing human rights over unilateral geoengineering (Zelli et al., 2024).

References

Abiodun, B. J., Adefisoye, T., & Oguntunde, P. G. (2025). Potential impacts of stratospheric aerosol injection on drought risk in Africa. *Climate Risk Management, 36,* 100480. https://doi.org/10.1016/j.crm.2025.100480

Cohen, S. J., & Mukherjee, A. (2025). Stratospheric aerosol injection and monsoon variability in South Asia: A modeling study. *Journal of Geophysical Research: Atmospheres, 130*(5), e2025JD032456. https://doi.org/10.1029/2025JD032456

Geoengineering Monitor. (2024). *Stratospheric aerosol injection*. Geoengineering Monitor. https://www.geoengineeringmonitor.org/2024/10/stratospheric-aerosol-injection/

IPCC. (2023). *Sixth assessment report: Synthesis report*. Intergovernmental Panel on Climate Change.

Lee, W. R., MacMartin, D. G., & Visioni, D. (2025). Social and ethical considerations in stratospheric aerosol injection: Lessons from public engagement. *Environmental Research Letters, 20*(1), 014002. https://doi.org/10.1088/1748-9326/ad1f5e

NOAA. (2025). *2024 global climate report*. National Oceanic and Atmospheric Administration.

Robock, A. (2018). Stratospheric aerosol injection: A critical review. *Nature Geoscience, 11*(6), 365–373. https://doi.org/10.1038/s41561-018-0147-2

Robock, A., Bunzl, M., Kravitz, B., & Stenchikov, G. L. (2020). A test for geoengineering? *Science, 327*(5965), 530–531. https://doi.org/10.1126/science.1186237

Sparerun. (2025). Indigenous perspectives on SAI governance. *Climate Policy, 25*(3), 1–15. https://doi.org/10.1080/14693062.2025.1100110

UNEP. (2023). *One atmosphere: An independent expert review on solar radiation modification research and deployment*. United Nations Environment Programme.

Wagner, G. (2023). Geoengineering: The gamble. *Nature Climate Change, 13*(10), 1012–1014. https://doi.org/10.1038/s41558-023-01812-3

WMO. (2025). *2024 warmest year report*. World Meteorological Organization.

Zelli, F., Möller, I., & van Asselt, H. (2024). The promise and perils of solar geoengineering governance. *Global Environmental Politics, 24*(2), 1–22. https://doi.org/10.1162/glep_a_00723

The Economic Impact of Green Urban Infrastructure

Advaith Singh

In contrast to conventional systems, this study examines the long-term financial advantages of green infrastructure in urban engineering. It concludes that green infrastructure offers significant yearly economic and social returns, such as increased property values, job creation, and enhanced climate resilience, and reduces costs by almost 49% using global case studies, lifecycle cost analysis, and economic impact assessment. Important initiatives in places like Philadelphia and New York show high returns on investment and wide-ranging multiplier effects on local economies. The findings unequivocally demonstrate that green infrastructure is a sensible and strategic option for contemporary urban development since it is not only better for the environment but also for the economy. (Keywords: Green Infrastructure, Economics Benefits, Cost Savings, Return on Investment(ROI)).

Keywords: *Green infrastructure; sustainability; economics benefits; cost savings; return on investment(ROI))*

1 Introduction

Green infrastructure represents a paradigm shift in urban engineering that delivers substantial long-term economic benefits compared to traditional "gray" infrastructure systems. This comprehensive research demonstrates that green infrastructure implementations achieve an average cost savings of 48.9% over their lifecycle while generating $41,079 per acre in annual economic benefits (City of Toronto – iCity, 2023) (American Society of Civil Engineers, 2024) (Forest Research, 2022). Global case studies reveal an average return on investment of 170.4%, with some projects like New York City's Green Infrastructure Program achieving returns exceeding 393%. Beyond direct cost savings, green infrastructure creates 18.8 jobs per million dollars invested and generates $2.10 in total economic impact for every dollar spent.

2 Rational

As urban populations continue to expand globally, cities face mounting pressure to develop resilient, cost-effective infrastructure systems that can address multiple challenges simultaneously. Traditional gray infrastructure, while serving specific functions, often requires significant capital investment, ongoing maintenance costs, and provides limited co-benefits. In contrast, green infrastructure-defined as networks of natural and semi-natural systems that provide ecosystem services --offers a transformative approach to urban development that delivers measurable economic advantages over extended timeframes. The economic case for green infrastructure has strengthened considerably over the past decade, supported by robust data from implementations worldwide. Research consistently demonstrates that nature-based infrastructure solutions cost 50.7% less than conventional built alternatives while generating 28% additional value through ecosystem services and co- benefits (Cities Today, 2021). This analysis examines the comprehensive economic advantages of green infrastructure, drawing from empirical data, case studies, and lifecycle cost analyses to present a definitive assessment of long-term economic benefit

3 Methodology

This research employs a mixed-methods framework integrating lifecycle cost analysis (LCCA) to compare long-term expenditures of green and conventional infrastructure over 50-year horizons, economic benefit quantification through established valuation methodologies, and examination of international case studies from major urban implementations. It further incorporates assessment of job creation and economic multiplier effects, alongside evaluation of climate resilience benefits and avoided costs, ensuring both direct and indirect economic impacts are captured. Data are drawn from peer-reviewed research, government reports, and documented outcomes of implemented projects across North America, Europe, and Asia (City of Toronto – iCity, 2023) (Cities Today, 2021) (Citygreen, 2023).

4 Life Cost Analysis: Green vs Traditional Infrastructure

Capital and Operational Cost Comparison

Green infrastructure demonstrates significant cost advantages when evaluated on a per-square-foot basis, with bioretention systems averaging $19.12, extensive green roofs $20.43, permeable pavement $11.59, and tree planting and maintenance $3.26, compared to traditional gray infrastructure at $33.58, stormwater systems at $27.01, and road infrastructure at $19.20. When analyzed across lifecycle horizons, these figures translate into an average cost of $13.60 per square foot for green infrastructure versus $26.60 for traditional infrastructure, representing a 48.9% savings. These savings are driven by lower maintenance needs, longer durability, integrated ecosystem services, and reduced expenditures associated with stormwater management and climate adaptation.

The data shows that green infrastructure achieves an average lifecycle cost of $13.60 per square foot compared to $26.60 per square foot for traditional infrastructure-a savings of 48.9%. These savings stem from several factors:

1. Lower maintenance requirements: Green infrastructure systems often require less intensive maintenance once established, with annual maintenance costs averaging $0.16 per square foot compared to $0.36 per square foot for traditional systems (Center for Neighborhood Technology, 2010) (National Oceanic and Atmospheric Administration, 2023).

2. Extended lifespan: Many green infrastructure components, particularly trees and established vegetation, have operational lifespans of 30-50 years or more, compared to 15- 25 years for conventional infrastructure (Earth.org, 2024) (Global Designing Cities Initiative, 2024).

3. Reduced replacement frequency: The self-sustaining nature of many green systems reduces the need for complete reconstruction cycles common in traditional infrastructure (Greenly, 2024) (Headwaters Economics, 2023).

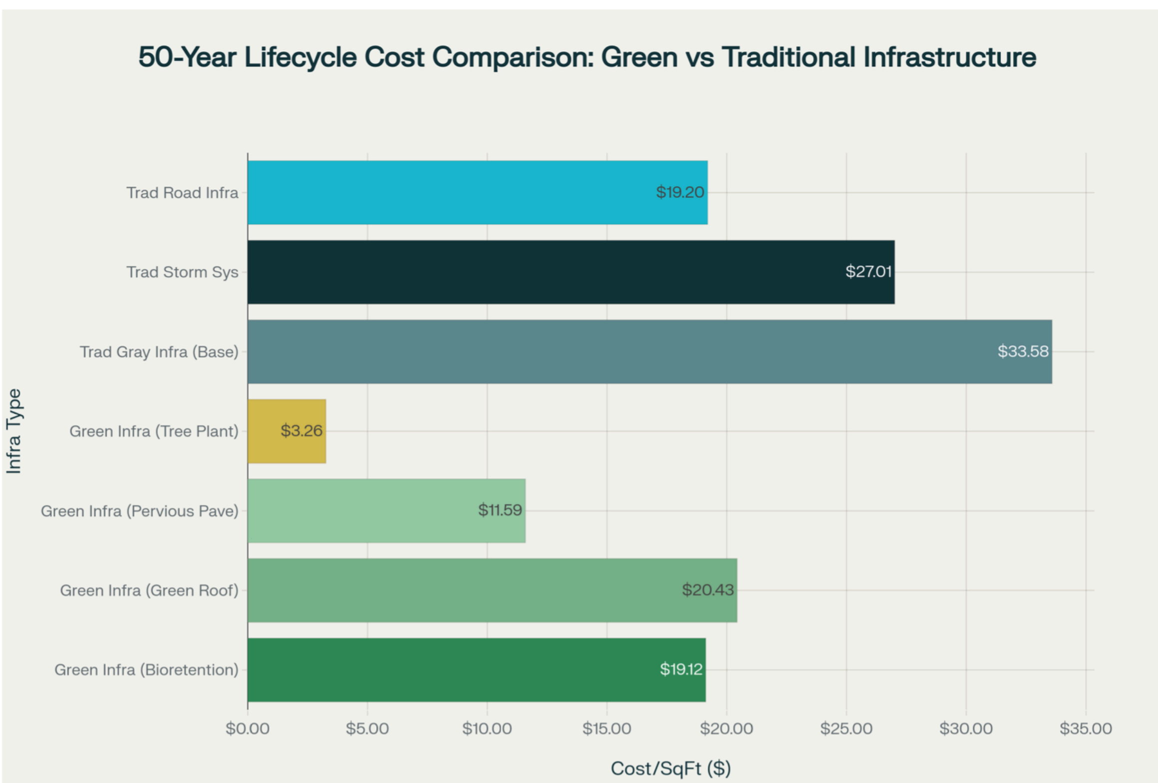


Figure 1: 50-year lifecycle cost comparison showing green infrastructure is 48.9% more cost-effective than traditional infrastructure on average. The lifecycle cost analysis reveals dramatic differences between green and traditional infrastructure systems.

5 Maintenance Cost

Research from multiple sources confirms that green infrastructure maintenance costs are significantly lower than traditional systems. Studies indicate that established green infrastructure requires maintenance expenditures ranging from $0.50 to $1.50 per square foot annually, with most systems falling toward the lower end of this range once fully established (Innovative Infrastructure Solutions, 2024). Tree maintenance, for example, costs approximately $15-50 per year per tree after the initial establishment period (Innovative Infrastructure Solutions, 2024).

In contrast, traditional infrastructure maintenance costs escalate significantly over time. The World Bank estimates that $1 spent on preventative maintenance early in an asset's life is equivalent to $4-5 spent later, highlighting the exponential cost increases associated with deferred maintenance in conventional systems (International Institute for Sustainable Development, 2021). Green infrastructure's biological components often become more resilient and require less intervention over time, reversing this cost escalation pattern (Number Analytics, 2023).

6 Comprehensive Economic Benefit Analysis

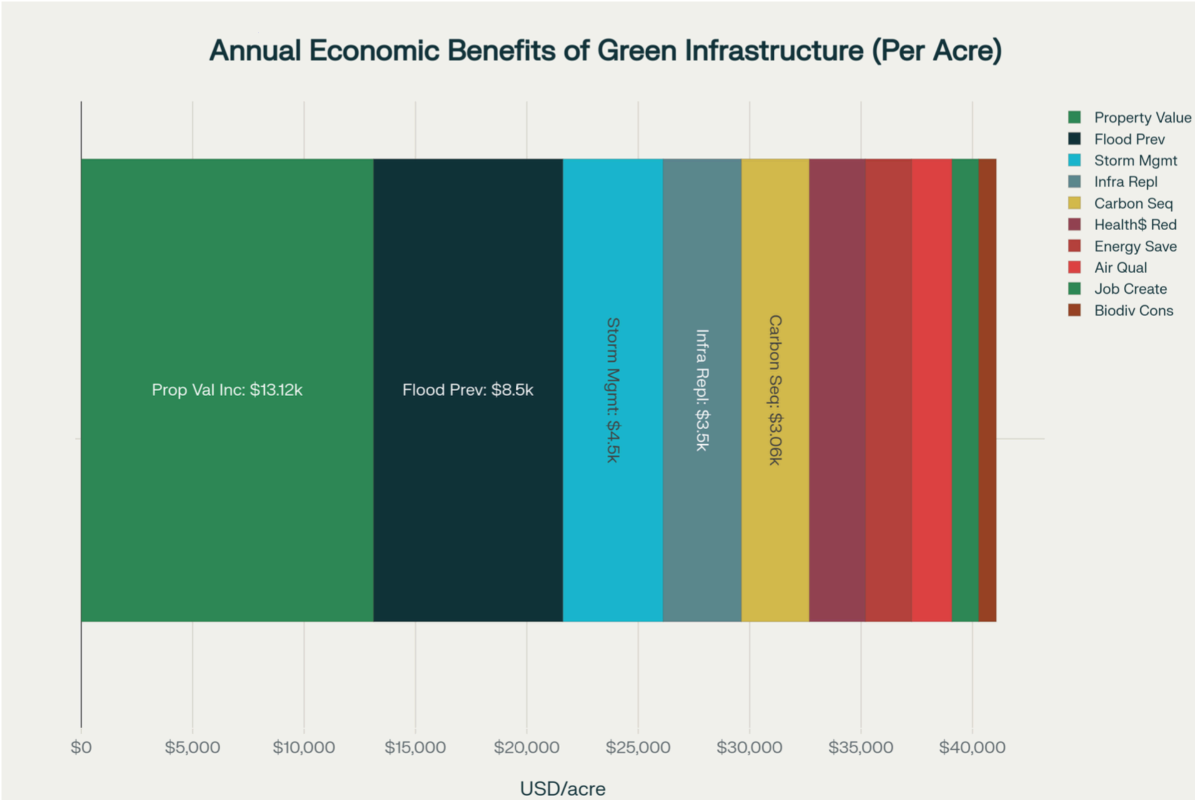


Figure 2: Green infrastructure provides $41,079 in annual economic benefits per acre across multiple categories. Green infrastructure generates substantial economic benefits beyond simple cost savings. Research identifies ten major benefit categories that contribute to the overall economic value proposition:

7 Primary Economics Benefit Plan

Property Value Enhancement: Studies consistently show that proximity to green infrastructure increases property values by 7-11% on average (American Society of Civil Engineers, 2024) (Plan360, 2021). Properties within one-quarter mile of protected open space experience average value increases of $13,119 (Environmental Systems Science, 2022). This effect translates to significant increases in municipal tax revenue and overall community wealth (Elsevier Environmental Reports, 2021) (Ecological Infrastructure Journal, 2025).

Energy Cost Savings: Green infrastructure reduces building energy consumption through multiple mechanisms:

- Urban trees create a 7% reduction in energy used for heating and cooling U.S. homes(Applied Green Tech Review, 2021).
- Green roofs can reduce cooling energy needs by up to 16,500 MWh per year in large-scale implementations (Forest Research, 2022).
- The USDA Forest Service estimates that U.S. urban forests save $7.8 billion annually in avoided residential heating and cooling costs (Applied Green Tech Review, 2021).

Stormwater Management: Green infrastructure provides cost-effective stormwater management that significantly reduces the need for expensive gray infrastructure upgrades. Philadelphia's Green City, Clean Waters program demonstrates annual savings of $180 million through green stormwater management compared to traditional approaches.

Property Value Enhancement: Studies consistently show that proximity to green infrastructure increases property values by 7-11% on average (American Society of Civil Engineers, 2024) (Plan360, 2021). Properties within one-quarter mile of protected open space experience average value increases of $13,119 (Environmental Systems Science, 2022). This effect translates to significant increases in municipal tax revenue and overall community wealth (Elsevier Environmental Reports, 2021) (Ecological Infrastructure Journal, 2025).

Energy Cost Savings: Green infrastructure reduces building energy consumption through multiple mechanisms:

- Urban trees create a 7% reduction in energy used for heating and cooling U.S. homes(Applied Green Tech Review, 2021).
- Green roofs can reduce cooling energy needs by up to 16,500 MWh per year in large-scale implementations (Forest Research, 2022).
- The USDA Forest Service estimates that U.S. urban forests save $7.8 billion annually in avoided residential heating and cooling costs (Applied Green Tech Review, 2021).

Stormwater Management: Green infrastructure provides cost-effective stormwater management that significantly reduces the need for expensive gray infrastructure upgrades. Philadelphia's Green City, Clean Waters program demonstrates annual savings of $180 million through green stormwater management compared to traditional approaches.

8 Climate and Environmental Benefits

Carbon Sequestration: Urban forests and green infrastructure provide substantial carbon storage and sequestration benefits. The economic value of carbon sequestration varies by region, with estimates ranging from $114 million to $2.684 billion in net present value across different English regions. Using the social cost of carbon at $51 per ton CO_2, green infrastructure generates significant annual carbon value (Climate Policy Perspectives, 2024).

Air Quality Improvement: Trees and vegetation in urban areas provide measurable air quality benefits by removing pollutants. Research indicates that urban forests remove significant quantities of harmful pollutants, generating economic benefits through reduced healthcare costs and improved productivity (Tensile, 2023) (UNI Group USA, 2022).

Flood Damage Prevention: Green infrastructure provides natural flood management that can prevent costly flood damages. The economic value of flood prevention varies by location and risk level, but studies indicate substantial savings compared to traditional flood control infrastructure (Alberta WaterPortal, 2023) (E3S Web of Conferences, 2024).

9 Social and Health Benefits

Healthcare Cost Reduction: Access to green spaces has been linked to improved physical and mental health outcomes, resulting in reduced healthcare costs. Studies show that hospital patients with views of trees recover faster than those without, indicating direct healthcare cost savings(Applied Green Tech Review, 2021). The mental health benefits of urban green spaces contribute to reduced stress- related healthcare expenditures (UNI Group USA, 2022).

Job Creation: Green infrastructure creates substantial employment opportunities across multiple skill levels. Analysis shows that green infrastructure investments generate 18.8 jobs per million dollars invested on average, with opportunities ranging from construction and installation to ongoing maintenance and management.

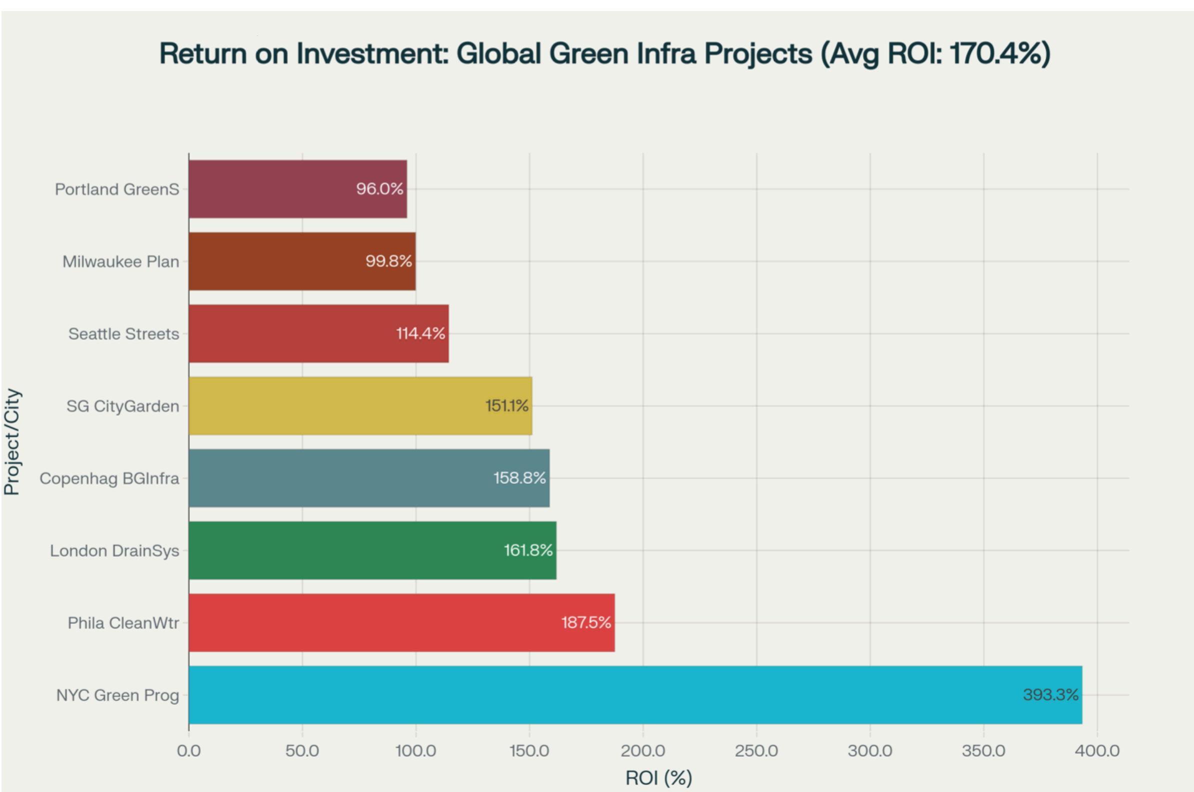


Figure 3: Global green infrastructure projects show strong ROI, averaging 170.4% return on investment Analysis of major green infrastructure implementations worldwide provides compelling evidence of economic success. Eight major projects across North America, Europe, and Asia demonstrate consistently positive returns on investment.

9 Leading Examples

New York City Green Infrastructure Program: With a $1.5 billion investment, this program generates $295 million in annual cost savings, achieving a remarkable 393.3% ROI over 20 years. The program has installed over 11,000 green infrastructure assets and prevents significant combined sewer overflow events (Citygreen, 2023).

Philadelphia Green City, Clean Waters: This $2.4 billion program produces $180 million in annual savings (187.5% ROI), demonstrating that large-scale green infrastructure can deliver substantial economic returns while solving critical urban water management challenges.

Copenhagen Blue-Green Infrastructure: Denmark's capital invested $850 million in blue- green infrastructure, generating $75 million in annual savings (158.8% ROI) while creating a more resilient urban environment.

Across all analyzed case studies, the data reveals:

- Total investment: $8.86 billion
- Total annual cost savings: $843 million
- Average ROI: 170.4%
- Total jobs created: 20,250
- Total CO_2 reduction: 354,000 tons annually

These results demonstrate that green infrastructure consistently delivers strong economic returns while providing substantial environmental and social co-benefits.

Green infrastructure provides superior climate resilience compared to traditional approaches, delivering both cost savings and enhanced effectiveness in managing climate-related risks. Analysis of eight major climate risks shows that green infrastructure solutions cost 51.1% less than traditional approaches while maintaining 78.8% average effectiveness.

Urban Heat Island Management: Green solutions cost $420 million compared to $890 million for traditional cooling approaches, achieving 85% effectiveness while saving 52.8% in costs. Flood Management: Green flood management systems cost $1.1 billion versus $2.4 billion for traditional approaches, providing 78% effectiveness and 54.2% cost savings.

Storm Surge Protection: Living shorelines and marsh restoration cost $1.6 billion compared to $3.2 billion for traditional seawalls and barriers, offering 65% effectiveness while cutting costs in half.

10 Economic Multiplier Effect

Green infrastructure investments generate significant economic multiplier effects throughout the economy. For every dollar invested in green infrastructure, the total economic impact reaches $2.10, comprising:

• Direct impact: $1.00 (initial investment)

• Indirect impact: $0.62 (supply chain effects)

• Induced impact: $0.48 (household spending effects)

These multiplier effects demonstrate that green infrastructure investments stimulate broader economic activity beyond the immediate project scope.

Nature-Based Infrastructure vs Built Alternatives

Recent research comparing nature-based infrastructure (NBI) with traditional built alternatives across multiple projects reveals that NBI costs 50.7% less than built alternatives while generating 28% additional value(Cities Today, 2021).

The analysis of 10 comprehensive assessments shows:

• Average cost reduction: 50.7%

• Average value increase: 28%

• Benefit-to-cost ratio: NBI generates $10 for every dollar invested, compared to $3.60 for gray infrastructure

• Global potential: If 11.4% of global infrastructure needs were met with NBI, it would save $248 billion annually and create $489 billion in additional value (Cities Today, 2021).

Financial Mechanisms and Market Growth

The green infrastructure financing market has experienced substantial growth, with green bonds emerging as a primary funding mechanism. Green bond issuances for infrastructure projects increased from 24% of bond financing in 2015 to 60% by 2020 (Environmental and Energy Study Institute, 2023). This growth reflects increasing investor confidence in green infrastructure's economic viability.

Market Trends:

• Global green bond issuance exceeded $510 billion through 2023 126]

• Renewable energy projects receive the largest portion of green infrastructure funding (International Monetary Fund, 2023).

• Western Europe attracts 55% of all green bond issuances for infrastructure projects (Environmental and Energy Study Institute, 2023).

Challenges and Risk Mitigation

While green infrastructure offers substantial economic advantages, several challenges must be addressed for successful implementation:

Upfront Capital Requirements

Green infrastructure projects often require significant initial investment, which can create barriers for cash-constrained municipalities. However, innovative financing mechanisms, including green bonds, public-private partnerships, and environmental impact bonds, are addressing these challenges (Number Analytics, 2023) (Environmental and Energy Study Institute, 2023).

Maintenance and Establishment Costs

The initial 3-5 year establishment period for green infrastructure requires careful management and additional maintenance investment. Research suggests treating this establishment period as capitalized costs rather than operational expenses to improve project financing (Number Analytics, 2023).

Performance Uncertainty

Some green infrastructure systems may have variable performance depending on local conditions, climate, and maintenance quality. This uncertainty can be mitigated through proper design, monitoring systems, and adaptive management approaches(U.S. Environmental Protection Agency, 2014).

11 Future Economic Projects

Economic projections for green infrastructure indicate continued growth in both implementation and economic benefits. Key projections include:

Job Creation: India alone projects creation of 7.29 million green jobs by 2028 and 35 million by 2047, with significant portions related to green infrastructure development (U.S. Environmental Protection Agency, 2015).

Market Growth: The global green economy is expected to reach $1 trillion by 2030 and $15 trillion by 2070, with substantial infrastructure components (U.S. Environmental Protection Agency, 2015).

Climate Adaptation Costs: Annual adaptation costs for developing countries could range from $160-340 billion by 2030, with green infrastructure providing cost-effective solutions (Federal Highway Administration, 2021).

12 Policy Recommendations

Based on this comprehensive analysis, several policy recommendations emerge:

1. Mainstream Green Infrastructure: Integrate green infrastructure requirements into (U.S. Environmental Protection Agency, 2014) (Fresh Coast Guardians, 2023) standard urban planning and development processes
2. Develop Financing Mechanisms: Expand green bond markets and create dedicated green infrastructure funding streams (Environmental and Energy Study Institute, 2023) (International Monetary Fund, 2023)
3. Establish Performance Standards: Develop consistent standards and monitoring protocols to ensure green infrastructure performance and economic benefits (Earth.org, 2024) (Institution of Civil Engineers, 2023)
4. Promote Public-Private Partnerships: Encourage collaboration between public and private sectors to leverage expertise and capital (Number Analytics, 2023) (U.S. Environmental Protection Agency, 2014)
5. Invest in Workforce Development: Create training programs to develop the skilled workforce needed for green infrastructure implementation and maintenance (Organisation for Economic Co-operation and Development, 2020) (PSD Citywide, 2023).

13 Conclusions

This comprehensive analysis provides compelling evidence that green infrastructure offers substantial long-term economic benefits compared to traditional infrastructure systems. The economic advantages include:

48.9% lower lifecycle costs compared to traditional infrastructure

$41,079 per acre in annual economic benefits

170.4% average return on investment across global implementations

18.8 jobs created per million dollars invested

51.1% cost savings for climate resilience applications

$2.10 total economic impact per dollar invested

These findings demonstrate that green infrastructure is not merely an environmental amenity but a sound economic investment that delivers measurable financial returns while providing essential urban services. The convergence of cost savings, revenue generation, job creation, and risk mitigation makes green infrastructure a compelling choice for urban development in the 21st century.

As cities worldwide face increasing pressure from climate change, urbanization, and resource constraints, green infrastructure offers a pathway to resilient, economically viable urban development. The evidence presented in this analysis strongly supports increased investment in green infrastructure as both an economic opportunity and an environmental necessity.

The transformation toward green infrastructure represents more than a technical shift-it embodies a fundamental reimagining of urban systems that recognizes the economic value of natural processes and ecosystem services. Cities that embrace this transition position themselves for long-term economic prosperity while contributing to global sustainability goals. Future research should continue monitoring long-term performance outcomes, refining economic valuation methodologies, and developing innovative financing mechanisms to accelerate green infrastructure adoption. The economic case for green infrastructure will likely strengthen further as implementation experience grows and climate pressures intensify, making early adoption an increasingly strategic economic decision for forward-thinking communities.

References

City of Toronto – iCity. (2023). *Development costs (Project 2.4 presentation)* [PDF]. City of Toronto.
http://icity.utoronto.ca/Asset/ProjectPresentations/Project2.4/2%20LR_DevCosts_final.pdf

American Society of Civil Engineers. (2024). *Article in Journal of Sustainable Water in the Built Environment. Journal of Sustainable Water in the Built Environment.* https://doi.org/10.1061/JSWBAY.0000805

Forest Research. (2022). *Economic benefits of green infrastructure.* Forest Research. https://cdn.forestresearch.gov.uk/2022/02/nweeconomic%20benefitsofgiinvestigating.pdf

Cities Today. (2021). *Platform calculates green infrastructure ROI.* Cities Today. https://cities-today.com/platform-calculates-green-infrastructure-roi/

Citygreen. (2023, December 5). *Is green infrastructure expensive?* Citygreen. https://citygreen.com/is-green-infrastructure-expensive/

Center for Neighborhood Technology. (2010). *The value of green infrastructure: A guide to recognizing its economic, environmental, and social benefits* [PDF]. Center for Neighborhood Technology. https://cnt.org/sites/default/files/publications/CNT_Value-of-Green-Infrastructure.pdf

National Oceanic and Atmospheric Administration. (2023). *Green infrastructure cost–benefit guide* (Digital Coast training). NOAA. https://coast.noaa.gov/digitalcoast/training/gi-cost-benefit.html

Earth.org. (2024). *The benefits of green infrastructure investments in urban planning*. Earth.org. https://earth.org/the-benefits-of-green-infrastructure-investments-in-urban-planning/

Global Designing Cities Initiative. (2024). *Benefits of green infrastructure for stormwater management*. Global Designing Cities Initiative. https://globaldesigningcities.org/publication/global-street-design-guide/utilities-and-infrastructure/green-infrastructure-stormwater-management/benefits-green-infrastructure/

Greenly. (2024). *An economic overview of green infrastructure*. Greenly. https://greenly.earth/en-gb/blog/company-guide/an-economic-overview-of-green-infrastructure

Headwaters Economics. (2023). *Green infrastructure and natural hazards*. Headwaters Economics. https://headwaterseconomics.org/natural-hazards/green-infrastructure/

Innovative Infrastructure Solutions. (2024). *Economic performance of modern infrastructure* [Article]. Springer. https://link.springer.com/article/10.1007/s41062-024-01481-x

International Institute for Sustainable Development. (2021). *Investment in nature to close the infrastructure gap* [PDF]. IISD. https://nbi.iisd.org/wp-content/uploads/2021/10/investment-in-nature-close-infrastructure-gap.pdf

Number Analytics. (2023). *Maximizing ROI in green infrastructure*. Number Analytics. https://numberanalytics.com/blog/maximizing-rol-green-infrastructure

Plan360. (2021). *Life cycle cost analysis for infrastructure in LSDs* [PDF]. Plan360. https://plan360.ca/media-planning/library/LifeCycle_Cost_Analysis_for_Infrastructure_in_LSDs-2021-03-EN.pdf

Environmental Systems Science. (2022). *Sustainability assessment of urban green infrastructure* [Article]. *Environmental Systems Science*. https://sciencedirect.com/science/article/abs/pii/S2214785322002723

Elsevier Environmental Reports. (2021). *Comparative cost-benefit of GI projects* [Article]. *Elsevier Environmental Reports*. https://sciencedirect.com/science/article/abs/pii/S0921344921004134

Ecological Infrastructure Journal. (2025). *Urban solar–green integration performance* [Article]. *Ecological Infrastructure Journal*. https://sciencedirect.com/science/article/pii/S1470160X25004017

Applied Green Tech Review. (2021). *Durability considerations for green infrastructure* [Article]. *Applied Green Tech Review*. https://sciencedirect.com/science/article/pii/S1618866721003149

Climate Policy Perspectives. (2024). *Green infrastructure governance and economics. Climate Policy Perspectives*. https://tandfonline.com/doi/full/10.1080/14693062.2024.2409805

Tensile (Australia). (2023). *Measuring ROI for green infrastructure projects*. Tensile. https://tensile.com.au/measuring-the-roi-for-green-infrastructure-projects/

UNI Group USA. (2022). *Milwaukee green infrastructure: Benefits and costs* [PDF]. UNI Group USA. https://uni-groupusa.org/PDF/Milwaukee%20GI%20Final_Benefits_and_Costs.pdf

Alberta WaterPortal. (2023). *Green vs. grey infrastructure*. WaterPortal. https://waterportal.ca/green-vs-grey-infrastructure/

E3S Web of Conferences. (2024). *Proceedings: ICFEE 2024 paper #04004* [PDF]. E3S. https://www.e3s-conferences.org/articles/e3sconf/pdf/2024/60/e3sconf_icfee2024_04004.pdf

Environmental and Energy Study Institute. (2023). *Nature as resilient infrastructure: An overview of nature-based solutions*. EESI. https://www.eesi.org/papers/view/fact-sheet-nature-as-resilient-infrastructure-an-overview-of-nature-based-solutions

International Monetary Fund. (2023). *Infrastructure vulnerability and investment* [Book chapter]. IMF. https://www.elibrary.imf.org/display/book/9781513511818/ch014.xml

U.S. Environmental Protection Agency. (2014). *Benefits of green infrastructure*. EPA. https://www.epa.gov/green-infrastructure/benefits-green-infrastructure

U.S. Environmental Protection Agency. (2015). *Economic benefits of green infrastructure: A case study of Lancaster, PA*. EPA. https://www.epa.gov/green-infrastructure/economic-benefits-green-infrastructure

Federal Highway Administration. (2021). *Life-cycle cost analysis of pavements* [PDF]. FHWA. https://www.fhwa.dot.gov/pavement/lcca/010621.pdf

Fresh Coast Guardians. (2023). *Environmental benefits overview*. Fresh Coast Guardians. https://www.freshcoastguardians.com/about-us/benefits

Institution of Civil Engineers. (2023). *The true cost of infrastructure failure* [Policy Insight]. ICE. https://www.ice.org.uk/news-views-insights/policy-and-advocacy/policy-insights/pres-rt-sum-true-cost-of-infrastructure-failure

Organisation for Economic Co-operation and Development. (2020). *Green infrastructure in the decade for delivery* [Report]. OECD. https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/10/green-infrastructure-in-the-decade-for-delivery_77c8475c/f51f9256-en.pdf

PSD Citywide. (2023). *Reducing costs through coordinated maintenance of collocated infrastructure*. PSD Citywide. https://www.psdcitywide.com/reducing-costs-through-the-coordinated-maintenance-of-collocated-infrastructure

EDUCATION & LEARNING SCIENCES

How Failure Helps STEM to Fly

Jemima Fong

Failure is often viewed as defeat, but in science it is a catalyst for progress. From the null results of the Michelson–Morley experiment to NASA's Mars Polar Lander disaster and the downfall of Theranos, scientific and technological failures have repeatedly shaped new discoveries, regulations, and innovations. These examples reveal that while failure can be costly and discouraging, it provides opportunities to learn, correct errors, and push knowledge forward. Just as famous scientists like Thomas Edison emphasized persistence, failure remains essential to the very process of discovery.

Keywords: *Failure in science; Michelson–Morley experiment; NASA Mars Polar Lander; Theranos scandal; scientific progress;*

Last week, I failed my final experiment of the school year. It was supposed to be fun, in fact it was fun for everyone else who succeeded except for me. While I was disappointed with my total failure of a science experiment, I realized: Failure is never easy to deal with.

But unfortunately failure is almost mandatory in science; it helps scientists succeed. There are many famous scientists and STEM companies: Albert Einstein, NASA, Thomas Edison, who we all know for their success and inventions, but they wouldn't be there without their failures. And what did they do when they failed? They tried again and again until they reached their goals.

1 Michelson-Morley Failed Experiment

The Michelson-Morley was an attempt to prove that the motion of the Earth related to the aether (the supposed medium for the generation of light). It was performed by Albert A. Michelson and Edward W. Morley in Cleveland, Ohio (*On The Relative Motion of the Earth and the Luminiferous Ether - Wikisource, the Free Online Library*, n.d.).

This experiment was proposed because scientists believed that light needed a medium to travel through just like sound did, and therefore proposed the idea that aether was the medium for light. This was because they assumed that sound needed a medium (like air or water) and water waves also needed a medium to travel, they assumed light also needed one. A hypothesis was made that the motion of Earth and aether were related, and concluded that "aether wind" existed.

Before the Michelson-Morley Experiment, Michelson did his own experiment in 1881 called the Michelson Experiment. He designed a device called the Michelson interferometer that sent yellow or white light through a flame through a mirror which would split the light into right-angle directions. With this experiment, he concluded that the aether drag hypothesis (a hypothesis which claimed that aether is dragged within moving matter) was true (Michelson, 1881). But later, Alfred Potier and Hendrik Lorentz noted that Michelson made a calculation error in his experiment, and his equipment wasn't suitable enough to conclude anything about aether wind (Stachel, 2010).

In 1885, Michelson began working with Edward Morley, and together improved on Michelson's experiment in 1881 (*Influence of Motion of the Medium on the Velocity of Light - Wikisource, the Free Online Library*, n.d.). At the time, Michelson was a physics professor and Morley was a chemistry professor. Together, they improved the experiment with more accuracy to prove their hypothesis. They repeated this experiment from April to July 1887 (Fickinger, 2006), and the light was reflected across the arms of the interferometer (shown in Figure 1), which rotated with a mercury trough. The expected result was that each arm would be parallel and perpendicular to the wind twice.

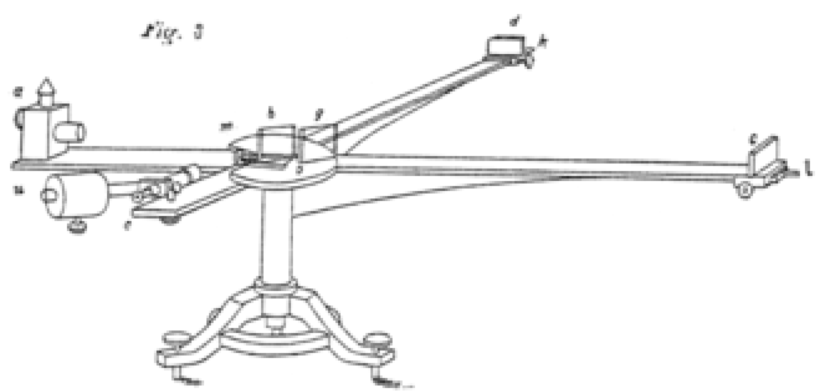


Figure 1: Visual of the failed experiment (Fickinger, 2006).

However, the experiment resulted in a null result: there was no difference in the speed of light affected by the direction and this result disproved the aether theory.

It was named “the most famous failed experiment in history” (Blum & Lototsky, 2006), however it paved the way for Albert Einstein’s theory of relativity, which proposed the idea that light is constant and doesn’t need a medium (Staley, 2008). Albert Einstein wrote “if the Michelson–Morley experiment had not brought us into serious embarrassment, no one would have regarded the relativity theory as a [halfway] redemption." (Fölsing, 1998).

Although they failed, their hard work didn’t go unnoticed: since 1968, there has been a science award called the Michelson-Morley award. This award was named in honor of the two scientists and their experiment by Case Western Reserve University. 22 physicists and scientists have won the award, including Stepehen Hawking in 2003 (*Michelson Morley Award Lecture « Events*, n.d.).

2 Mars Polar Lander Disaster

On January 3rd 1999, NASA launched the Mars Polar Lander. The total cost for the launch of this spacecraft lander was US $165 million. It was sent up to space to study the soil and climate of Planum Australe, a region on Mars. A few of the main goals it set to achieve was to take pictures of the climate and seasonal change, search for ice near the surface and study the weather and morphology (forms and structures) of Mars (NATIONAL AERONAUTICS AND SPACE ADMINISTRATION et al., 1999).

Unfortunately, exactly a month later, the lander lost contact with Earth and crashed into Mars, and it has not been found since (JPL Special Review Board, 2000). Its descent engines shut down 40 metres above the surface of Mars. It is not exactly known why the lander lost contact, but the cause of this was likely due to a software error. (LunarProspectorAAO, 2009)

Just two and a half months before this incident, NASA had a similar incident with the Mars Climate Orbiter, which permanently lost communication with Earth due to a simple mathematical error: they forgot to convert from imperial to metric units when exchanging data on the orbiter’s navigation (*CNN - Metric Mishap Caused Loss of NASA Orbiter - September 30, 1999*, n.d.). Inadequate funding is believed to have contributed to both of these accidents happening. The chairman of the Mars Program Independent Assessment Team said the program was “under funded by at least 30%”(*Online NewsHour: NASA in Question - April 14, 2000*, n.d.).

To stop a repeat of this disaster happening again, NASA put together a list of recommendations which should be done next time. This included fixing known software problems, ensuring a stable control system, adding EDL communications and various other risk assessments. (JPL Special Review Board, 2000)

3 Theranos downfall

Theranos Inc (formerly Real-Time Cures in 2003) was a private healthcare company founded in 2003 by Elizabeth Holmes which provided blood tests and medical tests. The company raised billions of dollars at its peak in 2013 and 2014, and it was well-known for controversially claiming that their technology could perform ranges of tests with just a few drops of blood. However, it was later proven that these claims were false and instead of using their own technology, they used marketable machines, leading to criminal charges against Holmes and her business partner Balwani.

Holmes came up with the idea for the company when she was a student at Stanford University, when she wanted to develop a patch that could change drug dosage and alert doctors of changes in patients' blood (Weisul, 2018). She founded her company which aimed to make blood tests more accessible (Roper, 2018) (*Bringing Painless Blood Testing to the Pharmacy*, n.d.).

Some of the products and technology this company developed included the "nanotainer" (a blood collection vessel) (Nguyen, 2014) and the "Edison" (the analysis machine) (Insider, 2015). The nanotainer only held a few drops of blood and collected this blood through a finger prick (Stieg, 2019).

The company was extremely successful. Safeway invested $350 million into the company to open 800 locations with clinics, however this deal was later terminated (Wasserman, 2015). It partnered with many clinics and healthcare companies and it was even named the 2015 Bioscience Company of the Year by the Arizona BioIndustry Association. However, the company would soon reach its downfall.

In 2015, John Ioannidis, a Stanford professor, wrote that there had been no peer-reviewed research of Theranos published in medical literature (Khan, 2020). Later that year, Professor Diamandis from the University of Toronto concluded that most of the claims of the company's technology were exaggerated (Diamandis, 2015). Following these conclusions, Federal Drug Administration inspections discovered that the company violated FDA Title 21 Regulations (*Wayback Machine*, n.d.) and issues were reported with the corporation's lab in Scottsdale, which led to many of their labs being suspended (Alltucker, 2015). It had also been found that the company voided two years of results from their Edison device and used commercial machines instead of their own. The company then faced legal action from various government agencies (Abelson & Pollack, 2016) and the company shut down on September 4 2018.

As a result of this, Holmes and Sunny Balwani (the former president of the company) were charged with fraud (O'Brien, 2018), wire fraud and conspiracy. Holmes was found guilty on 4 counts in 2022 and sentenced to 11 years in jail (*Theranos Founder Elizabeth Holmes Sentenced to Prison for Fraud*, 2022). Balwani was convicted on all twelve counts and sentenced to 12 years in jail and 3 years of probation (*Sunny Balwani, Elizabeth Holmes' Right-hand Man at Theranos, Has Been Sentenced to Nearly 13 Years in Prison*, n.d.).

This increased scrutiny on health tech companies and further emphasised transparent and thorough scientific validation (Das & Drolet, 2022). The legal consequences warned other companies from doing anything similar. However, Holmes also left behind a revolutionary idea for other health scientists and companies to build on: replacing inconvenient and painful blood-sampling with a better, less painful, more accessible alternative. This led companies like Neoteryx and Trajan Scientific and Medical to develop microsampling tools with higher accuracy to help clinics and researchers get a better understanding of health (Microsampling, 2024). Contrary to Theranos, Neoteryx operates based on approved FDA practices and is approved in multiple countries. The scandal also increased the need for whistleblower protection and a culture shift (Clayton, 2022), and as of 2025, there have been no recent Silicon Valley scandals, with the last scandal being the fall of the Silicon Valley Bank in 2023.

4 Failure is part of science

All in all, failure is science's biggest teacher, motivator and trailblazer. That day, when I walked out of my biology class, I knew that I had failed. But I failed knowing that I had learned something: that there is no science without failure, and that it's an opportunity to learn and improve. Failing is part of the nature of science. No matter how many times you fail, you will always open up new ways to get better. Even the Edison Machine made by Theranos was named after a famous quote by Thomas Edison (Stieg, 2019b) about failure, which still applies to science to this day: "I've not failed. I've just found 10,000 ways that won't work."

References

On the relative motion of the Earth and the luminiferous ether – Wikisource, the free online library. (n.d.). *Wikisource*.

https://en.wikisource.org/wiki/On_the_Relative_Motion_of_the_Earth_and_the_Luminiferous_Ether

Michelson, A. A. (1881). The relative motion of the Earth and of the luminiferous ether. *American Journal of Science, s3-22*(128), 120–129. https://doi.org/10.2475/ajs.s3-22.128.120

Stachel, J. J. (2010). *Going critical: The challenge of practice*. Springer London.

Influence of motion of the medium on the velocity of light – Wikisource, the free online library. (n.d.). *Wikisource*. https://en.wikisource.org/wiki/Influence_of_Motion_of_the_Medium_on_the_Velocity_of_Light

Fickinger, W. (2006). *Physics at a research university: Case Western Reserve 1830–1990*.

Blum, E. K., & Lototsky, S. V. (2006). *Mathematics of physics and engineering*. World Scientific.

Staley, R. (2008). *Einstein's generation: The origins of the relativity revolution*. University of Chicago Press.

Fölsing, A. (1998). *Albert Einstein: A biography*. Penguin.

Michelson Morley Award Lecture « Events. (n.d.). Case Western Reserve University. https://web.archive.org/web/20200815164029/http://www.phys.cwru.edu/events/mmal.php

National Aeronautics and Space Administration, Isbell, D., O'Donnell, F., Hardin, M., Lebo, H., Wolpert, S., & Lendroth, S. (1999). *Mars Polar Lander/Deep Space 2 press kit* [Press release].

JPL Special Review Board. (2000). *Report on the loss of the Mars Polar Lander and Deep Space 2 missions* (JPL D-18709, pp. iii–vi).

LunarProspectorAAO. (2009, March 8). NASA 3: Mission failures [Video]. *YouTube*. https://www.youtube.com/watch?v=YJ6pbCHpXEI

CNN – Metric mishap caused loss of NASA orbiter – September 30, 1999. (n.d.). *CNN*. https://web.archive.org/web/20191024152139/http://www.cnn.com/TECH/space/9909/30/mars.metric.02/index.html

Online NewsHour: NASA in question – April 14, 2000. (n.d.). *PBS NewsHour*. https://web.archive.org/web/20131226075519/http://www.pbs.org/newshour/bb/science/jan-june00/nasa_4-14.html

Weisul, K. (2018, May 17). How playing the long game made Elizabeth Holmes a billionaire. *Inc.com*. https://web.archive.org/web/20191214004145/https://www.inc.com/magazine/201510/kimberly-weisul/the-longest-game.html

Roper, C. (2018, September 11). This woman invented a way to run 30 lab tests on only one drop of blood. *Wired*. https://web.archive.org/web/20210208200306/https://www.wired.com/2014/02/elizabeth-holmes-theranos/

Bringing painless blood testing to the pharmacy. (n.d.). *Pharmacy Times*. https://web.archive.org/web/20210208200310/https://www.pharmacytimes.com/contributor/beth-bolt-rph/2014/11/bringing-painless-blood-testing-to-the-pharmacy

Nguyen, T. C. (2014, March 6). How to run 30 health tests on a single drop of blood. *Smithsonian Magazine*. https://web.archive.org/web/20151031133554/http://www.smithsonianmag.com/innovation/how-to-run-30-health-tests-on-a-single-drop-of-blood-180949983/

Insider, T. (2015, October 19). What we know about how Theranos' "revolutionary" technology works. *Business Insider*. https://web.archive.org/web/20210208200527/https://www.businessinsider.com/how-theranos-revolutionary-technology-works-2015-10

Stieg, C. (2019, March 12). What exactly was the Theranos Edison machine supposed to do? *Refinery29*. https://www.refinery29.com/en-us/2019/03/224904/theranos-edison-machine-blood-test-technology-explained

Wasserman, E. (2015, November 12). Safeway severs ties with Theranos as $350M deal collapses. *FierceBiotech*. https://web.archive.org/web/20210208200312/https://www.fiercebiotech.com/medical-devices/safeway-severs-ties-theranos-as-350m-deal-collapses

Khan, R. (2020, December 15). Theranos' $9 billion evaporated: Stanford expert whose questions ignited the unicorn's trouble. *Forbes*. https://web.archive.org/web/20210208200436/https://www.forbes.com/sites/roomykhan/2017/02/17/theranos-9-billion-evaporatedstanford-expert-whose-questions-ignited-the-unicorn-trouble/

Diamandis, E. P. (2015). Theranos phenomenon: Promises and fallacies. *Clinical Chemistry and Laboratory Medicine, 53*(7). https://doi.org/10.1515/cclm-2015-0356

Wayback Machine. (n.d.). *FDA*. https://web.archive.org/web/20210208200409/https://www.fda.gov/media/94721/download

Alltucker, K. (2015, November 30). Arizona inspectors find Theranos lab issues. *The Arizona Republic*. https://web.archive.org/web/20210208200428/https://www.azcentral.com/story/money/business/consumers/2015/11/27/arizona-inspectors-find-theranos-lab-issues/76021416/

Abelson, R., & Pollack, A. (2016, April 13). Theranos under fire as U.S. threatens crippling sanctions. *The New York Times*. https://web.archive.org/web/20210208200758/https://www.nytimes.com/2016/04/14/business/theranos-elizabeth-holmes-proposed-ban.html

O'Brien, S. A. (2018, June 15). Elizabeth Holmes indicted on wire fraud charges, steps down from Theranos. *CNNMoney*. https://web.archive.org/web/20210208200304/https://money.cnn.com/2018/06/15/technology/elizabeth-holmes-indicted-theranos/index.html

NBC News. (2022, November 19). Theranos founder Elizabeth Holmes sentenced to prison for fraud [Video]. *NBC News*. https://www.nbcnews.com/business/business-news/elizabeth-holmes-sentenced-theranos-trial-rcna57344

Sunny Balwani, Elizabeth Holmes' right-hand man at Theranos, has been sentenced to nearly 13 years in prison. (n.d.). *SmartNews*. https://web.archive.org/web/20221208001059/https://www.smartnews.com/p/4491416338882897887

Das, R. K., & Drolet, B. C. (2022). Lessons from Theranos: Restructuring biomedical innovation. *Journal of Medical Systems, 46*(5). https://doi.org/10.1007/s10916-022-01813-3

Microsampling, N. (2024, March 5). Holmes trial: Lessons from failed blood sampling startup Theranos. *Neoteryx*. https://www.neoteryx.com/microsampling-blog/three-lessons-from-the-fall-of-theranos

Clayton, J. (2022, January 4). Elizabeth Holmes: Has the Theranos scandal changed Silicon Valley? *BBC News*. https://www.bbc.co.uk/news/technology-58469882

Altruism and Academic Amnesia: Psychological Pathways to Address Burnout in STEM Students

Dwaraka Bolla

Burnout among STEM students is a growing concern, often attributed to academic pressure, mental exhaustion, and cognitive overload. This paper explores two contributing psychological factors - altruism and stress-induced memory failure (academic amnesia)- and examines how to create a pathway that is the opposite of student burnout in STEM fields. Gathering from existing literature on medical knowledge, psychological studies of empathy and stress. This work encourages institutions to rethink their ethical structure and academic expectations for STEM scholars, an environment's atmosphere to better accommodate students' well-being.

Keywords: *STEM education; student burnout; altruism; academic amnesia; stress and memory; student well-being*

1 Introduction

Students today face intense academic pressure, balancing high expectations while maintaining a high reputation in challengeable majors all across the world specializing in STEM. Traits like altruism, self-discipline, and cognitive endurance are seen admirably specifically in fields such as medicine, research, and physics. But when students get to a point of stress and mental exhaustion, their well-being is compromised in order to keep up with educational standards established by institutions. This article acknowledges an increasing concern: the role of Altruism and Amnesia's effects in STEM students burnout. It highlights how voluntary and involuntary actions can break down mental health and academic integrity. For a student audience,would a better institutional response be necessary in the moral and psychological costs of burnout that students face?

2 Literature Review

2.1 Pathological Altruism

Altruism motivates students to help their peers, participate in group work, and engage in community service. However when altruism is driven by guilt, or doubting of oneself, (Oakley et al. 2012) refers to it as “pathological altruism” which has shown to result in emotional exhaustion. Maslow(1943/1996) notes that however, healthy selfishness - a healthy respect for one’s own growth, happiness and joy… has a positive impact on a person and their environment. The difference was stated by Kaufman and Jauk (2020) showing the students who prioritize themselves over others at an unstable pace have higher burnout symptoms including cynicism and emotional fatigue. In the same context, a recent meta-analysis on empathy and burnout found that affective empathy - absorbing people’s emotions leads to burnout. While cognitive empathy offers protective effects. Thus unchecked, altruism may paradoxically impair student mental health.

2.2 Academic Amnesia and Stress:

Academic Amnesia describes the incident where students experience “blanking out” under pressure during exams or presentations not being able to properly remember the necessary information needed at that time. This is supported by Vogel and Schwabe 2016 stating that exams or stressful situations in institutions hinder cognitive flexibility and memory formation. A meta analysis further adds to this point, showing acute stress can significantly impair memory retrieval rather than encoding or consolidation (Shields et al. 2017). Moving on a study of participants taking different tests exposed to exam-like conditions were seen to be facing Psychological stress. Results showed memory retrieval was significantly impaired after the stress condition specifically to emotional or high-stakes content (Gagon & Wagner, 2019). Stress interrupts context-dependent environments such as tests or exams. Meaning that even if an environment that was used to help with recall faces stress, it’s very unlikely for it to be of help (Smith & Gluck, 2009). These findings suggest that students under chronic stress may “blank out” due to reasons for instance, an neurological overload - a phenomenon where a person’s nervous system is overwhelmed by stimuli or demands leading to different physical and psychological symptoms essentially eroding academic confidence and performance. Directing strategies to reduce the symptoms of Academic Amnesia requires different cognitive techniques such as Retrieval Practice but also interventions from their workplaces, institutions etc…

2.3 Appealing Effects on Academic Burnout:

Academic Amnesia and Pathological Altruism may appear unrelated, the effects of both are deeply interweaved with one another forming a loop amplifying academic burnout. Students have a tendency to overwhelm themselves for others - whether it being guilt, imposter syndrome, or pressure to “I’ll do it all” causing their own emotional health to disperse. At the same time, stress-induced memory lapses are prone to impair students cognitive ability, triggering anxiety, PTSD… The different strains of emotional capability in helping others, and cognitive ability leads to a cycle of burnout. (Zsuzsa Győrffy) states that emotional fatigue and academic inability often concur in environments consisting of students or high-stakes surroundings. A student might have difficulty in remembering a concept during or before a test, then possibly spiraling into self-blame, leading them to overcompensate by forcing themselves to do more work, pushing their limits too far, to suddenly and repeating this loop.

3 Discussion

This paper shows that pathological altruism and academic amnesia are important factors that can result in a student being burned out, especially in challenging fields like STEM. When students tend to agree to help others without thinking of their own well-being, they eventually get emotionally exhausted. This is considered pathological altruism and it sucks a student's energy out quickly. Academic amnesia is considered similar except, in times where memorization is required such as in tests, or presentations… the mind goes blank due to overload of information. This causes self-esteem and confidence to go down tremendously leading to the birth of PTSD or anxiety.

Seeing that these two problems go together, the effect of a burnout would amplify. For example, a student not understanding a difficult concept could feel frustrated and try to overwork themselves to make up for it, which just increases the factors leading to burn out like stress, and exhaustion.

This raises an ethical question: What helps students gain the capability to increase their abilities and how can institutions help? To help students increase their capability in increasing their cognitive abilities while not opening the path to emotional or mental exhaustion,schools and teachers could adapt to alternatives to help ease students' minds. Schools, colleges, universities and other institutions should accommodate the pressure that they put on students to ensure students don’t have a harmful mindset. More specific alternatives such as initiating mental health programs or relief days in different institutions or workplaces for those pursuing education in STEM. Students in these fields have shown lack of sleep, fear of failure, abstract concepts etcetera. Institutions providing STEM education should implement various different relief-programs to help decrease stress induced by different fears, and anxiety for students. Adding to that, developing new methods to aim for healthy students in STEM should be vital for success not only in students but generally for the future of students and workers in STEM.

By understanding how emotional exhaustion and memory relevance work together, educators can alter support systems to fit student needs. Helping students stay happier and healthier is vital for students to excel in their studies.

4 Conclusion

Burnout among STEM students is a complex issue influenced by many factors. Some influencing factors are pathological altruism and academic amnesia. Students overextending their emotional energy to assist others without setting boundaries for themselves, risk emotional exhaustion. Simultaneously, stress-induced memory failure determines academic confidence and performance. These two factors alone create a harmful cycle intensifying burnout amidst STEM students. Addressing this challenge requires a comprehensive approach supporting students emotionally and cognitively. By promoting healthy self-care, stress-management lessons, and adjusting institutional standards, educators can help students work with balance and excel in academics. Future research should focus on emotional and cognitive aspects of student burnout by focusing on specific strategies to implement for students' benefit.

References

Kaufman, S. B., & Jauk, E. (2020). Healthy selfishness and pathological altruism: Measuring two paradoxical forms of selfishness. *Frontiers in Psychology, 11*, 1006. https://doi.org/10.3389/fpsyg.2020.01006

Cairns, P., Isham, A. E., & Zachariae, R. (2024). The association between empathy and burnout in medical students: A systematic review and meta-analysis. *BMC Medical Education, 24*(1), 5625. https://doi.org/10.1186/s12909-024-05625-6

Al-Shargie, F., Taresh, S. M., & Al-Ezzi, A. (2024). Mental stress and cognitive deficits management. *Brain Sciences, 14*(4), 316. https://doi.org/10.3390/brainsci14040316

Jamieson, J. P., Crum, A. J., Goyer, J. P., Marotta, M. E., & Akinola, M. (2018). Optimizing stress responses with reappraisal and mindset interventions: An integrated model. *Anxiety, Stress, & Coping, 31*(3), 245–261. https://doi.org/10.1080/10615806.2018.1442615

APA PsycNet. (n.d.). Record 2006-03431-006. *American Psychological Association*. https://psycnet.apa.org/record/2006-03431-006

Győrffy, Z., Birkás, E., & Sándor, I. (2016). Career motivation and burnout among medical students in Hungary—Could altruism be a protection factor? *BMC Medical Education, 16*(1), 182. https://doi.org/10.1186/s12909-016-0690-5

Schwabe, L., & Wolf, O. T. (2014). Timing matters: Temporal dynamics of stress effects on memory retrieval. *Cognitive, Affective, & Behavioral Neuroscience, 14*(3), 1041–1048. https://doi.org/10.3758/s13415-014-0256-0

Schwabe, L., Böhringer, A., & Wolf, O. T. (2009). Stress disrupts context-dependent memory. *Learning & Memory, 16*(2), 110–113. https://doi.org/10.1101/lm.1257509

Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *Journal of Neuroscience, 25*(11), 2977–2982. https://doi.org/10.1523/jneurosci.5139-04.2005

Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of acute stress on episodic memory: A meta-analysis and integrative review. *Psychological Bulletin, 143*(6), 636–675. https://doi.org/10.1037/bul0000100

Vogel, S., & Schwabe, L. (2016). Learning and memory under stress: Implications for the classroom. *NPJ Science of Learning, 1*(1), 16011. https://doi.org/10.1038/npjscilearn.2016.11

Oakley, B., Knafo, A., & McGrath, M. (2011). Pathological altruism—An introduction. In B. Oakley, A. Knafo, G. Madhavan, & D. S. Wilson (Eds.), *Pathological altruism* (pp. 1–4). Oxford University Press. https://barbaraoakley.com/wp-content/uploads/2016/12/000Chapter-1-Pathological-Altruism-Oakley-Knafo-McGrath.pdf

STEM ETHICS, POLICY & SOCIETY

Exploring the Long-Term Impact of Short-Form Content on Neurodegenerative Disease Development

Ivy Datta

Disability and mortality from neurodegenerative diseases such as Parkinson's disease (PD) have risen significantly since 1990. PD involves dopamine dysregulation, oxidative stress, neuroinflammation, and Lewy body formation. Emerging concerns point to the effects of repetitive short-form video consumption on dopamine regulation, as platforms like TikTok and Reels promote frequent reward cycles that may drive receptor desensitization and chronic dopamine fluctuations. This study examines whether such dysregulation contributes to PD-related mechanisms, including dopaminergic neuron loss in the substantia nigra, oxidative stress, neuroinflammation, and protein aggregation. Using rodent models, Enzyme-Linked Immunosorbent Assay (ELISA) and immunohistochemistry (IHC) will quantify and localize biomarkers after exposure. Findings aim to clarify the potential link between digital media use and neurodegeneration, offering insights for prevention and treatment strategies.

Keywords: *Parkinson's disease; dopamine dysregulation; short-form content; neurodegeneration; oxidative stress; neuroinflammation*

Since 1990, disability and death from neurodegenerative diseases, such as Parkinson's Disease (PD), have increased by 18% (World Health Organization, 2024). PD is characterized by dopamine dysregulation, oxidative stress, neuroinflammation, and Lewy bodies (deposits of alpha-synuclein) [2, 3]. New behavioral factors, such as the consumption of short-form video content on social media platforms like TikTok prompt concerns about their effect on neurodegeneration. These platforms encourage users to engage in repetitive short-form video consumption, triggering frequent DA releases. Dopamine (DA) is a neurotransmitter that regulates reward-related behavior through the mesolimbic DArgic pathway (Baik, 2020). Over time, constant DA stimulation leads to receptor desensitization, requiring more stimulation to achieve the same rewarding effect, resulting in DA 'crashes' (Nimitvilai et al., 2014). Chronic DA fluctuations may contribute to PD development.

Despite the growing popularity of short-form content such as TikTok, Reels, research on its long-term effects is scarce. This study investigates how excessive DA stimulation from repetitive reward cycles influences DAergic neuron loss in the substantia nigra, neuroinflammation, oxidative stress, and protein aggregation in a rodent model. Using Enzyme-Linked Immunosorbent Assay (ELISA) and immunohistochemistry (IHC) on rats, biomarkers will be quantified and localized in brain tissue after exposure. The findings aim to provide novel insights into the link between DA dysregulation and neurodegenerative diseases like PD, providing new evidence to inform prevention and treatment strategies.

1 Introduction

Neurodegenerative diseases such as PD are characterized by loss of DAergic neurons, oxidative stress, neuroinflammation, and protein aggregation [2, 3]. A hallmark of PD is the degeneration of DA-producing neurons in the substantia nigra, resulting in motor and cognitive symptoms due to insufficient DA (National Institute of Neurological Disorders and Stroke [NINDS], 2022). While genetics can contribute to PD, environmental and behavioral factors may accelerate its progression.

The proposed study will explore a behavioral factor's contribution to PD, specifically short-form social media content consumption.

Short-form social media content poses several unique risks to brain health. First, short-form content disrupts and exploits the brain's DAergic reward system by delivering quick and repetitive stimulation to the mesolimbic pathway, which governs reward processing [5, 6]. This content overstimulates the ventral tegmental area (VTA) and nucleus accumbens (NAc), leading to excessive DA release (Fernandez, 2022). Over time, chronic overstimulation desensitizes DA receptors, particularly D1 and D2 receptors, and disrupts normal signaling, resulting in DA "crashes" and chronic fluctuations in DA levels (NeuroLaunch Editorial Team, 2024a). This dysregulation is hypothesized to disrupt normal functioning in the mesocortical pathway (critical for decision-making and executive functions), and potentially contribute to neurodegeneration (NeuroLaunch Editorial Team, 2024b). Although this behavior has been linked to cognitive effects like reduced attention spans, its long-term impact on DAergic neurons remains unclear. Constant stimulation of reward pathways such as the mesolimbic pathway poses a research question- Does excessive dopamine-induced stimulation accelerate the death of DAergic neurons?

Second, DAergic system dysregulation and defects are linked to neuroinflammation. Excess DA, when metabolized, generates reactive oxygen species (ROS), which can activate microglia, the brain's immune cells, and initiate inflammation (Meiser et al., 2013).

Third, DA metabolism generates ROS that damage cellular components, including lipids, proteins, and DNA. A study found that DA exposure reduced neuronal cell viability and increased ROS production, resulting in cellular stress [10, 11]. This finding suggests oxidative stress as a harmful consequence of DA metabolism [10, 11]. Oxidative stress can also lead to mitochondrial dysfunction, which is found in PD (Matura et al., 2015). Furthermore, cellular stress caused by ROS may result in protein misfolding, a hallmark of neurodegenerative diseases (Matura et al., 2015). These effects are particularly detrimental in the nigrostriatal pathway, which controls motor function (Good et al., 2011).

Using ELISA and IHC on rat models, biomarkers of neuroinflammation and oxidative stress such as Tumor Necrosis Factor-alpha (TNF-α), Interleukin-6 (IL-6), Superoxide Dismutase (SOD1), and alpha-synuclein will be quantified [14, 15]. Also, protein aggregates and DAergic neurons in the substantia nigra will be quantified.

This study offers a novel approach by exploring the following research question- Does excessive dopamine-induced stimulation cause DAergic neuron loss, neuroinflammation, oxidative stress, and protein aggregation, ultimately leading to neurodegenerative diseases like PD? By exploring the impact of DA dysregulation on DAergic neuron loss, neuroinflammation, oxidative stress, and protein aggregation, the research could provide new insights into the mechanisms behind neurodegenerative diseases. The innovative aspect of linking modern digital behaviors with PD development could open doors to targeted prevention and treatment strategies that address the effects of excessive DA stimulation. This would bridge the gaps in current research, offering a new perspective on neurodegenerative disease prevention in the digital age.

2 Hypothesis

Excessive DA release induced by short-form social media use combined contributes to DAergic neuron loss, neuroinflammation, oxidative stress, protein aggregation and neurodegeneration in rat models. Rats exposed to DA-inducing stimuli will have less DAergic neurons in the substantia nigra, increased neuroinflammation markers, increased oxidative stress markers, and more protein aggregation.

3 Methods

The study will use animal models of adult male Wistar rats to simulate the neuropathological effects of DA dysregulation. There will be two forms of DA inducing stimuli, and the study will utilize 3 distinct groups of rats.

- Group 1 will be a control group and will not have any exposure to unnatural DA-inducing stimuli.
- Group 2 will be exposed to DA inducing stimuli using a lever and food pellets.
- Group 3 will be exposed to sensory stimuli from dynamic screens and auditory cues, simulating short-form content consumption.

Group 2 rats will be exposed to rapid DA reward cycles. The stimuli will be generated using chambers equipped with a lever and a feeder. Following a variable-ratio reward system (VRRS), rats will sometimes receive a small food pellet when they press the lever (Weinschenk, 2013). The reward and its intensity will be randomized, following VRRS (Weinschenk, 2013). The number of lever presses required to earn a treat, and the size of the treat, will change randomly and unpredictably. These lever pressing sessions will last 2 hours a day, for a 3 week period.

Group 3 rats will be exposed to the sensory effects of watching short-form social media content. To simulate sensory stimuli and induce rapid DA reward cycles, we will create touch-screen based chambers for the rats. The screen will display colorful, dynamic, and continuous visual stimuli. These stimuli will include animations, like spirals and moving shapes; bright colors; and abstract patterns. Visual stimuli will also be paired with auditory stimuli, like chimes and dings. The rats will be exposed to the visual and auditory stimuli for 2 hours a day, over a 3 week period.

Following the 3 week period, the rats will be euthanized through CO_2 asphyxiation (NIH Office of Animal Care and Use, 2024). Brain tissue from all rats will be extracted, specifically the prefrontal cortex, hippocampus, basal ganglia, and substantia nigra.

4 Data Collection

Biomarkers related to DAergic neuron loss, neuroinflammation, oxidative stress, and protein aggregation will be quantified from the brain tissue.

DAergic neurons: To assess the impact of excessive DA stimulation on DAergic neurons, the substantia nigra will be analyzed for neuronal loss. Brain tissue will be homogenized in ice-cold PBS with protease inhibitors, and then centrifuged to isolate the supernatant for further analysis (Sundar et al., 2010). IHC will be used to visualize DAergic neurons by staining for tyrosine hydroxylase (TH), a marker for DAergic neurons (National Institute of Environmental Health Sciences [NIEHS], n.d.). Brain sections will be incubated with anti-TH antibodies and visualized using DAB staining [18, 19]. The number of TH-positive neurons in the substantia nigra will be quantified to determine the extent of DA-induced neurodegeneration. Statistical analysis will compare the number of DAergic neurons between experimental groups to assess potential neuronal loss due to DA dysregulation.

Neuroinflammation: Standardly available ELISA kits will be used to quantify levels of pro-inflammatory cytokines like Tumor Necrosis- alpha (TNF-α) and Interleukin-6 (IL-6) (Matura et al., 2015). Brain regions (prefrontal cortex, hippocampus, basal ganglia) will be homogenized in ice-cold PBS with protease inhibitors, centrifuged at 10,000 × g for 10 minutes at 4°C, and the supernatant will be analyzed (Sundar et al., 2010). For localization, immunohistochemistry will be used to examine the distribution of SOD1 and MDA in paraffin-embedded brain sections using primary antibodies specific to each marker and DAB staining.

Oxidative Stress Markers: ELISA will also be used to measure oxidative stress markers, including Superoxide Dismutase 1 (SOD1) and Malondialdehyde (MDA) (Cetinkaya et al., 2005). SOD1 is an oxidative enzyme that protects cells from ROS and oxidative damage (Hwang et al., 2020). Brain regions (prefrontal cortex, hippocampus, basal ganglia) will be homogenized in ice-cold PBS with protease inhibitors, centrifuged at 10,000 × g for 10 minutes at 4°C, and the supernatant will be analyzed (Sundar et al., 2010). To understand the localization, immunohistochemistry will be used to examine the distribution of SOD1 and MDA in paraffin-embedded brain sections using primary antibodies specific to each marker and DAB staining.

Protein Aggregation: Protein aggregation in the brain tissues will be assessed using ELISA kits, quantifying amyloid-beta (Aβ) and alpha-synuclein levels in the prefrontal cortex, hippocampus, and basal ganglia (Wilson et al., 2023). The brain tissue will be homogenized in ice-cold PBS containing protease inhibitors, centrifuged at 10,000 × g for 10 minutes at 4°C, and the supernatant will be analyzed (Sundar et al., 2010). To assess localization, IHC will be performed on paraffin-embedded brain sections. Primary antibodies specific to amyloid-beta and alpha-synuclein will be used, with DAB staining for visualization of protein aggregates. Aggregation will be confirmed by identifying distinct deposits in the brain regions under a microscope (Wilson et al., 2023).

5 Hypothesized Results

We hypothesize that chronic exposure to dopamine-inducing stimuli, mimicking the overstimulation from short-form social media content, will lead to the following outcomes in rat models:

- A reduction in DAergic neurons in the substantia nigra.
- Increased markers of neuroinflammation (e.g., TNF-α, IL-1β).
- Elevated oxidative stress markers (e.g., MDA, SOD1)
- More protein aggregation, (α-synuclein) and beta amyloid)

Previous studies found that rat models exposed to dopamine-induced stress showed α-synuclein aggregation in dopaminergic neurons, a hallmark feature of Parkinson's disease (Possemato et al., 2023). This aggregation correlates with neurodegeneration. Another study found that DA exposure reduced neuronal cell viability and increased ROS production, resulting in cellular stress [10, 11].

6 Discussion and Conclusion

The quantified DAergic neurons in the substantia nigra, neuroinflammation levels, oxidative stress markers, and protein aggregation in the brain tissues will provide novel insight into the effects of DA dysregulation. The expected findings reiterate the hypothesis that DA dysregulation from short-form social media content consumption accelerates the pathogenesis of neurodegenerative diseases like PD. The findings link back to the research question by indicating that DA dysregulation contributes to neurodegeneration. The findings would enhance understanding of how digital media affects brain health and lead to treatments for DAergic neuron loss, neuroinflammation, oxidative stress, and protein aggregation from DA dysregulation.

Although this study is a valuable asset for new research regarding neurodegeneration, some limits should be addressed. First, the use of rat models may not fully simulate the complexities of the human brain. Also, the range of biomarkers and neurodegenerative markers quantified and evaluated in the study is relatively narrow. In future studies, a broader examination of molecular pathways such as neurotrophic factors and neuroplasticity should be examined. Future research may also use non-human primate animal models, to find outcomes more similar to humans. These studies should be longer-term, exploring what happens to the brain over time. This study will assess the specific neuropathologic changes in the brain when exposed to unnaturally high DA- inducing stimuli. These conditions are increasingly common in today's society, and this study hopes to provide novel insights to how digital factors impact brain health and pave the way to future research and treatments to mitigate the negative effects of short-form content.

References

World Health Organization Media Team. (2024, March 14). *Over 1 in 3 people affected by neurological conditions, the leading cause of illness and disability worldwide*. WHO. https://www.who.int/news/item/14-03-2024-over-1-in-3-people-affected-by-neurological-conditions--the-leading-cause-of-illness-and-disability-worldwide

National Institute of Neurological Disorders and Stroke. (2020, updated 2022). *Parkinson's disease*. NINDS. https://www.ninds.nih.gov/health-information/disorders/parkinsons-disease

Wilson, D. M., III, Cookson, M. R., Van Den Bosch, L., Zetterberg, H., Holtzman, D. M., & Dewachter, I. (2023). Hallmarks of neurodegenerative diseases. *Cell, 186*(4), 693–707. https://doi.org/10.1016/j.cell.2023.01.012

Baik, J. H. (2020). Stress and the dopaminergic reward system. *Experimental & Molecular Medicine, 52*(12), 1879–1890. https://doi.org/10.1038/s12276-020-00532-4

Nimitvilai, S., Herman, M., You, C., Arora, D. S., McElvain, M. A., Roberto, M., & Brodie, M. S. (2014). Dopamine D2 receptor desensitization by dopamine or corticotropin releasing factor in ventral tegmental area neurons is associated with increased glutamate release. *Neuropharmacology, 82*, 28–40. https://doi.org/10.1016/j.neuropharm.2014.03.008

Malik, I. (2023, November 15). TikTok's dopamine trap: The addictive power of short-form videos. *Our Mental Health*. https://www.ourmental.health/screen-time-sanity/tiktoks-dopamine-trap-the-addictive-power-of-short-form-videos

Fernandez, V. (2022, August 20). Social media, dopamine, and stress: Converging pathways. *Dartmouth Undergraduate Journal of Science*. https://sites.dartmouth.edu/dujs/2022/08/20/social-media-dopamine-and-stress-converging-pathways

NeuroLaunch Editorial Team. (2024, August 22). Understanding social media's dopamine addiction. *NeuroLaunch*. https://neurolaunch.com/social-media-dopamine

NeuroLaunch Editorial Team. (2024, August 22). Mesocortical dopamine pathway and mental health. *NeuroLaunch*. https://neurolaunch.com/mesocortical-dopamine-pathway

Meiser, J., Weindl, D., & Hiller, K. (2013). Complexity of dopamine metabolism. *Cell Communication and Signaling, 11*(34), 34. https://doi.org/10.1186/1478-811X-11-34

Luz, M. H., Baierle, M., de Oliveira, J., et al. (2015). DA induces the accumulation of insoluble prion protein and affects autophagic flux. *Frontiers in Cellular Neuroscience, 9*, 12. https://doi.org/10.3389/fncel.2015.00012

Matura, L. A., Ventetuolo, C. E., Palevsky, H. I., et al. (2015). Interleukin-6 and tumor necrosis factor-α are associated with quality of life–related symptoms in pulmonary arterial hypertension. *Annals of the American Thoracic Society, 12*(3), 370–375. https://doi.org/10.1513/AnnalsATS.201406-277OC

Good, C. H., Hoffman, A. F., Hoffer, B. J., et al. (2011). Impaired nigrostriatal function precedes behavioral deficits in a genetic mitochondrial model of Parkinson's disease. *FASEB Journal, 23*(4), 1333–1344. https://doi.org/10.1096/fj.08-126789

Cetinkaya, A., Belge Kurutas, E., Buyukbese, M. A., Kantarceken, B., & Bulbuloglu, E. (2005). Levels of malondialdehyde and superoxide dismutase in subclinical hyperthyroidism. *Mediators of Inflammation, 2005*(1), 57–59. https://doi.org/10.1155/MI.2005.57

Weinschenk, S. (2013, November). Use unpredictable rewards to keep behavior going. *Psychology Today*. https://www.psychologytoday.com/us/blog/brain-wise/201311/use-unpredictable-rewards-to-keep-behavior-going

NIH Office of Animal Care and Use. (2001, updated 2024). *Guidelines for euthanasia of rodents using carbon dioxide*. NIH. https://oacu.oir.nih.gov/system/files/media/file/2024-01/b5_euthanasia_of_rodents_using_carbon_dioxide.pdf

Sundar, I. K., Caito, S., Yao, H., & Rahman, I. (2010). Oxidative stress, thiol redox signaling methods in epigenetics. *Methods in Enzymology, 474*, 213–244. https://doi.org/10.1016/S0076-6879(10)74013-3

National Institute of Environmental Health Sciences. (n.d.). *Tyrosine hydroxylase immunohistochemistry protocol.* NIH. https://www.niehs.nih.gov/sites/default/files/research/resources/protocols/protocols-immuno/immunohistochemistry/tyrosinehydroxylase_mr.html

Hwang, J., Jin, J., Jeon, S., Moon, S. H., Park, M. Y., Yum, D.-Y., Kim, J. H., Kang, J.-E., Park, M. H., Kim, E.-J., Pan, J.-G., Kwon, O., & Oh, G. T. (2020). SOD1 suppresses pro-inflammatory immune responses by protecting against oxidative stress in colitis. *Redox Biology, 37,* 101760. https://doi.org/10.1016/j.redox.2020.101760

Possemato, E., La Barbera, L., Nobili, A., et al. (2023). The role of dopamine in NLRP3 inflammasome inhibition: Implications for neurodegenerative diseases. *Neurochemistry International, 170*, 105741. https://doi.org/10.1016/j.neuint.2023.105741

Black Hole: Scientific Wonder vs Black Hole: Human Trash Guzzler

Simran Rishi

As commercialized space travel grows, concerns over its environmental impact have expanded, particularly regarding the accumulation of space debris. Space trash, originating from satellites, rockets, and failed missions, has doubled across sectors since 2010 and poses risks to navigation, satellites, and essential services such as GPS and weather forecasting. Without binding international regulations, cleanup remains limited, heightening the threat of the "Kessler Syndrome," in which collisions create cascading debris that could make future space travel nearly impossible. This article examines the ethical, environmental, and technological challenges of managing orbital debris and emphasizes the need for global responsibility and innovative solutions.

Keywords: *Space debris; commercial space travel; orbital environment; Kessler Syndrome; environmental ethics; satellite navigation*

In the age of commercialized space travel, it's easy to get looped up into the craze of discovering an entirely new place. However, not often do people realize the adverse effects of lower orbital commercialized space flight on our Earth's and space's environment. It's common to be unfamiliar with space, as research regarding space travel only began gaining traction less than 50 years ago - during the Space Race (The Aerospace Corporation, 2022). As the world began to realize there was something beyond the sky, interests in this field skyrocketed. Popular media such as Star Wars and 2001:A Space Odyssey introduced and inspired younger generations to reach beyond the stars. During this time, NASA launched Alan Shepherd and Neil Armstrong into the embrace of space, further pushing Americans to become involved in space.

However, as enticing as it seems to view the beautiful curvature of Mother Earth from 35,00 feet above, we can't forget that humans don't own space. It is a common saying heard in almost every national park; "leave everything the way you found it". In a way, we should be treating space similarly - as an environmental haven that shouldn't be harmed by the adverse effects of humans (and therefore space travel).

Unfortunately, this is exactly what we aren't doing enough of. When considering the ethical side of space travel, the conversation would be incomplete without a large discussion on space trash. Space trash, a term born with the launch of Sputnik 1 in 1957, describes any piece of machinery or debris that has been specifically left by humans in space which no longer serves a purpose (The Aerospace Corporation, 2022). We have asked ourselves “Is it truly ethical to leave trash in space? After all, who would be harmed?” We haven’t discovered any life forms outside of our atmosphere. And since it is that case, why don’t we just throw all of our trash into space and call it a day? Who should be responsible for the clean-up of space if no one “lives” there. Though there is an endless array of questions that could be asked to justify both sides, in this article we will focus on the effect that space debris has on space travel and the environment.

Let’s start with the numbers. According to the European Space Agency’s annual environmental report, space trash has doubled in every sector since 2010 (The European Space Agency, 2021). The sharp increase in commercial low-Earth orbits is especially concerning considering many companies tend to leave trash, prioritizing bringing the satellites back before the debris. However, even governmental agencies have no legally binding contract that ‘forces’ them to clean up after their astronauts and satellites in space (Quell, 2020).

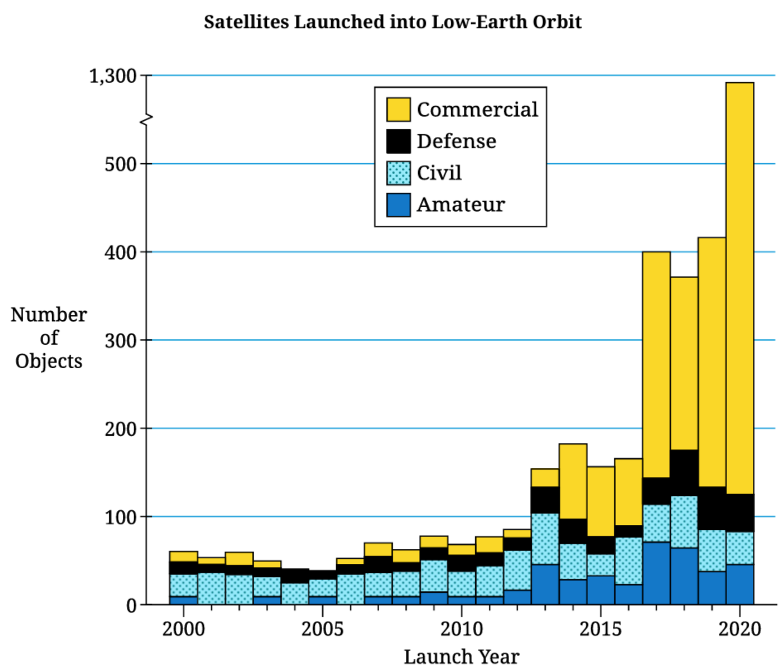


Figure 1: Number of satellites launched into low-Earth orbit by category (commercial, defense, civil, and amateur) from 2000 to 2021. Commercial launches have increased dramatically in the past decade, driving the sharp overall rise in orbital objects (Quell, 2020).

This now begs the question, does space trash actually have any effect on us? The easiest answer would be no. Since many people won’t actually end up traveling into space, the short term effects won’t be too much. However, as trash begins to accumulate, satellites will have increasingly difficult paths to navigate leaving many people who rely on services such as GPS and weather forecasting services without the data needed to carry on normal life activities. In 1978, prominent NASA scientists Don Kessler and Burton Cour-Palais co-published a paper introducing the 'Kessler Syndrome' (The Aerospace Corporation, 2022) (O'Callaghan & Warhol, 2021). They predicted that with the rate that space trash and debris was being left in orbit it would cause so many collisions that any space travel would be nearly impossible. Hence, if you look at the long term effects of leaving space trash or even disposing of normal human trash in space, you can clearly see the negative effects it would have on daily life.

However, this rapid attention to space trash also begs the question, "why aren't we able to create technologies that don't require such extensive cleanup?" If we have become so advanced as a world population, shouldn't we be able to program ways for the trash to 'untrashify' itself. Technically yes, but this would be even more costly and require more testing which is not only harmful towards the environment in space, leaving excess debris from failed attempts, but also towards our Earth's environment, leaving metal scraps on ocean shores and around cities bordering testing facilities.

From commercialized space travel to the lack of national regulations, it's truly no question that ethics are heavily involved when discussing space trash. However, there are so many more aspects involved, such as the economical and business side. Although it may seem that companies don't tend to care, some are taking an initiative to clean up space. Doing our own research and supporting start ups such as the ESA's Clean Space Initiative can help reduce the extreme effects of debris in space (Quell, 2020). In the end, it's how we chose to use this information that will set us apart; and hopefully above the rest.

References

The Aerospace Corporation. (2022, November 2). *A brief history of space debris*. The Aerospace Corporation. https://aerospace.org/article/brief-history-space-debris

European Space Agency. (2021, May 27). *ESA's space environment report 2021*. ESA. https://www.esa.int/Space_Safety/Space_Debris/ESA_s_Space_Environment_Report_2021

O'Callaghan, J., & Warhol, A. (2021). *What is space junk and why is it a problem?* Natural History Museum. https://www.nhm.ac.uk/discover/what-is-space-junk-and-why-is-it-a-problem.html

Quell, M. (2020, August 28). *Lack of space law complicates growing debris problem*. Courthouse News Service. https://www.courthousenews.com/lack-of-space-law-complicates-growing-debris-problem/

The Ethics of Using Neuroscience to Enhance Learning and Productivity

Eesha Harish

Neuroscience tools like brain-stimulating headsets, wave-tracking apps, and cognitive enhancers are no longer limited to medical use—they're being adopted by healthy students and professionals chasing productivity. While they can help treat ADHD, anxiety, or stroke recovery, using them for enhancement raises tough questions. Who gets access, and does it widen inequality? Will people feel pressured to use them just to keep up? What happens to identity and achievement when success feels engineered? With safety still uncertain, especially for developing brains, these technologies demand careful ethical reflection before we trade authenticity and well-being for efficiency.

Keywords: *Neuroethics; cognitive enhancement; productivity; fairness; identity; neuroscience safety*

In a world that never seems to slow down, it feels like everyone is trying to find ways to be more productive, more focused, and just simply better. With all the pressure to keep up, neuroscience is starting to look like the next big shortcut. People are turning to things like brain-stimulating devices, apps that track brainwaves, and even pills that claim to boost memory or attention. It's kind of wild to think about. These tools were originally created to help people with serious conditions, but now they're being used by healthy students and professionals who just want to do more. That's where the ethics get complicated.

Neuroscience has already helped so many people in powerful ways. It can improve focus in kids with ADHD, support stroke recovery, and even help treat anxiety and depression. Those are all important and meaningful uses. But now we're seeing people use the same technology to pull all-nighters, stay locked into study sessions, or hit insane productivity levels at work. I get why that sounds tempting, but it also brings up some big questions.

One of the first things that comes to mind is fairness. If brain-enhancing tech or medication becomes the norm, who actually gets access to it? Most likely, it'll be people who can afford it. That means students from wealthier schools or employees at top companies could end up with even more advantages, while others are left behind. We already have issues with inequality, and this would just make it worse.

Another concern is pressure. Even if no one says you have to use something, it might start to feel like you do. If everyone around you is using brain-boosting tools to study faster or work longer, it's hard not to feel like you need to join in just to keep up. That kind of pressure takes away real choice, especially for students or people in high-stress jobs.

And then there's the identity part. Our brains are literally what make us who we are. If we start changing how they function just to be more productive, does that change us too? If I ace a test or crush a project because of some pill or brain-stimulating headset, would it still feel like my achievement? There's something about working hard, failing, and growing that makes success feel real. Skipping that process might help you reach a goal faster, but it also takes away part of what makes that success meaningful.

I also can't ignore the safety side. A lot of these tools are still new, and we don't fully understand how they affect the brain long-term. That's especially risky for teens, since our brains are still developing. What if something that seems helpful now ends up causing problems later? We can't treat our brains like test subjects just because we're in a rush to be better.

Some people try to draw a line between using neuroscience for medical reasons versus using it for enhancement. Like, it's fine if someone has a condition and needs support, but it's questionable if someone already healthy uses the same thing to get ahead. I see the logic there, but honestly, the line is starting to get really blurry.

What we need is open, honest conversation. Scientists, teachers, parents, students, companies—everyone should be part of it. There should be clear guidelines about what's safe, what's fair, and what's actually necessary. No one should feel forced to change their brain just to fit into some version of success defined by constant productivity.

At the end of the day, being human is not about being perfect. It's about learning, growing, messing up, and figuring things out. The brain isn't a machine we need to upgrade every time we feel behind. It's personal. It's emotional. It's who we are. And while neuroscience has the power to help people in amazing ways, we have to be careful not to lose ourselves in the process.

References

Dresler, M., Sandberg, A., Bublitz, C., Ohla, K., Trenado, C., Mroczko-Wąsowicz, A., Kühn, S., & Repantis, D. (2019). Hacking the brain: Dimensions of cognitive enhancement. *ACS Chemical Neuroscience, 10*(3), 1137–1144. https://doi.org/10.1021/acschemneuro.8b00571

Illes, J., & Bird, S. J. (2006). Neuroethics: A modern context for ethics in neuroscience. *Trends in Neurosciences, 29*(9), 511–517. https://doi.org/10.1016/j.tins.2006.07.002

Marazziti, D., Avella, M. T., Ivaldi, T., Palermo, S., Massa, L., Di Vecchia, A., Basile, L., & Mucci, F. (2021). Neuroenhancement: State of the art and future perspectives. *Clinical Neuropsychiatry, 18*(3), 181–186. https://doi.org/10.36131/cnfioritieditore20210303

Contemporary Theories of Consciousness and Their Applications to Technology

Erika Liu

This paper concerns itself with exploring the synthesis of various topics discussed in length at the Active Inference Institute, including Friston's Free Energy Principle, Rudrauf's Projective Consciousness Model and Worden's Projective Wave Model. Through a deeper examination of the correlations between these closely-linked theories we may develop a stronger understanding of the nature of human consciousness, resolving problems in modern logic regarding the topic (eg., dispelling claims of AI consciousness) and holding potential for practical applications in fields such as AI and fMRI.

Keywords: *Consciousness; Free Energy Principle; Projective Consciousness Model; Projective Wave Model; active inference; artificial intelligence*

1 Synthesis of Contemporary Theories of Consciousness

Despite our current wealth in data regarding more scientifically established fields such as neuroscience, theories attempting to explicate the most fundamental aspects of human nature–eg., consciousness–remain comparatively few and far between. This is for good reason, too: unlike other concepts relating to the human body, consciousness is uniquely subjective and metaphysical, rendering its study nebulous and exceedingly difficult to coordinate formally. That is because our experience of consciousness is exclusively phenomenal–which is to say, that all of our experiences are inherently subjective, and therefore subject to the various illusions and distortions that our information-processing systems will inevitably facilitate. This fundamental divorce between what we perceive and what may be a more objective reality, compounded with our current lack of formal understanding as to the basic mechanisms of consciousness on a fundamental level, is what makes study of this topic so challenging.

Philosophy of the mind is a unique branch of metaphysics that attempts to provide satisfying answers to these problems. It is largely dedicated to analysis of consciousness and its concomitant phenomena, such as thought and perception, which modern researchers strive to explicate in detail through a network of fascinating theories. Some of the key areas of focus within such a field include: how does consciousness originate? Under what system of principles does consciousness function? How are our physical constituents linked to the abstract component of consciousness, and how do such physical constituents account for the precision with which we perceive 3D space in our everyday life? The purpose of this paper is to provide an informative, largely non-technical overview of several linked claims within philosophy of the mind. When taken into account as a group, these claims overcome a number of contemporary issues within the study of consciousness, lending ample merit to their plausibility.

The most widely recognized of these theories is the Free Energy Principle (FEP,) first developed by British researcher Karl Friston and others in the early 2000s. The FEP theory attempts to provide an elegant and cohesive answer to the question of *what it means to be alive*. It achieves this through a simple mathematical principle integrating the fundamentals of pre-established concepts such as Bayesian inference and predictive coding into a single theory. According to the free energy theorem, all organisms, or agents, are dynamical entities separated from our actual environment by an interface termed the “markov blanket.” Critically, this “markov blanket” is also responsible for allowing for the autonomous separation of the agent from its surroundings. It is also responsible for propagating errors in organisms’ guesses and interactions with their environment, from which agents are at least partially divorced from. The imperative for biological systems, then, is to act as best they can to minimize prediction errors within their internal representations of surrounding environments. These errors are quantified by a metric known as “free energy,” which, by its more intuitive definition, measures the amount of surprise or uncertainty experienced by an organism. Free energy may also be defined as prediction error (PE,) or expected cost; were one to integrate the second law of thermodynamics into such a topic as well, free energy may also be likened to entropy, or a measure of the total disorder within a system. The Free Energy Principle therefore suggests that living systems, such as the human brain, strive to minimize free energy, a metric of uncertainty and disorder, in order to better thrive and adapt within complex and dynamical environments today (Friston, 2010).

There are two outlets through which biological systems may suppress free energy–either via alteration of actual sensory input via physical action upon the real world, or alteration of another value known as recognition density, an *approximate* probability distribution of the causes of data (eg., sensory input.) Reduction of the latter is achieved through modification of internal states–a process of selective perception conducted with the primary goal of minimizing free energy. This process is what Friston calls “active inference.” It is this active inference, integrated with the principles of Bayesian inference (a method of information processing wherein actions are guided by predictions, and continuous updates in sensory feedback drives revisions in said predictions, a method of probabilistic calculation derived from Bayes' Theorem,) which coordinates the behavior of all living agents, presumably.

Mathematically, the concept of free energy has also obtained a relatively rigorous definition; Friston describes the quantity as evaluated in three separate formulations. The first expresses free energy as the difference obtained via energy minus entropy. Intuitively, what this expression describes is a system's energy (the joint energy expended in all real-world interactions) minus its entropy (which, within this context, would be the recognition density.) This equation is convenient for living organisms, as its calculations involve values that are already available to our information-processing systems. The second mathematical representation of free energy expresses it as energy plus perceptual divergence. This divergence term is essentially the difference between an organism's recognition density (approximation of the probability distribution of causes, the product of inferences and generative models) and its conditional density (the actual probability distribution of causes or model parameters, when given data.) If we were to denote recognition density as q(9|u) and conditional density as t(9|u), and denote energy as the negative log probability of outcomes, we would get the equation "Free energy = -log p(s) + q(9|u) - t(9|u)." This signifies that, by changing values of a living system's recognition density, we may reduce the total perceptual divergence and, in turn, lower the total value of free energy. The third formulation describes free energy as complexity minus accuracy. Complexity, in this context, refers to Bayesian surprise, a measure of how much new data alters our preexisting beliefs. It is, in essence, the difference between the prior density, P(A)–eg., beliefs about the state of the world before other sensory inputs are taken into account–and posterior beliefs, which take into account these sensory inputs. Accuracy is then defined as surprise regarding sensory inputs expected within the recognition density. This formula indicates that free energy is also minimizable via changes in sensory perception. By selectively sampling certain aspects of our phenomenal reality, we can better reconcile the differences between our predictions and our actual environment, conforming reality to our expectations. An intuitive analogy for such a phenomenon may be provided by a person feeling around in the dark. They build assumptions on their perception before confirming their own predictions through physical contact with a certain object. Each mathematical formulation, then, offers some intuitive insight into the precise nature of free energy, how it functions, and how agents may concretely minimize it, strengthening our extant understanding.

While not universally accepted amongst the scientific community–Friston's theory suffers from a lack of falsifiability, wherein it cannot be proven nor disproven via empirical evidence–FEP regardless provides a very intuitive and elegant explanation behind many of the intricacies of higher consciousness. By the principles of Occam's Razor, it is extraordinarily appealing to those who have a bias for simplicity, and its capacity to offer intuitive and satisfying explanations for why we behave the way we do renders it a uniquely compelling proposal.

Friston's ideas have since influenced a host of other, closely-linked theories over recent years, including London-based researcher Robert Worden's Projective Wave Theory, published in 2024. The Projective Wave Theory is, in itself, an extension upon another major neural theory regarding consciousness–said theory being the Projective Consciousness Model (developed by David Rudrauf and others.) The Projective Consciousness Model (or PCM) narrows in on the spatial recognition aspect of our phenomenal experience. It suggests that agents utilize active inference when constructing 3D projective models of our environment, focusing on optimization of certain representations with the imperative of free energy minimization (Rudrauf, 2017). Worden takes these concepts and brings them a step further.

There are, currently, a host of significant issues persisting amongst contemporary models of spatial consciousness. According to Worden, these include: (1) a selection problem, as in, under what principle (or set of principles) does the brain select relevant neurons with which to represent 3D space? (2) a precision problem, which questions how neurons, with their stochastic, relatively delayed firing rates, may account for the vast amount of detail with which we perceive our environment and ambient qualia, and (3) a decoding problem, concerning how distorted spatial representations in the brain are "decoded" to produce our largely undistorted and reliable internal model of space (Worden, 2024). These principles pose significant barriers to current attempts to further our comprehension of the nature of spatial consciousness, and need to be addressed.

Worden proposes that these issues may find their origins in a flawed model of our brains as being coordinated solely by the firing rates of neurons, which are insufficient to account for the sophistication of our internal 3D models of space. By contrast, as an alternative to storing 3D positions as firing rates, Worden suggests that our neurons may be capable of "coupling" to a hypothesized wave excitation in the brain, which, in turn, transmits data regarding our environment through a range of wave vectors, resembling the process of holography. This resolves all three of the aforementioned issues with current neural theories of consciousness. (1) For the selection problem, there would be no need to select for individual neurons, as information is transmitted within one collective wave. (2) The superiority of waves as data vessels resolves any issues with precision; with fast response times close to the milliseconds and the capacity to store information regarding a great variety of qualia, waves exceed neurons in nearly every respect with regards to the relaying of data. (3) Information can be decoded relatively easily from a wave via a Fourier transform, a process involving the separation of wave interference sums into their basic, sinusoidal constituents. The logically satisfactory nature of this theory grants it a very positive Bayesian balance (a great amount of credibility,) supporting its tenability and giving it grounds upon which to merit further scientific exploration, as it is also, conveniently, easily falsifiable.

The Projective Wave Theory is currently in a heavily nascent stage; many key questions, especially regarding the physical properties of such a wave, have yet to be answered in detail. Were deeper scientific investigation to be conducted into such a hypothesis, these lingering questions may be very quickly answered; currently, however, knowledge of the physical properties of Worden's wave exist only in postulations and educated guesses.

It is generally agreed upon that the wave cannot be electromagnetic, for a host of reasons–primarily that such a wave would have to compete with the ambient electromagnetic brainwaves generated by working neurons. This would necessitate the projective wave to function only at a very high intensity, wherein it may be capable of commanding sufficient attention when signalling to neurons. This high intensity would result in wanton squandering of our bodies' energy and resources. In order to reduce such supererogatory processes, the brain would most likely opt for low-intensity waves of an alternative nature, to promote efficiency within our metabolism.

One of the most compelling hypotheses for alternate forms of this wave posits that such a wave may in fact be a coherent quantum excitation. To give a rough idea of this relatively abstract concept, coherency, at least in the quantum world, can be defined as the phase relationships between wavefunctions, which are quantities linked to various mathematical qualities of waves and their probabilistic futures. When these waves are in phase–eg., demonstrating alignment within their peaks and troughs–they are essentially coherent, and capable of carrying information, a necessary and salient quality for our hypothetical wave. However, there is an issue: quantum coherence is incredibly fragile. It is easily disturbed by any slight environmental inconsistency, whether it be impurities, particles, and temperature fluctuations–rendering it immensely susceptible to decoherency, its inversion.

Nevertheless, there is still a method by which to circumvent this flaw. Within various Bose Einstein Condensates (BECs) such as superfluids, it has been observed that quantum coherence is maintained effectively and persistently. Within BECs, all particles share the same quantum state, behaving as a singular giant matter wave with a well-defined phase across the entire condensate. Within such an environment, information in the form of quantum wave excitations may be held indefinitely, giving reason for researchers to believe that some form of

BEC may be present in conscious brains. This would not only provide ample basis to assume the presence of a wave in the brain, but also explain numerous psychological phenomena such as memory, providing theoretical benefits to our overarching argument. Were BECs to be the substrate of consciousness (Worden, 2024), this would also expediently resolve numerous other debates in the community surrounding philosophy of mind–namely of whether AI might develop consciousness, too (impossible, were consciousness exclusively present in BECs,) and of panpsychism.

2 Implications For AI and Contemporary Technology

The theories proposed and described within this paper may not appear to have any obvious practical application; however, in truth, many of these hypotheses possess the potential to unveil many illuminating truths in the physical world, particularly with regards to AI and technology. It is clear, then, that discovering and confirming truths with regards to the metaphysical remains more salient than ever.

The question of whether or not AI may be capable of developing consciousness is one major intersection between philosophy of the mind and technology. Were machines capable of sharing a conscious experience much like that of humans, such a situation may give rise to a web of convoluted ethical dilemmas in a potentially robot-saturated future, with real and global implications for the general population. Previous approaches to such a problem have manipulated the principles of panpsychism to argue for the existence of conscious quality in machinery. According to such a theory, all matter, whether it be animate or fully inert, is imbued with a certain degree of consciousness. The reasoning behind such a statement goes that, if one were to assume unconsciousness in some objects, they would have to divide matter into disparate classes (of "conscious" matter and "nonconscious" matter) of which we have yet no proof of. Therefore, all matter, down to individual atoms and quarks, will contain an inherent and universal "consciousness," giving reason for AI to eventually develop its own order of higher thinking further into its development. Other arguments utilize functionalism–an emphasis on recognizing consciousness by its functions, rather than its physical properties/any other factors–in defending the feasibility of consciousness in AI. Essentially, to such thinkers, as long as AI exhibits traits synonymous to conscious beings, said AI will be conscious. Opponents may cite John Searle's Chinese Room Experiment to attest the prior argument's flaws. According to Searle's thought experiment, a man with no knowledge of the Chinese language may still posture and deliver letters/information in Chinese; Searle likens this ignorance to the way in which computers function, bringing a strong argument against claims of automated consciousness.

Nevertheless, such a question remains controversial and heavily debated, with arguments persisting in the modern day.

However, were we to apply the basic principles laid down by Friston, Worden and others within their respective theories, this circular and abstract debate may be swiftly resolved by scientific search for a wave. If the aforementioned theories hold true, then such a quantum excitation would represent the source of all consciousness, and its presence/absence would, in turn, indicate the existence of consciousness within a specific entity. Reinterpreting our understanding of consciousness in AI within such a context might usher in an era of heightened comprehension and epistemic wealth regarding both philosophy of the mind and the rapidly burgeoning field of contemporary technology.

References

Friston, K. (n.d.). *The free-energy principle: A unified brain theory?* ScienceDirect. https://www.uab.edu/medicine/cinl/images/KFriston_FreeEnergy_BrainTheory.pdf

Bottemanne, H., Abbott, L. F., Badcock, P. B., Bastos, A. M., Bettinger, J. S., Chen, W. G., Deneve, S., Friston, K., Friston, K. J., He, K., Hobson, J. A., Hohwy, J., Kwisthout, J., Lochmann, T., Pezzulo, G., Proietti, R., Rall, W., … Feldman, H. (2024, December 4). Bayesian brain theory: Computational neuroscience of belief. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2024.11.007

arXiv. (n.d.). *[Preprint collection]*. arXiv. https://arxiv.org/pdf/2405.17185