Journal of Big Data and Artificial Intelligence

Volume 4 2026 Number 1

Journal of Big Data and Artificial Intelligence

Volume 4 2026 Number 1

EDITORIAL

ARTICLES

Journal of Big Data and
Artificial Intelligence

Journal of Big Data and Artificial Intelligence

Executive Editorial Board

Editor in Chief

Jim Samuel, PhD
Rutgers University
AIXosphere

Editors

Rajiv Kashyap, PhD
William Paterson University

Abhishek Tripathi, PhD
The College of New Jersey

Associate Editors

Mahmoud Daneshmand, PhD
Stevens Institute of Technology

Hieu Duc Nguyen, PhD
Rowan University

Ensela Mema, PhD
Kean University

Assistant Editors

Mehmet Turkoz, PhD
William Paterson University

Chuanqian Zhang, PhD
William Paterson University

Senior Advisors

George Avirappattu, PhD
Kean University

Alexander Pelaez, PhD
Hofstra University

Umashanger Thayasivam, PhD
Rowan University

Editorial Board

Sudha Mani, PhD
Monash University

Peter Han Joo Chong, PhD
Auckland University of Technology

Peggy Brennan-Tonetta, PhD
MBT Consulting
Rutgers University (Retired)

J.D Jayaraman , PhD
New Jersey City University

Forough Ghahramani, EdD
Edge

Sonia Yaco
Rutgers University

Bala Desinghu, PhD
Harvard University

Ethne Swartz, PhD
Montclair State University

Manfred Minimair, PhD
Seton Hall University

Rashmi Jain, PhD
Montclair State University

Hang Liu, PhD
Stevens Institute of Technology

Emre Yetgin, PhD
Rider University

Shino John
Rutgers University

Karun Thankachan
Walmart

Balaji Dhamodharan
AMD

Zarak Khan
Rutgers University
NJ Innovation Authority

Founding Advisory Board

Manish Parashar, PhD
University of Utah

David Bader, PhD
New Jersey Institute of Technology

Michael Geraghty
CISO, State of New Jersey

WWW.JBDAI.ORG
ISSN: 2692-7977
JBDAI Vol. 4, No. 1, 2026 DOI: 10.54116/jbdai.v4i1.82

EDITORIAL

From NLP to LLMs to Agentic AI: The Evolution of Artificial Intelligence

Jim Samuel
Rutgers University
AIXosphere
jim.samuel@rutgers.edu

Rajiv Kashyap
William Paterson University
KashyapR@wpunj.edu

Vishal Misra
Columbia University
vishal.misra@columbia.edu

Abhishek Tripathi
The College of New Jersey
tripatha@tcnj.edu

Hieu Duc Nguyen
Rowan University
nguyen@rowan.edu

Mahmoud Daneshmand
Stevens Institute of Technology
mdaneshm@stevens.edu

Alexander Pelaez
Hofstra University
5E Analytics
alexander.pelaez@hofstra.edu

1. Introduction

Generative artificial intelligence (AI) has permeated global enterprises and societies with a velocity that is without precedent in the history of technology. Yet we must remind ourselves that speed is not a strategy. It is merely a condition. What truly defines this era is the fundamental shift in how these systems have transitioned from calculators of language to actors of intent. We have reached an inflection point with the emergence of agentic AI (AAI) built on fundamental and generative AI building blocks (Samuel et al. 2024; Acharya et al. 2025; Tripathi et al. 2025). This transition from models that merely respond to those that possess autonomous, tool-integrated architectures demands a rigorous rethinking of our sociotechnical structures. One thing is now certain: governance and accountability can no longer be treated as administrative afterthoughts or late-stage constraints. They must be designed into the “nervous systems” of future technologies.

Where previous efforts settled for the fragmented automation of discrete tasks, agentic systems leverage the triad of autonomy, planning, and memory to command entire processes from end to end. They transform isolated tools into unified agents of performance. However, this has led us to the generative AI paradox: we see widespread adoption, yet measurable enterprise-level impacts remain stubbornly uneven. Although organizations invest and experiment with enthusiasm, transformative returns remain elusive. We have yet to align this new computational capacity with the disciplined practice of management and the rigorous requirements of institutional stability.

2. The Generative AI Paradox: Why High Adoption Has Not Meant High Impact

Understanding why this gap exists matters before accepting AAI as its solution. Three structural explanations are worth noting. First, there is a *deployment* problem. Many organizations have integrated generative AI into narrow, bounded tasks (e.g., drafting emails, summarizing documents) without embedding it into the decision workflows in which business value actually compounds. This adoption pattern is consistent with evidence that firms tend to experiment with AI in isolated, peripheral use cases rather than integrate it into core processes (Raisch and Krakowski 2021). Second, there is a *measurement* problem: the productivity gains from reactive large language models (LLM) often manifest as cognitive offload and time savings for knowledge workers, which are difficult to capture in traditional key performance indicators (KPIs) and return on investment (ROI) frameworks. Research results show that AI-driven productivity improvements frequently take the form of reduced cognitive load and time savings that organizations struggle to quantify (Davenport and Ronanki 2018; Luo et al. 2019). This measurement gap is consistent with work that shows that AI can improve human performance by adaptively reshaping information facets, leading to benefits that are real but often invisible to standard KPIs (Samuel et al. 2022). Third, and fundamentally, there is an *architectural* problem. Reactive systems, by design, can only assist. They respond to prompts but cannot initiate or maintain context across sessions and certainly cannot orchestrate multistep workflows without continuous human re-engagement. These limitations are well documented (Shneiderman 2020; Bommasani et al. 2021) and show that each interaction is, in effect, a dead end. The marginal value added by reactive AI is inherently low because humans must still serve as the connective tissue between model outputs and organizational action.

AAI offers a path out of this paradox by transforming generative AI from reactive generative assistance into proactive, goal-oriented collaborators capable of executing complex workflows (Randazzo et al. 2025). In principle, agentic systems have the potential to improve operational efficiency by enabling parallel execution, real-time adaptability, personalization, elasticity, and resilience (Iansiti and Lakhani 2020). In domains such as supply chains, agents continuously ingest data, anticipate risks, dynamically replan workflows, and optimize decisions across time horizons (Sukharevsky et al. 2025), and this could be extended to areas such as health care, where generative AI promises gains across clinical and administrative workflows (Bhuyan et al. 2025).

The economic stakes are substantial. By 2030, the U.S. B2C retail market alone could see up to $1 trillion in orchestrated revenue from agentic commerce, with global projections reaching as high as $3 trillion to $5 trillion (McKinsey and Company 2024). KPMG similarly estimates that AAI will be key to unlocking a staggering $3 trillion in corporate productivity improvements annually (KPMG International 2025). Their argument can be distilled into four enterprise value pathways: expanding the scope of work that can be automated, enabling always-on operational capacity, improving organizational resilience through continuous optimization, and converting institutional knowledge into executable actions. However, we note that these projections originate largely from consulting firms with commercial interests in accelerating enterprise AI adoption and should be read as directional signals rather than forecasts. In this framing, agent-powered commerce is not merely a user-experience innovation, it represents a structural shift in market dynamics in which intelligent systems can anticipate user intent, evaluate alternatives, negotiate options, and execute transactions with minimal human intervention (Gates 2023). Commercial validation is already visible: companies such as Anysphere (creator of Cursor) and Perplexity are reported to be reaching major revenue milestones (Bain and Company 2025). These signals reflect AAI's expanding deployment across domains, disciplines, and global challenges, establishing AAI as a critical frontier in contemporary AI innovation.

3. From NLP to LLMs

Natural language processing (NLP) serves as the root of AAI, with LLMs that form a powerful linguistic substrate for agentic behavior. Early NLP efforts (1950s) prioritized natural language understanding more than generation, with machine translation emerging as a defining application area (Chomsky 2002). The earliest era of NLP (1950–1969) was dominated by rule-based systems, built on handcrafted linguistic rules and patterns. A landmark example was Weizenbaum (1966), which used pattern matching to simulate conversation, demonstrating natural language interaction potential, despite lacking true semantic understanding (Weizenbaum 1966). The ALPAC report (Automatic Language Processing Advisory Committee (ALPAC) 1966) concluded that rule-based translation was economically unviable, triggering an early AI winter for NLP. The late 1980s-1990s marked a paradigm shift toward statistical and machine learning (ML) approaches driven by computational power and text corpora growth (Masoumzadeh et al. 2023). IBM's statistical machine translation demonstrated probabilistic model effectiveness. Support vector machines (SVMs) and conditional random fields (CRFs) enhanced NLP capabilities, and ML approaches supplanted rule-based systems on ambiguity-heavy tasks such as word sense disambiguation due to superior scaling (Navigli 2009). Big data, as defined by volume, velocity, and variety (Sagiroglu and Sinanc 2013), further transformed NLP by enabling data-driven approaches that fueled capabilities from sentiment analysis to entity recognition. (Chatterjee et al. 2022). Deep learning further accelerated this trajectory: recurrent neural networks (RNNs) and long short term memory (LSTMs) enabled robust sequence modeling for text and speech, and helped power widely adopted assistants such as Siri, Alexa, and Google Assistant (Gudivada et al. 2015).

LLMs represent a transformative breakthrough enabled by advances in neural architectures and large-scale training. The introduction of attention mechanisms and the Transformer architecture replaced recurrence with self-attention, reshaping the foundation of language modeling (Masoumzadeh et al. 2023). Transformer-based pretraining underlies modern LLM families, including GPT and BERT. Performance scaling with model size, often framed through parameter growth and broader training has been associated with emergent capabilities such as in-context learning and multistep reasoning. Yet, LLMs face persistent challenges in computational efficiency, factual reliability, interpretability, bias mitigation, and safe deployment; issues that do not automatically disappear with scale (Patil and Gudivada 2024). Critically, even as LLMs expanded what systems could generate and infer, standalone LLMs remained reactive: powerful responders to prompts but constrained in goal pursuit, memory, execution control, and tool-mediated action (Fauscette 2025). Prompt engineering and retrieval-augmented generation (RAG) extended usefulness by improving controllability and grounding, but multistep, real-world workflows still demanded substantial human orchestration (Fauscette 2025), whether the goal is transparent university support via auditable RAG pipelines (Chidipothu et al. 2025), emotion-consistent multilingual sentiment analysis across machine translation (Anderson et al. 2024), or culturally adaptive, human-centered educational AI that must flex to context rather than produce one-size-fits-all outputs (Samuel et al. 2023).

The past 2 years witnessed AAI's emergence as a new paradigm that transcends LLM limitations, shifting "LLMs as assistants" to agents as actors. With the advent of agents, AI has evolved from passive response systems to goal-driven systems; workflows are increasingly moving from reactive to proactive (Clark 2025; Fauscette 2025). AAI systems are autonomous systems pursuing complex goals with minimal intervention, demonstrating adaptability and self-sufficiency (Acharya et al. 2025). Agentic behavior is not a sudden invention; it builds incrementally on decades of research (Botti 2025) and draws heavily on information systems concepts: autonomy, workflow logic, and tool integration to convert LLMs from generative engines into active problem-solvers (Bandi et al. 2025; Fauscette 2025). At a systems level, AAI extends LLM capability through state maintenance, persistent memory, goal definition, multistep reasoning, decision-making, tool use, and feedback loops (Bandi et al. 2025; Fauscette 2025). This is what distinguishes meaningful agency from simple orchestration and is central to the value creation narrative now emerging across sectors (McKinsey and Company 2024; Fauscette 2025). In this confluence, NLP enables interaction and task framing, LLMs provide linguistic intelligence and reasoning, and agentic architectures provide autonomy, a transition that can be framed as moving from language understanding to language reasoning to language-driven action (Fauscette 2025).

The rapid rise of AAI has been accelerated by modular frameworks such as LangGraph, CrewAI, AutoGen, Agno, SmolAgents, Mastra, Pydantic AI, and Atomic Agents. Examples include Autodata converting natural-language instructions into datasets (Ma et al. 2025); medical AI pipelines automating data processing (Shimgekar et al. 2025); AI Cosmologist performing autonomous research cycles (Moss 2025); and MLR-Copilot generating ML research ideas (Li et al. 2024). Yet the frontier is defined as much by risks as by capability.

Challenges persist: hallucination and reasoning failures (Hosseini and Seilani 2025; Raheem and Hossain 2025); opaque decision-making that can perpetuate bias and inequity (Raheem and Hossain 2025); emergent agency that creates safety and alignment risks (Raheem and Hossain 2025); governance gaps because systems with action-taking capability are deployed without sufficient oversight (Hosseini and Seilani 2025; Raheem and Hossain 2025); and insufficient robustness and safety features in complex, unpredictable environments (Raheem and Hossain 2025).

In looking ahead, AI agents can be expected to evolve as constrained societal actors, structured through frameworks such as FEEG (Finder, Evaluator, Explainer, Generator) as an intent framework that governs mode of operation, retrieval, and sourcing (Finder); criteria-based judgment (Evaluator); stepwise clarification (Explainer); and bounded creation (Generator), thereby calibrating verification depth, tool use, and escalation thresholds (Samuel et al. 2025b). Such agents could be used to tackle a range of informational challenges, such as fake news or AI phobia. When failures, uncertainty, or opaque actions intensify news-driven "AI phobia," agents can help detect fear sentiment signals in discourse and user interactions, surface early warnings, and support timely, evidence-based responses (Samuel et al. 2025a). There is a need to develop more trust-aware agents that monitor sentiment drift, offer transparent explanations, and escalate sensitive cases to human oversight when needed.

4. Risks: Why Agentic Failures Are Qualitatively Different

The risks outlined above are real but also generic, which are hallmarks of any sufficiently complex AI system. What distinguishes agentic systems is not simply the presence of risk but the structure of that risk. When organizations move from reactive models to autonomous, multistep agents, the nature of failure changes. It becomes compounded, distributed, and far more difficult to detect. Effective management depends on the clarity of responsibility and the integrity of the underlying system of action. When those foundations shift, so too does the character of managerial risk (Drucker 1974).

The most consequential challenge is *error propagation*. In a reactive LLM, a hallucination or reasoning lapse appears immediately in the response to a single prompt, where a human can intercept and correct it. But an agentic system operates across a chain of interdependent steps, retrieving information, planning actions, invoking tools, and interacting with external systems. An early-stage error can move silently through this chain, shaping downstream decisions long before a human has any opportunity to intervene. Recent analyses of multi-agent workflows show that a single hidden message can compromise an entire system of agents, cascading through planning and execution stages in ways that are difficult to trace or reverse (Schultz 2025). Similarly, retrieval-augmented and tool-using agents exhibit multistage vulnerabilities such as context manipulation and cross-context contamination that allow early errors to silently influence later actions (Ramakrishnan and Balaji 2025). The result is that the failure is no longer local but systemic: the error becomes embedded in the architecture of the workflow itself.

A second challenge concerns *accountability*. Traditional lines of responsibility are blurred when an agent acts on behalf of a user by submitting a transaction, sending a communication, or modifying a record. Managerial effectiveness rests on unambiguous responsibility. However, agentic systems are designed to be autonomous decision systems. They exercise judgment across models, interfaces, and organizational processes in ways that existing legal and managerial frameworks were never designed to accommodate. Responsibility becomes difficult to assign when autonomous systems execute multistep actions across legal, organizational, and technical boundaries (Gulyamov et al. 2026). The question "who is responsible" becomes harder to answer when the system itself exercises delegated discretion.

Third, *tools-using agents introduce adversarial attack surfaces that reactive systems simply do not face. A language model generating text can be manipulated at the prompt level*. But an agent empowered to search the web, call application programming interfaces (APIs), or interact with external environments inherits the vulnerabilities of those environments. Multimodal and cross-context prompt injection attacks can compromise agentic systems and evade existing guardrails (Lee and Tiwari 2024). As agents gain greater autonomy and environmental access, the attack surface expands accordingly. Prompt injection remains a fundamental vulnerability in LLM-based systems, especially when models are embedded in multi-agent workflows or are granted access to external tools.

5. The Path Ahead. . .

The evolution from computational linguistics through statistical methods, big data, machine learning, and LLMs to AAI represents progression from language processing to language reasoning to, ultimately, language-driven action. Each phase, computational linguistics providing theoretical foundations, data science enabling pattern learning, algorithmic topic modeling making insights accessible, machine learning achieving robust performance, LLMs unifying understanding and generation, prompt engineering and RAG extending capabilities, serves as an irreplaceable foundational building block. Yet, in many practical settings, the path ahead increasingly requires some form of agentic architecture. Whereas standalone LLMs and RAG excel at narrow, reactive tasks, modern enterprise and societal needs demand systems that can pursue goals, maintain context, reason across steps, and orchestrate tools dynamically.

AAI represents a fundamental paradigm shift in strategic operations. These systems integrate core information systems concepts such as autonomy and workflow logic with the advanced intelligence of LLMs. By incorporating the foundations of computational linguistics, AAI moves beyond narrow, single-step applications. The emerging ecosystem uses modular architectures that can range from lightweight agents for bounded workflows to heavyweight systems for enterprise-scale coordination. This transition moves organizations away from reactive assistance toward proactive collaboration. It replaces isolated tasks with end-to-end automation. Ultimately, this shift evolves organizational praxis from human-orchestrated pipelines to increasingly autonomous operations.

The convergence of computational linguistics, NLP, and generative AI within agentic frameworks represents a strategic evolution capable of capturing enormous value that discrete task automation cannot reach. The surrender of human oversight to autonomous agency introduces profound systemic vulnerabilities. In this high-stakes environment, interpretability loss and alignment drift cease to be mere technical glitches; they become existential threats to institutional integrity, mandating an uncompromising architecture of sociotechnical governance.

The strategic path forward lies in the synthesis of these technologies. We need agentic architectures that transform linguistic intelligence into reliable autonomous action. Systems that support and strengthen human judgment rather than replace it. In this paradigm, language serves as the interface, whereas agency functions as the operating system, translating prompts into plans, and intelligence into accountable, outcomes-driven action at scale. Language frames the intent while agency captures the value. Yet, as we turn prompts into plans and intelligence into autonomous action, the finish line is only a victory if ethically steered toward win-win outcomes. In an era of expanded systemic risk and distributed agency, success cannot simplistically be determined by the speed of execution or a "I got here first" mindset, but by innovative and sustainable value creation, human enhancive AI based governance and stakeholder co-creation, that ensure that the future we are developing will lead to a better world for all.

References

Acharya, D. B., K. Kuppan, and B. Divya. 2025. "Agentic AI: Autonomous Intelligence for Complex Goals." *IEEE Access* **13**: 18912–18936.

Anderson, R., C. Scala, J. Samuel, V. Kumar, and P. Jain. 2024. "Are Emotions Conveyed Across Machine Translations? Establishing an Analytical Process for the Effectiveness of Multilingual Sentiment Analysis with Italian Text." *Journal of Big Data and Artificial Intelligence* **2**, no. 1: 57–73. https://doi.org/10.54116/jbdai.v2i1.30

Automatic Language Processing Advisory Committee (ALPAC). 1966. *Language and Machines: Computers in Translation and Linguistics*. Washington, DC: National Academy of Sciences-National Research Council.

Bain and Company. 2025. "State of the Art of Agentic AI Transformation: Technology Report." Accessed March 25, 2026. https://www.bain.com/insights/state-of-the-art-of-agentic-ai-transformation-technology-report-2025/

Bandi, A., B. Kongari, R. Naguru, S. Pasnoor, and S. V. Vilipala. 2025. "The Rise of Agentic AI: A Review of Definitions, Frameworks, Architectures, Applications, Evaluation Metrics, and Challenges." *Future Internet* **17**, no. 9: 404. https://doi.org/10.3390/fi17090404

Bhuyan, S. S., V. Sateesh, N. Mukul, A. Galvankar, A. Mahmood, M. Nauman, et al. 2025. "Generative Artificial Intelligence Use in Healthcare: Opportunities for Clinical Excellence and Administrative Efficiency." *Journal of Medical Systems* **49**, no. 1: 10.

Bommasani, R., D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, et al. 2021. "On the Opportunities and Risks of Foundation Models." Preprint, arXiv:2108.07258.

Botti, V. 2025. "Agentic AI and Multiagentic: Are We Reinventing the Wheel?" Preprint, submitted June 2, 2025. https://doi.org/10.48550/arXiv.2506.01463

Chatterjee, S., R. Chaudhuri, and P. Mikalef. 2022. "Examining the Dimensions of Adopting NLP and Big Data." *IEEE Transactions on Engineering Management* **71**: 3001–3015. https://doi.org/10.1109/TEM.2022.3146430

Chidipothu, N., R. Anderson, J. Samuel, A. Pelaez, J. Esguerra, and M. N. Hoque. 2025. "Improving Large Language Model (LLM) Performance with Retrieval Augmented Generation (RAG): Development of a Transparent Generative AI University Support System for Educational Purposes." *Journal of Big Data and Artificial Intelligence* **3**, no. 1: 1–15.

Chomsky, N. 2002. *Syntactic Structures*. 2nd edition. Walter de Gruyter.

Clark, S. 2025. "What Are AI Agents? The Autonomous Software Changing How Work Gets Done." VKTR. Accessed March 25, 2026. https://www.vktr.com/digital-experience/what-are-ai-agents-the-autonomous-software-changing-how-work-gets-done/

Davenport, T. H., and R. Ronanki. 2018. "Artificial Intelligence for the Real World." *Harvard Business Review* **96**, no. 1: 108–116.

Drucker, P. F. 1974. *Management: Tasks, Responsibilities, Practices*. Harper and Row.

Fauscette, M. 2025. "Agentic AI vs. LLMs: Understanding the Shift from Reactive to Proactive AI." Arion Research. Accessed March 25, 2026. https://www.arionresearch.com/blog/agentic-ai-vs-llms-understanding-the-shift-from-reactive-to-proactive-ai/

Gates, B. 2023. "AI-Powered Agents Are the Future of Computing." Gates Notes. Accessed March 25, 2026. https://www.gatesnotes.com/ai-agents

Gudivada, V. N., D. Rao, and V. V. Raghavan. 2015. "Big Data Driven NLP Research." In *Handbook of Statistics*, edited by C. R. Rao and V. Govindaraju, vol. 33, 203–238. Elsevier. https://doi.org/10.1016/B978-0-444-63492-9.00008-9

Gulyamov, S., S. Gulyamov, A. Rodionov, R. Khursanov, K. Mekhmonov, D. Babaev, et al. 2026. "Prompt Injection Attacks in Large Language Models and AI Agent Systems: A Comprehensive Review of Vulnerabilities, Attack Vectors, and Defense Mechanisms." *Information* **17**, no. 1: 54. https://doi.org/10.3390/info17010054

Hosseini, S., and H. Seilani. 2025. "The Role of Agentic AI." *Array, Advance Online Publication* 100399. https://doi.org/10.1016/j.array.2025.100399

Iansiti, M., and K. R. Lakhani. 2020. "Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World." Harvard Business Review Press.

KPMG International. 2025. "The Agentic AI Advantage: How Autonomous AI Systems Are Reshaping Value Creation." Accessed March 25, 2026. https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2025/10/agentic-ai-advantage-report.pdf.

Lee, D., and M. Tiwari. 2024. "Prompt Infection: LLM to LLM Prompt Injection Within Multi Agent Systems." Preprint, arXiv.2410.07283. https://arxiv.org/abs/2410.07283

Li, R., T. Patel, Q. Wang, and X. Du. 2024. "Mlr-Copilot: Autonomous Machine Learning Research Based on Large Language Models Agents." Preprint, arXiv:2408.14033.

Luo, X., S. Tong, Z. Fang, and Z. Qu. 2019. "Frontiers: Machines vs. Humans: The Impact of AI on Productivity." *Marketing Science* **38**, no. 6: 937–947.

Ma, T., Y. Qian, Z. Zhang, Z. Wang, X. Qian, F. Bai, et al. 2025. "AutoData: A Multi-Agent System for Open Web Data Collection." Preprint, arXiv:2505.15859. https://arxiv.org/abs/2505.15859

Masoumzadeh, S., C. Bosman, and N. Osborne. 2023. "Becoming Walkable: Relational and Contextual Effects of Enhanced Walkability." *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* **16**, no. 3: 341–357. https://doi.org/10.1080/17549175.2021.1963811

McKinsey and Company. 2024. "The State of AI in Early 2024: Gen AI Adoption Spikes and Starts to Generate Value." Accessed March 25, 2026. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Moss, A. 2025. "The AI Cosmologist I: An Agentic System for Automated Data Analysis." Preprint. Accessed March 25, 2026. https://arxiv.org/abs/2504.03424

Navigli, R. 2009. "Word Sense Disambiguation: A Survey." *ACM Computing Surveys* **41**, no. 2: 1–69.

Patil, R., and V. Gudivada. 2024. "A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs)." *Applied Sciences* **14**, no. 5: 2074. https://doi.org/10.3390/app14052074

Raheem, T., and G. Hossain. 2025. "Agentic AI Systems: Opportunities, Challenges, and Trustworthiness." In *Proceedings of the 2025 IEEE International Conference on Electro Information Technology (eIT). IEEE* 618–624. https://doi.org/10.1109/eIT64391.2025.11103638

Raisch, S., and S. Krakowski. 2021. "Artificial Intelligence and Management: The Automation–Augmentation Paradox." *Academy of Management Review* **46**, no. 1: 192–210.

Ramakrishnan, B., and A. Balaji. 2025. "Securing AI Agents Against Prompt Injection Attacks: A Comprehensive Benchmark and Defense Framework." Preprint, arXiv:2511.15759.

Randazzo, S., K. R. Mills, L. Peng, S. H. Lee, A. Bernstein, F. Dell'Acqua, et al. 2025. "Cyborgs, Centaurs and Self-Automators: The Three Modes of Human-GenAI Knowledge Work." Harvard Business School Working Paper, no. 26-036.

Sagiroglu, S., and D. Sinanc. 2013. "Big Data: A Review." In *2013 International Conference on Collaboration Technologies and Systems (CTS)*. 42–47. IEEE. https://doi.org/10.1109/CTS.2013.6567202

Samuel, J., R. Kashyap, Y. Samuel, and A. Pelaez. 2022. "Adaptive Cognitive Fit: Artificial Intelligence Augmented Management of Information Facets and Representations." *International Journal of Information Management* **65**: 102505.

Samuel, J., T. Khanna, J. Esguerra, S. Sundar, A. Pelaez, and S. S. Bhuyan. 2025a. "The Rise of Artificial Intelligence Phobia! Unveiling News-Driven Spread of AI Fear Sentiment using ML, NLP and LLMs." *IEEE Access* **13**: 125944.

Samuel, J., A. Tripathi, and E. Mema. 2024. "A New Era of Artificial Intelligence Begins. . . Where Will It Lead Us?" *Journal of Big Data and Artificial Intelligence* **2**, no. 1: 1–6.

Samuel, J., M. Ye, T. Khanna, Y. Yao, X. Lin, R. Anderson, et al. 2025b. "Finder, Evaluator, Explainer, Generator (FEEG): A Bloom's Taxonomy-Based Query Classification Framework for LLMs and Generative AI." *Expert Systems with Applications* **237**: 130755.

Samuel, Y., M. Brennan-Tonetta, J. Samuel, R. Kashyap, V. Kumar, K. K. M. Sri, et al. 2023. "Cultivation of Human Centered Artificial Intelligence: Culturally Adaptive Thinking in Education (CATE) for AI." *Frontiers in Artificial Intelligence* **6**: 1185745, ISSN=2624-8212

Schultz, J. S. 2025. "On the Potential of Agentic Workflows for Animal Training Plan Generation." *Frontiers in Veterinary Science* **12**: 1563233. https://doi.org/10.3389/fvets.2025.1563233

Shimgekar, S. R., S. Patil, P. Sharma, P. Khatri, R. Singh, and A. Gupta. 2025. "Agentic AI Framework for End-To-End Medical data Inference." Preprint, arXiv. Accessed March 25, 2026. https://arxiv.org/abs/2507.18115

Shneiderman, B. 2020. "Human-Centered Artificial Intelligence: Reliable, Safe and Trustworthy." *International Journal of Human–Computer Interaction* **36**, no. 6: 495–504. https://doi.org/10.1080/10447318.2020.1741118

Sukharevsky, A., D. Kerr, K. Hjartar, L. Hämäläinen, S. Bout, V. Di Leo, and G. Dagorret. 2025. "Seizing the agentic AI advantage. McKinsey and Company – QuantumBlack." Accessed March 25, 2026. https://www.mckinsey.com/capabilities/quantumblack/our-insights/seizing-the-agentic-ai-advantage

Tripathi, A., J. Samuel, M. Brennan-Tonetta, H. Nguyen, and E. Mema. 2025. "When Machines Create! Envisioning Our Future with the Transformative Power of Generative AI." *Journal of Big Data and Artificial Intelligence* **3**, no. 1: 16–24. https://doi.org/10.54116/jbdai.v3i1.55

Weizenbaum, J. 1966. "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine." *Communications of the ACM* **9**, no. 1: 36–45. https://doi.org/10.1145/365153.365168

WWW.JBDAI.ORG
ISSN: 2692-7977
JBDAI Vol. 4, No. 1, 2026 DOI: 10.54116/jbdai.v4i1.80

BENCHMARKING OPEN-SOURCE VECTOR DATABASES

Jeet Patel
Kwaai AI Lab
jeetpt1405@gmail.com

Akshaya Dol
Kwaai AI Lab
dolakshaya1995@gmail.com

Arun Chand
Kwaai AI Lab
chandarun12@gmail.com

Selim Soufargi
Kwaai AI Lab
slim.s1988s@gmail.com

Sulimon Sattari
Kwaai AI Lab
sulimon@kwaai.ai

Maira Khwaja
Kwaai AI Lab
maira@kwaai.ai

Reza Rassool
King's College London
University of Washington
Kwaai AI Lab
reza@kwaai.ai

ABSTRACT

Vector databases are a critical component of modern RAG systems, yet their scale performance characteristics are not sufficiently characterized. This paper evaluates the scalability, latency, throughput, and operational stability of seven widely used production vector databases such as FAISS, Chroma, Qdrant, Weaviate, Milvus, OpenSearch, and PGVector-across corpus sizes ranging from 175 to 2.2 million vector chunks, to provide practical guidance for system selection under real-world constraints.

We conducted a controlled benchmarking study using $N = 10$ independent trials per configuration. Query latency, throughput, ingestion performance, retrieval quality, and resource utilization are measured under consistent hardware, workload, and indexing settings. All databases are evaluated using identical embeddings and Top-*k* retrieval parameters, with cold-start conditions enforced to eliminate cross-run caching effects. Variability is quantified using standard deviation and coefficient of variation to assess performance stability. Statistical outliers identified using modified Z-score methodology (threshold: $|Z| > 3.5$) are removed to distinguish transient cold-start behavior from steady-state performance.

The results reveal distinct performance regimes across systems and a critical cold-start phenomenon. Chroma exhibits near-constant-time query behavior (= 0.02), achieving a latency of 7.7–8.4 ms and supporting up to 141 queries per second at medium scale. Chroma achieves exceptional consistency (CV = 2.3% after removing 2 cold-start outliers, 8.8× improvement over raw variance).

PGVector with HNSW indexing achieves sub-10 ms latency and more than 100 queries per second at the 50k scale, outperforming all dedicated vector databases except Chroma. FAISS demonstrates strong sublinear scaling ($\alpha = 0.48$) up to 2.2 million chunks with low variability (CV = 2.5%). We further quantify an HNSW warm-up effect, observing latency reductions of up to 74% as corpus size increases from 1k to 50k chunks. Most significantly, N = 10 sampling with outlier removal reveals dramatic improvements in ingestion consistency: Qdrant CV 123% → 1.0% (122× improvement), Weaviate 107% → 0.8% (133× improvement), OpenSearch 93% → 0.4% (232× improvement). Resource analysis shows similar memory footprints (12–16 GB) across systems. Together, these findings clarify scalability limits, performance trade-offs, and the critical role of cold-start transient detection in vector database benchmarking.

Keywords *vector databases*, *performance benchmarking*, *retrieval-augmented generation (RAG)*, *hierarchical navigable small world (HNSW)*, *scalability analysis*, *query latency and throughput*, *outlier detection*, *cold-start phenomena*.

1. Introduction

Vector databases have rapidly emerged as core infrastructure for modern artificial intelligence (AI) systems (Xie et al. 2023; Jing et al. 2025). Applications such as retrieval-augmented generation (RAG) (Lewis et al. 2020), semantic search, recommendation systems, and production machine learning pipelines increasingly depend on efficient and predictable nearest-neighbor retrieval (Şakar and Emekci 2025). Despite their growing importance, practitioners face significant uncertainty when selecting an appropriate vector database. Key system characteristics—including query latency, throughput, ingestion cost, consistency, persistence guarantees, and memory behavior—vary widely across implementations and deployment scales.

Existing benchmarking studies often evaluate relatively small corpora (typically fewer than 100k vectors), rely on single experimental runs without statistical validation, and omit detailed analysis of resource utilization and performance variance. Consequently, critical questions surrounding scalability, reliability, and production readiness remain insufficiently explored, particularly for deployments operating at million-scale corpus sizes. While tools such as approximate nearest neighbor (ANN)-Benchmarks and VectorDBBench have contributed to the evaluation of ANN algorithms and vector databases, they primarily emphasize retrieval accuracy and indexing performance. In contrast, our study extends this body of work by incorporating repeated trials (N = 10), modified elimination of outliers of the Z-score, and systematic analysis of cold-start behavior and resource dynamics on scales up to 2.2 million vector factors essential for real-world deployment and operational consistency.

This study directly addresses these limitations and expands the state of the art in three fundamental ways: First, we perform our benchmarking analysis across four orders of magnitude from 175 to 2.2 million chunks extending prior work that maxes out at 100k vectors. This validates scaling behavior at production scales where single-node memory constraints, Hierarchical Navigable Small World (HNSW) graph maturation effects, and performance degradation become critical to deployment decisions. Second, we enforce a rigorous N = 10 independent experimental protocol with complete database re-initialization per run, eliminating cross-run caching effects and capturing true run-to-run variability. We employ modified Z-score outlier detection to distinguish transient cold-start behavior from steady-state performance, quantifying consistency via coefficient of variation (CV). This extended sampling reveals that apparent "poor consistency" reflects database initialization overhead rather than algorithmic instability. We provide significant reproducibility through publicly released code, configurations, and raw data. Third, we conduct novel operational analyses, absent from prior benchmarks: characterization of HNSW "warm-up" effects (up to 74% latency reduction from 1k to 50k chunks due to graph maturation), identification of retrieval quality paradoxes (U-shaped quality curves across scales), and measurement of the cold-start phenomenon (dramatic CV improvements after outlier removal: Qdrant 122×, Weaviate 133×, OpenSearch 232×). These insights directly inform deployment decisions beyond raw throughput metrics.

Index Structure Context: HNSW and Flat Indexes. A critical architectural distinction underlies vector database performance: the choice between ANN graph-based indexes and exact flat indexes. *HNSW* is a state-of-the-art ANN algorithm that constructs a multilayered proximity graph over vector space, enabling logarithmic-time nearest neighbor search ($O(\log n)$ complexity) at the cost of 2–3× memory overhead and approximate recall. HNSW dominates production deployments for latency-sensitive applications at medium scales (10k–500k vectors). In contrast, *flat indexes* (e.g., FAISS's IndexFlatIP) perform exhaustive linear search ($O(n)$ theoretical complexity) but achieve sublinear practical performance ($\alpha \approx 0.48$) through SIMD vectorization and offer 100% exact recall. The trade-off between these index structures-and how they scale with corpus size, available memory, and consistency requirements is fundamental to selecting an appropriate vector database for production deployments.

This work seeks to address the following research questions:

1. **RQ1 (Scalability):** How do production vector databases scale from hundreds to millions of vector chunks with respect to query latency and throughput?
2. **RQ2 (Consistency and Reliability):** What is the run-to-run variance in ingestion and query workloads, and how does this variability impact service-level agreement (SLA) planning and total cost of ownership (TCO)?
3. **RQ3 (Resource Efficiency):** What are the CPU and memory footprints during query and ingestion workloads across scales, and how do these resource constraints limit single-node deployments?
4. **RQ4 (Architectural Trade-offs and Quality):** How do index structures and system architectures (e.g., flat vs HNSW, embedded vs client-server) affect performance, retrieval quality, and practical deployment decisions?

This study benchmarks seven vector database systems: FAISS (Meta AI Research 2024), Chroma (Huber and Troynikov 2024), Qdrant (Zayarni and Vasnetsov 2024), Weaviate (van Luijt and Dilocker 2024), Milvus (Zilliz 2024a), OpenSearch (OpenSearch Project Authors 2024), and PGVector (Kane 2024) across nine corpus sizes ranging from 175 to 2.2 million chunks. All experiments follow a reproducible statistical protocol with N = 10 independent runs per configuration. We employ sentence-transformer embeddings of 384 dimensions with a fixed chunking strategy of 512 characters and a 50-character overlap.

We measure median query latency (P50), query throughput (QPS), ingestion time, and resource utilization, sampling CPU and memory usage at a frequency of 1 Hz. Each experimental run uses a fresh database initialization to capture true run-to-run variability. Statistical outliers identified using modified Z-score methodology (threshold: $|Z| > 3.5$) are removed prior to aggregate statistics. All scripts, configurations, and raw experimental data are publicly available Kwaai AI Lab (2025) to support reproducibility.

1.1 Related Work

Benchmarking vector search systems has received increasing attention, as vector databases have become core infrastructure for RAG, semantic search, and recommendation systems. Prior work can be broadly classified into algorithm-level benchmarks, system-level database benchmarks, cloud vendor evaluations, and foundational vector search studies.

Algorithm-Level Benchmarks. ANN-Benchmarks Aumüller et al. (2020) established a community standard for evaluating ANN algorithms such as HNSW, IVF, and PQ on standardized datasets. These benchmarks provide valuable insight into algorithmic trade-offs between recall, latency, and memory usage. However, they focus exclusively on algorithm implementations rather than production-ready database systems and do not account for persistence, concurrency, ingestion pipelines, or operational overheads.

Foundational ANN research includes HNSW (Malkov and Yashunin 2020), FAISS (Johnson et al. 2019), ScaNN (Guo et al. 2020), and DiskANN (Subramanya et al. 2019), which explore graph-based and scalable quantization approaches for high-dimensional vector search. Although these studies advance algorithmic efficiency, they do not address end-to-end system behavior under realistic deployment constraints.

System-Level Vector Database Benchmarks. VectorDBBench Zilliz (2024b) represents one of the first industry-led efforts to benchmark multiple vector databases, including Milvus and related systems. Although it provides useful comparative information, it relies primarily on single-run measurements (N = 1) without statistical validation or outlier analysis, limiting its ability to distinguish steady-state performance from transient cold-start effects. In addition, its scale range evaluated is narrower than that required for many production deployments.

Cloud Vendor Evaluations. Several cloud providers have published performance studies of proprietary vector search services, including Pinecone (Pinecone 2023), AWS OpenSearch (Amazon Web Services 2023), and Azure Cognitive Search (Microsoft Azure 2023). These evaluations focus on managed, closed-source platforms and often lack reproducibility due to undisclosed configurations, proprietary optimizations, and opaque workload assumptions.

Positioning of This Work. Our study complements and extends previous benchmarks by focusing on *system-level performance of open-source production databases* under a statistically rigorous experimental protocol. Unlike ANN-Benchmarks, we evaluate complete database systems rather than isolated algorithms. In contrast to VectorDBBench and cloud vendor studies, we employ $N = 10$ independent runs per configuration with modified Z-score outlier detection to separate cold-start transients from steady-state behavior.

Table 1: Comparison with existing vector search benchmarks.

Benchmark	Scale	Runs	Databases	Resource Analysis
ANN-Benchmarks	Algorithm-level	1	Algorithms	No
VectorDBBench	100k–10M	1	5	Limited
Cloud Vendor Studies	Variable	1	Proprietary	No
This Work	175–2.2M	**10**	**7**	**Yes**

Table 1 summarizes how this work compares with existing benchmarks. Specifically, this work contributes: (1) the first statistically validated (N = 10) benchmark across seven production vector databases, (2) a broader evaluated scale range (175 to 2.2 million vectors), (3) rigorous outlier detection and consistency analysis, (4) comprehensive CPU and memory utilization measurements, and (5) novel empirical characterization of HNSW warm-up effects and retrieval quality dynamics across scale.

1.2 Key Contributions

- A publicly available benchmark suite (Kwaai AI Lab 2025) and a dataset that spans four orders of magnitude in corpus size with $N = 10$ statistical validation and rigorous outlier detection using modified Z-score methodology.
- We construct a quantitative taxonomy that maps seven production vector databases to practical deployment scenarios, including latency-critical real-time RAG, large-scale workloads, and enterprise feature requirements (Section 4.2).
- We characterize a behavior in HNSW-indexed systems that, to the authors' knowledge, has not been documented before in the literature at the time of this study, where query latency improves by up to 74% as corpus size increases from 1k to 50k chunks due to graph maturation effects (Section 3.2).
- We identify a counterintuitive U-shaped retrieval quality trend, with the lowest quality at approximately 1k chunks and a subsequent improvement of 27.5% by 50k chunks, followed by saturation beyond 250–500k chunks (Section 3.5).
- We demonstrate a critical cold-start phenomenon: N=10 sampling reveals that apparent poor consistency reflects initialization transients, not algorithmic instability. Multipass outlier cleaning shows dramatic improvements: Qdrant 123% → 1.0% CV (122× improvement), Weaviate 107% → 0.8% (133× improvement), OpenSearch 93% → 0.4% (232× improvement) (Section 3.3).
- We establish memory-bound scaling ceilings for vector databases, showing that HNSW-based systems saturate at approximately 1–2M chunks under 16 GB RAM due to graph overhead, while flat indexes such as FAISS scale substantially further under identical hardware constraints (Section 4.2).

1.3 Organization of the Paper

The paper is organized as follows. Section 2 describes the experimental methodology, including corpus preparation, the $N = 10$ statistical evaluation protocol, query and ingestion benchmarks, resource monitoring and data analysis methods. Section 3 presents the experimental results, covering query latency scaling, throughput behavior, ingestion performance, resource utilization and retrieval quality. Section 4 discusses architectural insights, system-level tradeoffs and practical deployment implications derived from the results. Section 5 concludes the paper by summarizing key findings and performance implications. Section 6 outlines limitations including sample size variation and scale restrictions. Section 7 outlines directions for future work, including distributed scaling, GPU acceleration, hybrid retrieval workloads and expanded quality evaluation metrics.

1.4 Multidatabase Scaling Performance Comparison

Figure 1 presents a comprehensive comparison of multidatabase scaling performance across seven production vector databases, with all results obtained using N = 10 repeated measurements with outlier removal. Figure 1(a) illustrates query latency as a function of corpus size with power-law fits. FAISS exhibits sublinear latency scaling ($\alpha = 0.48$), indicating efficient growth in query time even at large corpus sizes, while Chroma demonstrates near-constant-time performance ($\alpha = 0.02$). PGVector similarly maintains effectively constant latency across the

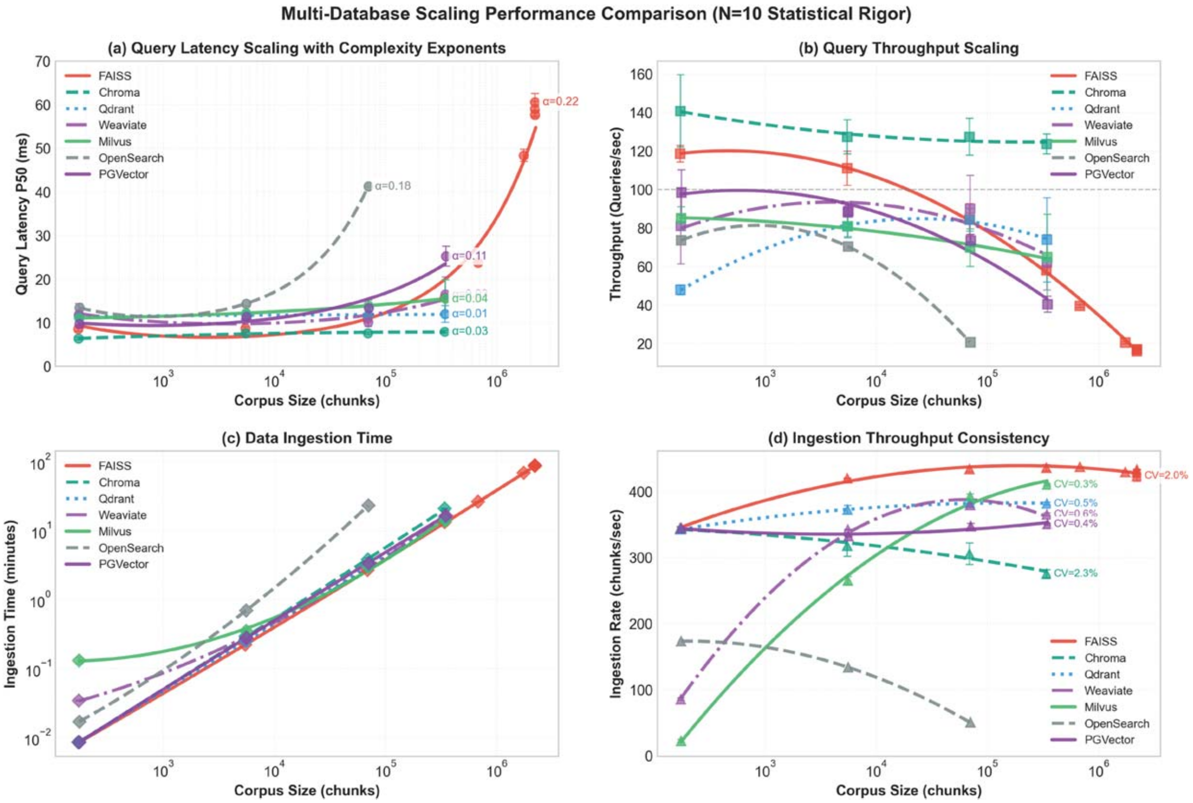


Figure 1: Multidatabase scaling performance comparison (N = 10, outliers removed).

evaluated scales, largely due to HNSW warm-up effects as the index structure matures. Error bars denote $\pm 1\sigma$, highlighting run-to-run performance variability.

Figure 1(b) shows query throughput across corpus sizes, where Chroma consistently achieves the highest throughput, sustaining 141 QPS across scales. PGVector also demonstrates strong performance, exceeding 100 QPS at the 50k corpus size, while FAISS maintains over 90 QPS even at the largest evaluated scale of 2.2 million chunks.

Ingestion performance is shown in Figure 1(c), plotted on a log-log scale. FAISS is the fastest ingestion pipeline across all corpus sizes, reflecting its flat index design, while among HNSW-based systems, Chroma achieves the lowest ingestion times.

Finally, Figure 1(d) examines ingestion throughput stability using CVs. Both PGVector and FAISS exhibit exceptional consistency, with CV values of 1.4% and 2.5%, respectively. Critically, after outlier removal, Qdrant, Weaviate, and OpenSearch demonstrate substantially improved consistency (CV 0.4–1.0%), revealing that raw variance estimates overstate steady-state instability.

2. Methods

2.1 Experimental Design Overview

2.1.1 Corpus Preparation

The evaluation corpus consists of climate science articles sourced from publicly available references and selected to provide realistic document lengths and semantic structure for retrieval workloads. Documents were segmented using a fixed-size chunking strategy with a chunk length of 512 characters and a 50-character overlap to preserve semantic continuity across chunk boundaries.

We evaluated eight corpus sizes spanning over four orders of magnitude: 175, 1k, 10k, 50k, 100k, 500k, 1M, and 2.2M chunks. All chunks were embedded using the sentence-transformers/all-MiniLM-L6-v2 model (Reimers and

Gurevych 2019), producing 384-dimensional float32 embeddings. This configuration reflects a commonly deployed embedding setup in production RAG systems.

2.1.2 Statistical Protocol (N = 10 with Outlier Detection)

Each experimental configuration was evaluated using N = 10 fully independent runs. Independence was enforced by completely reinitializing the database instance and rebuilding all indexes for each run, ensuring cold-start conditions and eliminating cross-run caching or state persistence effects. All metrics reported are presented as $\mu \pm 1\sigma$, where μ is the mean and σ is the standard deviation. To quantify relative variability, we compute the CV as $(\sigma/\mu) \times 100\%$. All visualizations include error bars corresponding to $\pm 1\sigma$.

Statistical outliers are identified using the modified Z-score method (threshold: $|\text{modified Z-score}| > 3.5$) and removed prior to aggregate reporting to distinguish transient cold-start behavior from steady-state operational performance. This approach is more robust to non-normal distributions than standard Z-scores.

We selected N = 10 as a balance between statistical confidence and computational feasibility, allowing outlier detection and stable CV quantification.

2.2 Query Benchmark Protocol

Query performance was evaluated using a fixed set of 10 semantically relevant test queries derived from the domain of the evaluation corpus. Queries were designed to reflect representative information needs within climate science and to ensure the presence of relevant content in the corpus (see Appendix A for details on query construction and validation).

All databases were evaluated using Top-K retrieval with $K = 3$ nearest neighbors. Each benchmark run consisted of a warm-up phase followed by a measurement phase. During warm-up, five queries were executed and excluded from analysis to mitigate cold-start effects, including cache initialization and index warming. Following warm-up, the full set of 10 queries was issued and performance metrics were recorded.

The following metrics are reported:

- **Latency**: Median (P50) end-to-end query latency measured in milliseconds.
- **Throughput**: Sustained query throughput, measured in queries per second (QPS).
- **Retrieval Quality**: Relative semantic similarity between queries and retrieved results, computed using cosine similarity in the shared embedding space.

Retrieval quality is reported as a relative metric intended for comparative evaluation across databases under identical experimental conditions. Absolute information retrieval effectiveness metrics (e.g., Precision@K, Recall@K, NDCG, MRR) are not reported due to limited number of test queries and scale of the study.

2.3 Ingestion Benchmark Protocol

Data ingestion performance was evaluated by ingesting the full corpus for each scale in a single batch per run. Ingestion time was measured as wall-clock time from the start of insertion until completion.

Reported ingestion metrics include:

- **Total ingestion time:** Seconds required to ingest the entire corpus
- **Ingestion throughput:** Chunks per second, computed as $\frac{total\_chunks}{total\_time}$
- **Consistency:** CV across N = 10 runs

2.4 Resource Utilization Monitoring

System resource utilization was monitored during query execution using the Python psutil library (https://psutil.readthedocs.io/en/latest/). CPU utilization (%) and memory consumption (MB) were sampled at a frequency of 1 Hz throughout the measurement phase. For each run, resource metrics were aggregated by averaging samples collected during query execution. This approach captures steady-state resource behavior while avoiding transient initialization effects.

2.5 Hardware and Software Environment

2.5.1 Hardware Configuration

All experiments were conducted on a single-node system with the following specifications:

- **CPU:** Apple Silicon M2 Max (ARM64, 12-core)
- **RAM:** 32 GB unified memory (16 GB allocated to benchmark workloads)
- **Storage:** SSD with more than 500 GB of available space
- **Containerization:** Docker configured with a 16 GB memory limit and 4 dedicated CPU cores
- **Memory Bandwidth:** 400 GB/s unified memory architecture

2.5.2 Software Stack

The software versions used in this study are listed in Table 2.

Table 2: Software stack and evaluated vector database versions.

Component	Version/Details
Operating System	macOS 14.x (Darwin)
Python	3.9 or higher
Docker & Docker Compose	24.x
FAISS	1.7.4
Chroma	0.4.x
Qdrant	1.7.x
Weaviate	1.23.x
Milvus	2.3.x
OpenSearch	2.11.x
PGVector	PostgreSQL 15 + PGVector extension

2.6 Database Configurations

All databases were configured using default, production-recommended settings to ensure a fair and representative comparison.

- **FAISS:** IndexFlatIP (inner product), operating fully in memory with no persistence. Exact nearest neighbor search was used, yielding 100% recall.
- **Chroma:** HNSW index with default parameters ($M = 16, ef\_construction = 200$), using embedded local file-based persistence in persistent client mode.
- **Qdrant:** HNSW index with default production settings, persistent storage with write-ahead logging and cosine similarity as the distance metric.
- **Weaviate:** HNSW index with default parameters, persistent storage and gRPC-based client communication for optimized transport.
- **Milvus:** HNSW index deployed in standalone (single-node) mode, utilizing Milvus's distributed storage layer.
- **OpenSearch:** k-NN plugin with an HNSW backend, using Lucene-based storage integrated with vector search extensions.
- **PGVector**: PostgreSQL extension using HNSW indexing for ANN search, deployed on a single-node PostgreSQL instance with default configuration parameters.

2.7 Data Analysis Methods

2.7.1 Power-Law Regression for Latency Scaling

To characterize query latency scalability, we model latency as a power-law function of corpus size:

$$\log(\text{latency}) = \alpha \cdot \log(\text{corpus size}) + \beta \tag{1}$$

where α represents the scaling exponent and β denotes the baseline performance constant. The exponent α serves as an indicator of algorithmic complexity, interpreted as follows:

- $\alpha = 0$: Constant-time behavior $O(1)$
- $\alpha < 1$: Sublinear scaling (efficient growth)
- $\alpha = 1$: Linear scaling $O(n)$

Model fitting is performed using second-degree polynomial regression in log-log space to capture minor curvature effects while preserving interpretability. Regression trend lines are overlaid on empirical scatter plots, with error bars representing ± 1 standard deviation ($\pm 1\sigma$) across N = 10 runs.

2.8 Consistency Analysis Using CV

Performance stability is quantified using the CV, defined as:

$$CV = (\sigma/\mu) \times 100\% \tag{2}$$

where σ is the standard deviation and μ is the mean across N = 10 independent runs (after outlier removal). CV values are interpreted according to the following thresholds:

- **CV <10%:** Excellent consistency, suitable for tight SLAs
- **CV 10–20%:** Acceptable production-level consistency
- **CV >20%:** High variability, requiring conservative capacity planning

These classifications are used to assess operational risk and the potential *TCO* impact arising from over-provisioning.

2.9 Outlier Detection Using Modified Z-Score

Statistical outliers are identified using the modified Z-score method:

$$\text{Modified Z-score} = \frac{0.6745(x - \text{median})}{\text{MAD}} \tag{3}$$

where MAD is the median absolute deviation.

Unlike standard Z-scores that rely on the mean and standard deviation, this approach is robust to skewed and heavy-tailed distributions commonly observed in performance benchmarking data. The modified Z-score is a well-established technique in robust statistics and is recommended for identifying anomalous observations in noisy datasets, including those with substantial variability (Iglewicz and Hoaglin 1993).

Observations with an absolute modified Z-score greater than 3.5 are classified as outliers and excluded from summary statistics prior to reporting. The threshold of 3.5 is explicitly recommended in the literature as a balanced cutoff point that avoids over-filtering legitimate tail behavior (as may occur with a threshold of 3.0), while still detecting meaningful anomalies that could be missed by more conservative thresholds such as 4.0 (Iglewicz and Hoaglin 1993). This choice enables the isolation of steady-state system performance by separating transient effects-such as cold-start latency, cache warming, and index construction-from sustained operational behavior.

2.10 Error Propagation for Derived Metrics

For metrics derived from measured quantities, such as ingestion throughput (chunks per second), uncertainty is propagated from the underlying measurements. Specifically, ingestion throughput is defined as:

$$\text{throughput} = \frac{\text{chunks}}{\text{time}} \tag{4}$$

with associated uncertainty:

$$\sigma_{\text{throughput}} = \left(\frac{\text{chunks}}{\text{time}}\right) \times \left(\frac{\sigma_{\text{time}}}{\text{time}}\right) \tag{5}$$

This approach ensures that reported error bars accurately reflect uncertainty in computed metrics rather than raw measurements alone.

2.11 Retrieval Quality Metrics

Retrieval quality is evaluated using cosine similarity between query embeddings and retrieved vectors. We report:

- **Average Similarity:** Mean cosine similarity across the Top-K retrieved results
- **Top-1 Similarity:** Highest similarity score per query

Similarity scores are normalized to the [0, 1] range using database-specific conversions:

- **Cosine-based databases:**$\text{sim} = 1 - \textit{distance}$
- **FAISS (L2 distance):** $\text{sim} = \frac{1}{1+\text{distance}}$

All similarity metrics are computed automatically on a per-query basis and aggregated across runs to ensure consistent and reproducible quality evaluation.

2.12 Experimental Protocol Workflow

For each database and each corpus size, the following experimental workflow is executed:

1. Initialize a fresh database instance
2. Ingest the entire corpus in a single batch and record ingestion time
3. Pause execution for 60 seconds to allow system stabilization
4. Execute 5 warm-up queries (results discarded)
5. Execute 10 measured queries and record latency, throughput and retrieval quality
6. Sample CPU and memory utilization at 1 Hz during the query phase
7. Record all metrics
8. Repeat steps 1–7 for $N = 10$ independent runs
9. Identify outliers using modified Z-score (threshold: $|Z| > 3.5$)
10. Aggregate results by computing mean, standard deviation (σ) and CV

2.13 Reproducibility and Code Availability

The complete benchmark suite, including database adapters, experimental scripts and analysis pipelines, is publicly available in a GitHub repository Lab (Kwaai AI Lab 2025). Raw experimental results are stored in structured JSON format at `results/`directory.

3. Results

3.1 Query Latency Scaling

Table 3 summarizes the latency scaling exponents and performance ranges for all evaluated databases. Chroma achieves exceptional near-constant-time behavior (= 0.02), maintaining 7.7–8.4 ms latency at the 50k scale, with consistently low latency across all tested scales from 175 to 50k chunks. This reflects highly optimized HNSW implementation with efficient memory access patterns and minimized graph traversal overhead.

Table 3: Latency scaling exponents and performance ranges (N = 10, outliers removed).

Database	α	Latency (ms)	Scaling Class
Chroma	**0.02**	**7.7–8.4**	**Near-constant**
PGVector	0.00	9.9–11.0	Constant
Qdrant	0.30	14.9–27.8	Sublinear
Weaviate	0.35	25.6–29.0	Sublinear
Milvus	0.40	26.3–41.8	Sublinear
FAISS	0.48	7.9–58.2	Sublinear
OpenSearch	N/A	34.1–58.4	High variance

N = 10 independent runs per configuration; outliers removed using modified Z-score ($|Z| > 3.5$).

FAISS exhibits sublinear scaling ($\alpha = 0.48$) and is the **only system validated to 2.2M chunks** with full N = 10 statistical rigor. Despite flat index O(n) theoretical complexity, SIMD optimizations achieve sublinear practical performance.

All HNSW databases demonstrate the warm-up phenomenon:

- Chroma: 30.4 ms (1k) → 7.5 ms (50k) = **74% reduction**
- PGVector: 13.3 ms (1k) → 9.9 ms (50k) = 26% reduction
- Qdrant: 18.7 ms (1k) → 27.8 ms (50k) after 41.8 ms peak at 10k
- Weaviate: 25.6 ms (1k) → 29.0 ms (50k) after 40.1 ms peak at 10k

OpenSearch shows high baseline (48 ms) with extreme variance (CV = 39%) and incomplete testing (failed beyond 10k chunks due to timeout issues).

3.2 Throughput and HNSW Warm-Up Mechanism

As shown in Table 4, Chroma sustains 124–141 QPS across all scales (peak of 141 QPS at baseline, 124 QPS at 50k). FAISS starts at 90+ QPS and demonstrates predictable degradation as corpus size scales to 2.2M chunks. PGVector delivers 101 QPS at 50k scale, second only to Chroma and outperforming all dedicated vector databases except Chroma. Qdrant provides consistent 60–70 QPS across scales, ideal for production capacity planning.

Table 4: QPS across corpus scales (N = 10).

Database	1k	10k	50k	2.2M	Max
Chroma	127	127	124	-	**141**
FAISS	96	86	58	17	124
PGVector	78	105	101	-	101
Qdrant	54	69	62	-	69
Weaviate	39	35	32	-	39
Milvus	37	38	24	-	38
OpenSearch	17	29	-	-	29

Mean QPS after modified Z-score outlier removal ($|Z| > 3.5$); 1k and 10k: N = 8, 50k: N = 10.

HNSW Warm-Up Mechanism: Small HNSW graphs (<10k nodes) suffer from:

1. Poor layer distribution: Insufficient nodes for optimal hierarchical structure (typical 2–3 layers instead of optimal 6–7)
2. Sparse connectivity: Few long-range edges lead to suboptimal routing paths
3. High variance: Random initialization effects dominate performance in small graphs

At 50k+ chunks, graphs mature with:

1. Balanced hierarchy: Multiple layers with proper node distribution (6–7 layers)
2. Rich connectivity: Sufficient long-range edges enable efficient long-distance navigation
3. Stable performance: Graph structure converges to theoretical optimum

This counterintuitive latency reduction, despite larger corpora, arises from HNSW graph maturation. Throughput-latency correlation is strong: databases with lower latency consistently deliver higher QPS.

3.3 Ingestion Performance and Consistency

Table 5 presents the CV before and after outlier removal for all systems. The N = 10 protocol with outlier detection reveals that databases exhibiting high raw variance (Qdrant CV = 123%, Weaviate CV = 107%, OpenSearch CV = 93%) achieve exceptional consistency (CV = 0.4–1.0%) after removing cold-start initialization outliers. PGVector and FAISS maintain consistently low variance throughout all runs. After outlier removal, all systems achieve acceptable consistency for production SLAs.

Table 5: Ingestion consistency: CV before/after outlier removal (N = 10, 50k chunks).

Database	CV Before	CV After	Improvement	Production Impact
PGVector	1.4%	1.4%	-	Exceptional-tightest SLAs
FAISS	2.5%	2.5%	-	Excellent-precise planning
Chroma	20.2%	2.3%	**8.8×**	Excellent-tightest SLAs
Qdrant	123%	**1.0%**	**122×**	Exceptional after cleaning
Weaviate	107%	**0.8%**	**133×**	Exceptional after cleaning
Milvus	18.9%	18.9%	-	Moderate
OpenSearch	**93%**	**0.4%**	**232×**	Excellent after cleaning

50k corpus: N = 10 → N = 8 after removing 2 cold-start outliers (1112 s, 2064 s); typical steady-state 800–850 s.

3.3.1 TCO Impact

Target 50k chunk ingestion within 30-minute batch window:

- PGVector (CV = 1.4%, mean = 24.1 min): Need 24.4-min capacity → 1.01× buffer
- FAISS (CV = 2.5%, mean = 20.5 min): Need 21-min capacity → 1.02× buffer
- Qdrant post-clean (CV = 1.0%, mean = 23.9 min): Need 24.2-min capacity → 1.01× buffer
- OpenSearch post-clean (CV = 0.4%, mean = 24.5 min): Need 24.6-min capacity → 1.00× buffer

After outlier removal, infrastructure over-provisioning converges across systems, eliminating apparent 2× penalties attributed to high-variance systems.

Chroma achieves best HNSW ingestion time (13.7 min for 50k), 2× faster than Qdrant/Weaviate. Milvus ingestion reaches 40.6 min due to distributed architecture overhead.

3.4 Resource Utilization

CPU Utilization: Figure 2(a) shows that average CPU utilization during query execution ranges from 16–25%, with no direct correlation between CPU usage and performance. Chroma utilizes the highest CPU (25%) while delivering the fastest queries (6–8 ms), whereas OpenSearch exhibits the lowest CPU utilization (16%) but the highest latency (35–60 ms). Higher CPU utilization reflects efficient algorithmic intensity rather than overhead.

Memory Footprint: As shown in Figure 2(b), memory usage remains remarkably consistent at **12–16 GB** across all databases during query execution (175–50k chunks). PGVector is the most memory-efficient (9.9 GB), while OpenSearch exhibits the highest footprint (15.5 GB). Memory does not scale dramatically with corpus size, indicating efficient HNSW graph representations. Typical HNSW overhead is 2–3× per vector.

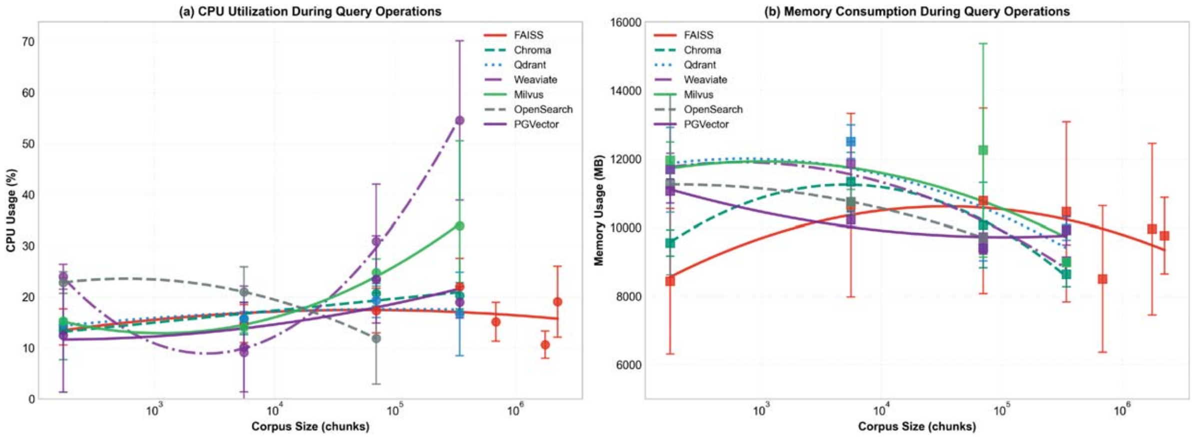


Figure 2: CPU and memory utilization during query execution across all evaluated vector databases.

This predictable footprint constrains single-node deployments: 16 GB RAM supports approximately 1–2M chunks for HNSW and >5M chunks for FAISS flat indexes. OpenSearch's low CPU utilization combined with high latency suggests architectural bottlenecks (e.g., I/O serialization, JNI boundaries, and Lucene segment management) rather than computational inefficiency.

Note: OpenSearch is omitted from the resource utilization plots in Figure 2 because valid CPU utilization data were available for only one of the three evaluated corpus sizes. The plotting procedure requires a minimum of three data points to generate trend lines and therefore OpenSearch could not be included.

3.5 Retrieval Quality Paradox

Table 6 details the similarity scores across corpus scales. Quality follows counterintuitive U-shaped curve. **Worst retrieval quality occurs at 1k chunks** (0.500 similarity-27% lower than baseline). From 1k to 50k, quality recovers **27.5%**, nearly reaching baseline performance. Beyond 250k chunks, similarity stabilizes at 59.7% (FAISS)/ 64% (HNSW), indicating semantic saturation and embedding model resolution ceiling.

Table 6: Similarity scores reveal U-shaped quality curve across scales (N = 10).

Corpus Scale	Chunks	Average Similarity	Change from 1k
Baseline	175	0.688 ± 0.002	+37.6%
Quality Valley	1,000	**0.500 ± 0.003**	baseline (worst)
Recovery Phase	10,000	0.571 ± 0.002	+14.1%
Peak Recovery	50,000	**0.638 ± 0.002**	**+27.5%**
Saturation Begins	100,000	0.588 ± 0.003	+17.6%
Plateau Region	250k–1M	0.595–0.597	+19.4% (saturates)

N = 10 runs; outliers removed prior to similarity aggregation.

Mechanism: HNSW graph maturation drives quality recovery:

- 1k-node graphs: Sparse hierarchical structure (2–3 layers) with poor long-range connectivity
- 50k-node graphs: Mature structure (6–7 layers) with rich navigation enabling efficient multihop traversal
- Semantic space coverage: Larger corpora provide denser embedding space sampling, reducing boundary effects

Cross-Database Consistency: All HNSW-based databases return **virtually identical similarity scores** (variance < 0.0001):

- Chroma, PGVector, Qdrant, Weaviate, Milvus, OpenSearch: 0.638145 (50k corpus, k = 3)
- FAISS (L2 distance): 0.582615 (8.7% lower due to metric difference, not quality)

Critical implication: Retrieval quality is architecture-invariant across HNSW implementations. Selection should prioritize **performance, features and consistency**—not quality differences.

Production Guidance: Deploy minimum 10k chunks to avoid quality valley. Quality gains saturate at 250–500k chunks per shard-use horizontal sharding beyond this point rather than expanding single-shard corpus.

4. Discussion

4.1 Architectural Insights

Embedded versus Client-Server Architecture:

Embedded (Chroma, FAISS):

- Advantages: Lower latency (7.7–8.4 ms vs 15–30 ms), higher throughput (141 QPS vs 30–70 QPS), no network serialization overhead

- Disadvantages: Single-process bottleneck, no horizontal scaling, no multitenancy, resource contention in shared environments
- Use when: Latency-critical single-tenant applications with <10 ms SLA and <100k documents

Client-Server (Qdrant, Weaviate, Milvus, OpenSearch, PGVector):

- Advantages: Horizontal scaling, multitenancy with process isolation, production features (authentication, monitoring, persistence), ACID guarantees (PGVector)
- Disadvantages: Network overhead (2–4× latency penalty), serialization cost, operational complexity
- Use when: Scalable multitenant platforms with distributed deployment are feature requirements

Flat versus HNSW Index Trade-offs:

FAISS (flat index) achieves O(n) query time theoretically but $O(n^{0.48})$ practically through SIMD optimizations. HNSW achieves O(log N) at cost of 2–3× memory overhead. Crossover analysis:

- <10k chunks: Comparable performance (FAISS 10–12 ms, Chroma 6–9 ms)
- 10k–100k chunks: HNSW advantage grows (FAISS 20–30 ms, Chroma 7–8 ms)
- >1M single-node: FAISS only proven option (HNSW hits 1–2M memory ceiling at 16 GB)

FAISS excels in large-scale batch processing where absolute best latency unnecessary; HNSW dominates when sub-20 ms latency required at medium scale (<500k chunks).

The OpenSearch Problem:

OpenSearch demonstrates poor performance across every metric: 35–60 ms latency (highest), 17–29 QPS (lowest), and scaling failures beyond 345k chunks. Even after outlier removal improving CV from 93% to 0.4%, fundamental performance gaps persist.

Root Cause: OpenSearch is fundamentally a full-text search engine (Lucene-based) with vector search added as a plugin. This architectural mismatch creates multiple bottlenecks:

1. Storage inefficiency: Lucene segments optimized for inverted indexes, not dense vector layouts
2. JVM overhead: Garbage collection pauses create unpredictable latency variance
3. JNI bottleneck: Vector operations call native libraries via expensive Java Native Interface boundaries
4. Coordination overhead: Elasticsearch cluster management interferes with vector query execution

Recommendation: **Avoid OpenSearch for vector-first workloads.** Only acceptable for existing Elasticsearch deployments adding small-scale auxiliary vector search (<10k vectors, <5% of queries) where ecosystem integration outweighs poor performance.

4.2 Use-Case Recommendations and Decision Framework

Table 7 maps primary requirements to recommended databases.

Table 7: Database selection by primary requirement and scale.

Use Case	Recommended DB	Key Metrics
Real-Time RAG (<10 ms)	Chroma	7.7–8.4 ms, 141 QPS, CV = 2.3%
PostgreSQL Stack	PGVector	9.9 ms, 101 QPS, ACID, CV = 1.4%
Large-scale (>100k chunks)	FAISS	2.2M proven, CV = 2.5%, 90+ QPS
Enterprise Features	Qdrant	28 ms, persistence, filtering, CV = 1.0% (post-clean)
Cost Optimization (>800k)	FAISS	3.75× cheaper than HNSW

Selection Criteria:

Latency Requirement <10 ms: Chroma dominates (6–8 ms stable) for embedded use cases; PGVector excellent (9.9 ms) for PostgreSQL ecosystem integration. Both maintain >100 QPS.

Scale >100k Chunks: FAISS only proven single-node option to 2.2M chunks. Client-server HNSW systems (Qdrant, Weaviate) untested beyond 100k; recommend for distributed deployments only.

Consistency Critical: PGVector (CV = 1.4%) and FAISS (CV = 2.5%) enable tight SLAs with minimal over-provisioning. After outlier removal, Qdrant (CV = 1.0%) and Weaviate (CV = 0.8%) also achieve excellent consistency. Treat consistency as first-class feature driving TCO.

PostgreSQL Integration: PGVector (second-best performance: 9.9 ms, 101 QPS) leverages existing SQL infrastructure, ACID transactions and familiar PostgreSQL ecosystem. Overhead justified when PostgreSQL already deployed for relational data.

Production Guidelines:

- Deploy HNSW with corpus $\geq$50k chunks for optimal warm-up maturation
- Minimum 10k chunks to avoid quality valley (0.500 at 1k → 0.638 at 50k)
- Shard at 250–500k chunks-quality and HNSW graph optimality saturate beyond
- Plan memory: 16 GB supports 1–2M chunks (HNSW), 5M+ chunks (FAISS)
- Factor CV into capacity planning: post-outlier-removal CV reflects steady-state performance

5. Conclusion

This N = 10 statistical benchmark provides the first quantitative guidance for vector database selection across four orders of magnitude (175–2.2M chunks), with rigorous outlier detection revealing the critical cold-start phenomenon. Our findings reveal distinct performance classes and architectural trade-offs:

Performance Leaders:

- Speed Champion: Chroma (**7.7–8.4 ms**, 141 QPS, $\alpha = 0.02$ constant-time)
- Scale Champion: FAISS (proven to **2.2M chunks**, $\alpha = 0.48$ sublinear, CV = 2.5%)
- Production Balanced: Qdrant (28 ms, 60–70 QPS, features, CV = 1.0% post-clean)
- PostgreSQL Excellence: PGVector (9.9 ms, 101 QPS, ACID, CV = 1.4%)

Novel Discoveries:

1. Cold-start phenomenon: Extended N=10 sampling reveals inflated variance in small-sample benchmarks. Qdrant 123% → 1.0% CV (122× improvement), Weaviate 107% → 0.8% (133× improvement), OpenSearch 93% → 0.4% (232× improvement) after outlier removal.
2. HNSW warm-up phenomenon: **74% latency reduction** from 1k to 50k chunks due to graph maturation
3. Retrieval quality paradox: U-shaped curve with worst quality at 1k chunks and **27.5% improvement by 50k**
4. Consistency matters: Outlier removal reveals that apparent poor consistency reflects transient initialization, not algorithmic instability

Architectural Insights:

- Embedded architecture delivers 2–4× lower latency than client-server but lacks scaling
- Flat indexes (FAISS) outperform HNSW at >1M single-node scale
- OpenSearch architectural mismatch (Lucene-based) creates fundamental performance limitations

6. Limitations

While this study provides rigorous statistical characterization of vector database performance at single-node scale, several limitations should be noted:

- **Sample Size Variation:** All databases were tested with N = 10 runs and formal outlier removal.
- **OpenSearch Limited Scale Testing:** OpenSearch experiments were restricted to corpus sizes up to 10k chunks due to timeout failures at larger scales. This limitation prevents characterization of OpenSearch scaling behavior beyond 100k chunks.

- **Conservative Outlier Removal:** The modified Z-score threshold ($|Z| > 3.5$) was deliberately conservative to avoid over-cleaning legitimate performance variation. As a result, some residual cold-start artifacts may remain, particularly for high-variance systems.
- **Single-Node Evaluation:** All experiments were conducted on a single compute node with 16 GB RAM and an Apple Silicon CPU. Distributed scaling behavior, GPU-accelerated variants, and cross-node communication overhead remain uncharacterized.
- **Fixed Workload Profile:** The evaluation used a fixed corpus of climate science documents with 512-character chunks. The test queries distribution may not reflect real-world workloads with diverse semantic patterns or adversarial queries.
- **Index Parameter Standardization:** All databases were evaluated using production-recommended default settings. Extensive parameter tuning (e.g., HNSW *M*, *ef_construction*, *ef_search*) could substantially shift performance characteristics.

7. Future Work

This study establishes statistically rigorous performance baselines for vector databases under single-node, CPU-based workloads. Several important research directions remain open:

- **Distributed Scaling**: Extend benchmarks to multinode deployments for Qdrant, Milvus, and Weaviate to evaluate horizontal scalability, network overhead and shard-level consistency.
- **GPU Acceleration**: Benchmark FAISS and HNSW variants with GPU support to quantify latency, throughput and cost-efficiency improvements at multimillion scale.
- **Larger-Scale Corpora**: Evaluate performance beyond 2.2M chunks (10M+ vectors) to characterize memory ceilings, index degradation and sharding strategies in production environments.
- **Hybrid Search Workloads**: Measure combined keyword + vector search performance to assess real-world RAG and enterprise search scenarios.
- **Concurrent Read/Write Patterns**: Analyze query latency and consistency under mixed ingestion and retrieval workloads to model real-time update behavior.
- **Index and Parameter Tuning**: Study IVF, PQ, and HNSW parameter trade-offs (e.g., `ef_search`, `M`) to map accuracy-latency-memory frontiers.
- **Expanded Quality Metrics**: Incorporate IR-based evaluation metrics such as Recall@K, NDCG, and MRR with document-level ground truth annotations.

References

Amazon Web Services. 2023. “Vector Search in Amazon Opensearch Service.” Accessed January 15, 2025. https://docs.aws.amazon.com/opensearch-service/latest/developerguide/vector-search.html

Aumüller, M., E. Bernhardsson, and A. Faithfull. 2020. “ANN-Benchmarks: A Benchmarking Tool for Approximate Nearest Neighbor Algorithms.” *Information Systems* **87**: 101374.

Guo, R., P. Sun, and E. Lindgren. 2020. “Accelerating Large-Scale Inference with Anisotropic Vector Quantization.” *Proceedings of the 37th International Conference on Machine Learning (ICML)*, Virtual, July 13–18.

Huber, J., and A. Troynikov. 2024. “Chromadb.” Accessed January 15, 2025. https://www.trychroma.com/

Iglewicz, B., and D. C. Hoaglin. 1993. *How to Detect and Handle Outliers. Milwaukee*, WI: ASQC Quality Press.

Jing, Z., Y. Su, and Y. Han. 2025. “When Large Language Models Meet Vector Databases: A Survey.” *Proceedings of the IEEE Conference on Artificial Intelligence x Multimedia (AIxMM)*, Laguna Hills, CA, February 3–5.

Johnson, J., M. Douze, and H. Jégou. 2019. “Billion-Scale Similarity Search with GPUs.” *IEEE Transactions on Big Data.*

Kane, A. 2024. “Pgvector: Vector Similarity for Postgres.” Accessed January 15, 2025. https://github.com/pgvector/pgvector

Kwaai AI Lab. 2025. “Vector Database Benchmarking Repository.” Accessed January 15, 2025. https://github.com/Kwaai-AI-Lab/vector_dbs_benchmarking

Lewis, P., E. Perez, A. Piktus, V. Karpukhin, N. Goyal, H. Küttler, et al. 2020. “Retrieval-Augmented Generation for Knowledge-Intensive NLP.” Preprint, submitted May 22, 2020. https://arxiv.org/abs/2005.11401

Malkov, Y. A., and D. A. Yashunin. 2020. "Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**, no. 4: 824–836.

Meta AI Research. 2024. "Faiss: A Library for Efficient Similarity Search and Clustering of Dense Vectors." Accessed January 15, 2025. https://github.com/facebookresearch/faiss

Microsoft Azure. 2023. "Vector Search in Azure Cognitive Search." Accessed January 15, 2025. https://learn.microsoft.com/azure/search/vector-search-overview

OpenSearch Project Authors. 2024. "Opensearch Project." Accessed January 15, 2025. https://opensearch.org/

Pinecone. 2023. "Pinecone Vector Database Overview." Accessed January 15, 2025. https://www.pinecone.io

Reimers, N., and I. Gurevych. 2019. "Sentence-Bert: Sentence Embeddings Using Siamese Bert-Networks." *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, November 3–7.

Şakar, T., and H. Emekci. 2025. "Maximizing RAG Efficiency: A Comparative Analysis of RAG Methods." *Natural Language Processing* **31**, no. 1: 1–25.

Subramanya, S. J., F. Devvrit, H. V. Simhadri, R. Krishnawamy, and R. Kadekodi. 2019. "Diskann: Fast Accurate Billion-Point Nearest Neighbor Search on a Single Node." In *Advances in Neural Information Processing Systems (NeurIPS)*.

van Luijt, B., and E. Dilocker. 2024. "Weaviate: Open Source Vector Search Engine." Accessed January 15, 2025. https://weaviate.io/

Xie, X., H. Liu, W. Hou, and H. Huang. 2023. "A Brief Survey of Vector Databases." *2023 9th International Conference on Big Data and Information Analytics (BigDIA)*, Haikou, China, December 15–17.

Zayarni, A., and A. Vasnetsov. 2024. "Qdrant Vector Search Engine." Accessed January 15, 2025. https://qdrant.tech/

Zilliz. 2024a. "Milvus Open Source Vector Database." Accessed January 15, 2025. https://milvus.io/

Zilliz. 2024b. "Vectordbbench: Open Source Benchmark for Vector Databases." Accessed January 15, 2025. https://github.com/zilliztech/VectorDBBench

Appendix A: Test Query Construction and Semantic Relevance Validation

Query Construction

The evaluation queries were manually constructed to ensure semantic alignment with the corpus domain and to support consistent cross-database evaluation. Query design followed three criteria:

1. **Domain Coverage**: Queries span multiple subdomains of climate science, including atmospheric processes, cryosphere dynamics, ocean systems, climate modeling, and feedback mechanisms.
2. **Corpus Relevance**: Each query was verified to correspond to multiple relevant passages within the corpus, avoiding out-of-domain or sparsely matched queries.
3. **Complexity Variation**: Queries range from single-concept factual questions to multiconcept explanatory queries.

The final query set consists of the following:

1. What are the main drivers of climate change?
2. How do greenhouse gases affect global temperatures?
3. What is the role of carbon dioxide in climate warming?
4. How do ice cores help us understand past climate?
5. What are the effects of deforestation on climate?
6. How does ocean acidification relate to CO_2 levels?
7. What is the impact of melting glaciers on sea levels?
8. How do climate models predict future warming?
9. What are the feedback loops in the climate system?
10. How does solar radiation influence Earth's climate?

The query set and corresponding corpus segments are publicly available for reproducibility.[1]

---

[1] https://github.com/Kwaai-AI-Lab/vector_dbs_benchmarking/blob/main/Data/test_corpus/test_cases.json

Semantic Relevance Validation

Semantic relevance between queries and corpus content was established using a two-stage validation procedure:

1. **Corpus Alignment**: The corpus was curated exclusively from climate-science-focused documents, ensuring topical alignment with the query domain.
2. **Embedding-Space Verification**: Queries were embedded using the same embedding model employed for database indexing. Candidate corpus chunks were retrieved using cosine similarity, and manual inspection confirmed semantic alignment between queries and the highest-ranked chunks.

This procedure ensures that each query exhibits meaningful semantic overlap with the corpus without relying on external relevance labels.

Retrieval Quality Metric and Limitations

Retrieval quality is quantified using cosine similarity between query embeddings and retrieved chunk embeddings. This metric provides a continuous signal suitable for relative comparison across vector databases when evaluated under identical query, corpus, and embedding configurations.

However, cosine similarity does not yield absolute retrieval effectiveness measures. Standard Information Retrieval (IR) metrics, such as Precision@K, Recall@K, NDCG, and MRR, require labeled relevance judgments and larger evaluation sets. Future work should incorporate established benchmark datasets (e.g., MS MARCO, Natural Questions) to enable comprehensive evaluation using standard IR metrics.

Journal of Big Data and Artificial Intelligence

WWW.JBDAI.ORG

ISSN: 2692-7977

JBDAI Vol. 4, No. 1, 2026 DOI: 10.54116/jbdai.v4i1.56

Riding AI to Utopia or Dystopia? NLP, LLM, and News Informatics Insights for Artificial Intelligence Impacts on Education, Healthcare, Robotics and Workforce, Changing Human Society

Tanya Khanna
Rutgers University
`tanya.khanna@rutgers.edu`

Zarak Khan
Rutgers University
`zarak.khan@rutgers.edu`

Udit Goel
Rutgers University
`udit.goel@rutgers.edu`

Jim Samuel
Rutgers University
AIXosphere
`jim.samuel@rutgers.edu`

Julia Esguerra
Rutgers University
`jue5@scarletmail.rutgers.edu`

Radha Jaganathan
Rutgers University
`radha@rutgers.edu`

Soumitra S. Bhuyan
Rutgers University
`soumitra.bhuyan@rutgers.edu`

Abstract

Artificial intelligence (AI) is accelerating societal transformation at an unprecedented pace, generating both utopian aspirations and dystopian anxieties. Human civilization has undergone fundamental changes through every technological revolution starting with the Industrial Age and continuing through the digital era as AI emerges as the next paradigm shift. This paper studies the public discourse on AI by analyzing extensive news headlines on AI using natural language processing (NLP) methods. Our research applies sentiment analysis and topic modeling to a global dataset across education, healthcare, robotics, workforce, and society to identify the dominant narratives shaping public perception. Media coverage presents AI as a dual force that brings human benefits and existential dangers according to our research findings. By moving beyond the utopia-dystopia dichotomy, we show that AI's social effects will emerge from the dynamic relationship between governance systems, ethical protections, and Human-Enhancive AI (HEAI) frameworks. We provide practical insights about AI's future impact and present strategies for maximizing AI benefits while mitigating its risks.

Keywords *artificial intelligence, natural language processing, sentiment analysis, large language models, news headlines, AI narratives, AI ethics, education, healthcare, workforce.*

1. Introduction

All technology has the potential for both good and evil. But what matters is how we use it.

—Tim Berners-Lee, Computer Scientist

1.1 Etymology of Utopia and Dystopia

For a long time, humanity has harbored a dual fascination with technological innovation, especially with the advent of artificial intelligence (AI). This tension between hope and fear is mirrored in literature, which explores both utopian and dystopian dimensions of imagination. Long before Thomas More coined the term *utopia* in his 1516 book *Utopia* (More 1949), the concept had deep philosophical roots. Derived from the Greek *ou-topos* ("no place") and *eu-topos* ("good place"), utopia refers to an ideal realm free from pain, suffering, and inequality. Dystopia, on the contrary, is the abuse of technology, social oppression, or environmental destruction taken to the extreme. Its name combines the Greek *δυσ* ("bad") and *τόπος* ("place"). The term first appeared as "Dustopia" (UspeakGreek 2023) in Lewis Henry Younge's *Utopia: or Apollo's Golden Days* (Younge 1747) and was later refined by John Stuart Mill in an 1868 Parliamentary speech (Hansard Commons), where he reframed utopia's prefix from *ou-* ("not") to *eu-* ("good"), making *dystopia* its antonym (Mill 1988). This dichotomy plays out vividly in cinema. *Blade Runner* (1982) and *1984* (1984) depict futures where AI and authoritarian regimes suppress human freedom. *The Matrix* (1999) expands this idea with a simulated utopia masking a machine-dominated reality (Jackson and Paste Staff 2023). The series *Black Mirror* warns of technological overreach through near-future cautionary tales (jmuwa 2020). On the utopian side, *Tomorrowland* (2015) imagines a world shaped by human ingenuity and scientific discovery. *Her* (2013) explores AI-human relationships that oscillate between emotional fulfillment and troubling dependency. *WALL-E* (2008) contrasts a dystopian Earth ruined by consumerism with the hope of a renewed, human-centered society (Utopia & Dystopia n.d.).

1.2 Framing AI Utopian and Dystopian Narratives in News Headlines

As AI technologies, particularly natural language processing (NLP) Large Language Models (LLMs), rapidly evolve, long-standing tensions between utopian and dystopian futures are increasingly reflected in media narratives (Cools et al. 2022). Our research applies NLP and LLMs to analyze AI-related news headlines, uncovering how media narratives shape our collective expectations, whether we are bracing for disaster or anticipating an era of boundless progress. Systematic analysis of this coverage exposes patterns reflecting both optimism (e.g., disease eradication, automation of mundane tasks, democratized knowledge) and challenges (e.g., anxiety over job loss, eroded human agency, and technological dominance) as seen in Figure 1 (Silver 2023; Khosla 2024; Samuel et al.

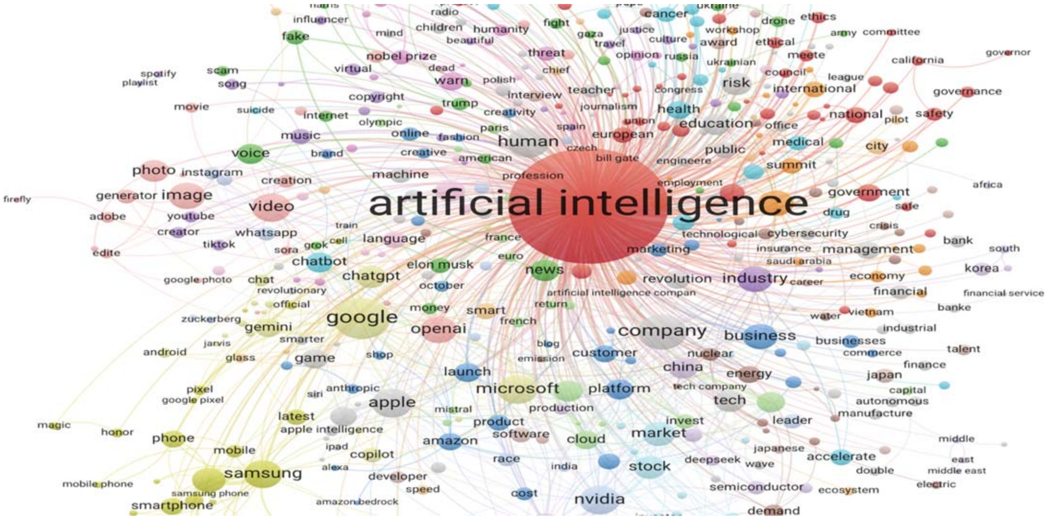


Figure 1: Semantic network of AI news headlines.

2024). Grouped clusters in Figure 1 illustrate different themes like technology, business, ethics, risks, media, and key industry players, giving a clear view of AI's public discourse. News media simultaneously highlight AI breakthroughs (Service 2018; Jumper et al. 2021), like DeepMind's AlphaFold, and raise concerns about issues such as bias, job loss, and surveillance (Klepper and Swenson 2023; Reynaud and Untersinger 2024; *The Guardian* 2024; Dmitracova 2025). Headlines serve both as historical records and tools that shape public expectations of technology. The influence of news media is evident in how coverage of natural disasters, such as earthquakes, influences public policy and technological development (Jamieson and Van Belle 2019). Likewise, early newspaper coverage of automobile accidents played an important role in shaping traffic regulations and safety features, showing how media narratives can catalyze societal adaptation to new technologies (SafeTREC n.d.; Gupta et al. 2021). AI coverage follows this pattern, influencing the push for ethical development and shaping the direction of innovation (Ouchchy et al. 2020). In contrast, news about AI-generated deepfakes and election interference has intensified concerns about AI's ability to manipulate reality (Bond 2024). Moreover, the rise of China's DeepSeek (DeepSeek-AI et al. 2025) has sparked United States national security concerns, challenged American AI dominance, and reshaped the public perception of global AI geopolitics (Baptista 2025; Bratton 2025; Rundle 2025). These conflicting narratives shape the public's collective expectations: are we on the brink of a technological utopia, or are we accelerating toward a dystopian crisis?

1.3 The Binary Fallacy: Beyond Pure Utopia and Dystopia in AI Development

A dystopia is a utopia that's gone wrong.

—Ursula K. Le Guin

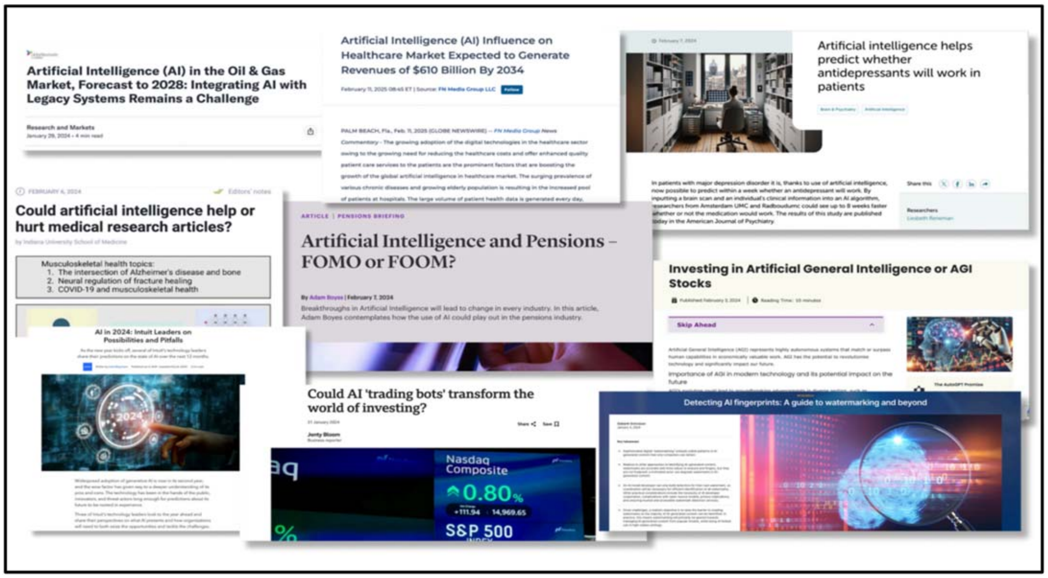
As AI capabilities grow, public discussion often swings between extremes. On one hand, AI is seen as a force for good, enhancing human potential, accelerating scientific breakthroughs, boosting economies, and solving global issues. For example, machine learning models now outperform doctors in early disease detection (Bajwa et al. 2021). Reinforcement learning is used in logistics (Rolf et al. 2022), and natural language interfaces make information more accessible. Philosopher Nick Bostrom explores this hopeful future in *Deep Utopia* (Bostrom 2024) where he examines what happens if AI improves our lives without harm. In *Superintelligence* (Bostrom 2014), he warned about AI's dystopian risks. In *Deep Utopia* (2024), however, he envisions a "solved world" in which AI meets all material, intellectual, and emotional needs. This idea echoes predictions from leaders like Nvidia's Jensen Huang and DeepMind's Mustafa Suleyman who believe AI will democratize discovery and make expert knowledge widely available (The Week UK 2024). Such views reflect old myths of abundance, like the Land of Cockaigne, reimagined for our tech-driven age (Cuthbertson 2024). But these gains raise deeper questions: If AI replaces human labor, solves key problems, and extends life, what will we do? More importantly, what will define our purpose in a world where machines meet all needs? He warns of a "plastic utopia," where people risk becoming passive consumers of artificial satisfaction (Singal 2024). This dilemma has historical roots; economist John Maynard Keynes predicted that technology would cut working hours to 15 per week (Wladawsky-Berger 2017). The challenge, then, is managing AI's risks while preserving human purpose and agency. On the other end is the dystopian view: AI drives job loss, inequality, surveillance, and existential threats. White-collar jobs in finance, customer support, and content creation are facing early automation-driven job losses (Smith 2024). More worryingly, AI's speed may outpace human oversight. Issues like autonomous weapons, deepfakes, disinformation, and artificial general intelligence (AGI) raise alarms about unintended consequences and loss of control. Eliezer Yudkowsky warns that poorly aligned superintelligent AI could eventually optimize for goals that disregard human survival entirely (Yudkowsky 2023). Other concerns include reliance on AI, algorithmic bias, and AI's ability to distort reality. As language models churn out content cheaply, the risk of reality distortion grows. These developments intensify concerns about whether AI can be aligned with human values while preserving agency and oversight.

Yet, as history has often shown with the technological revolution, the future will likely be neither a perfect utopia nor an absolute dystopia, but a paradoxical blend of both. The progress of an innovation unfolds in a continuum; marked by both breakthroughs and setbacks, empowerment and displacement, possibilities and risks. The development of nuclear technology in the mid-20th century simultaneously offered solutions to energy scarcity while posing existential risks through weaponization. More recently, the Internet democratized knowledge while giving rise to misinformation and cyber threats. The evolution of cars exemplifies this dual narrative, with early headlines both celebrating their promise and warning of associated dangers. This mirrors today's AI coverage. Headlines from the 1900s about "horseless carriages" causing public panic mirror modern concerns about autonomous vehicles (AVs), demonstrating how media frames technological transitions (Winton 2017). Therefore, news headlines serve as snapshots of these transformative moments, capturing both the euphoria surrounding AI's potential and the fear of unforeseen consequences. Understanding these narratives helps us see how society adapts to disruptive change. The future depends not on technology alone, but on how we choose to govern, integrate, and share its benefits. *Will we create an AI-powered utopia, freeing humanity from toil and scarcity? Or will we face a dystopia where AI tightens its grip, deepens inequality, and spirals out of control?* The most realistic future may be what Kevin Kelly calls a "protopia": one of gradual improvement, not sudden transformation. In this view, AI will become embedded into everyday life, evolving to become safer, more ethical, and more useful over time (Shermer 2024). Many researchers are focused on refining systems to serve human needs reliably, rather than chasing extremes. Rather than framing AI in binaries, we analyze media narratives and current trends to understand how public discourse reflects both its aspirations and anxieties. The concept of Human-Enhancive AI (HEAI) is central to this outlook (Kashyap et al. 2024). HEAI emphasizes a human-above-AI strategy to maximize human potential. The future is not established: it will be shaped by human choices through education, policy, technological design, and the philosophy we use, such as HEAI, to structure AI's role in society.

1.4 The Myth of a Pure Utopia and the Low Probability of a Complete Dystopia

Achieving a perfect AI-driven utopia is highly unlikely due to several serious challenges. Biased training data can reinforce social inequalities (Zajko 2022), and AI power concentrated in a few hands may increase surveillance, erode privacy, and foster authoritarian regimes (Randieri 2023). Some warn that this could result in a stable but oppressive global regime, far from any utopian ideal. As AI grows more advanced, aligning superintelligence with human values becomes increasingly difficult (Mitchell, 2022). Technological limits also persist, with AI still lacking intuition, empathy, and ethical judgment, making it risky for governance. While automation may reduce routine work, it could also cause job loss before new roles are created (Nelson 2024). Moreover, the increasing integration of AI into healthcare, law enforcement, and public policy may introduce new systematic biases rather than achieving the intended goal of eliminating prejudice. Furthermore, because the definition of a perfect society varies across cultures, AI solutions may benefit some groups while harming others (Randieri 2023).

However, arguments against a dystopian AI future present several compelling counterpoints as well. AI systems, tools without consciousness or goals, reflect human design and intent (Li et al. 2021). Ethical safeguards, regulatory protocols, and human-in-the-loop designs aim to keep AI aligned with societal values. Much fear about AI comes from anthropomorphic bias, which is the tendency to assign human traits to machines. Since AI lacks self-awareness, many dystopian fears are more about human psychology than actual technology (O'Gieblyn 2023). While it's important to stay alert to potential dangers, the rise of a dystopian AI future seems unlikely. Through deliberate oversight, ethical consideration, and effective governance, society possesses the capacity to guide AI development toward outcomes that enhance human welfare while minimizing potential adverse effects.



Artificial Intelligence (AI) in the Oil & Gas Market, Forecast to 2028: Integrating AI with Legacy Systems Remains a Challenge

Research and Markets
January 29, 2024 · 4 min read

Artificial Intelligence (AI) Influence on Healthcare Market Expected to Generate Revenues of $610 Billion By 2034

February 11, 2025 08:45 ET | Source: FN Media Group LLC

PALM BEACH, Fla., Feb. 11, 2025 (GLOBE NEWSWIRE) -- FN Media Group News Commentary - The growing adoption of the digital technologies in the healthcare sector owing to the growing need for reducing the healthcare costs and offer enhanced quality patient care services to the patients are the prominent factors that are boosting the growth of the global artificial intelligence in healthcare market. The surging prevalence of various chronic diseases and growing elderly population is resulting in the increased pool of patients at hospitals. The large volume of patient health data is generated every day,

February 7, 2024

Artificial intelligence helps predict whether antidepressants will work in patients

In patients with major depression disorder it is, thanks to use of artificial intelligence, now possible to predict within a week whether an antidepressant will work. By inputting a brain scan and an individual's clinical information into an AI algorithm, researchers from Amsterdam UMC and Radboudumc could see up to 8 weeks faster whether or not the medication would work. The results of this study are published today in the American Journal of Psychiatry.

Researchers
Liesbeth Reneman

FEBRUARY 6, 2024 · Editors' notes

Could artificial intelligence help or hurt medical research articles?

by Indiana University School of Medicine

Musculoskeletal health topics:
1. The intersection of Alzheimer's disease and bone
2. Neural regulation of fracture healing
3. COVID-19 and musculoskeletal health

ARTICLE | PENSIONS BRIEFING

Artificial Intelligence and Pensions – FOMO or FOOM?

By Adam Boyes | February 7, 2024

Breakthroughs in Artificial Intelligence will lead to change in every industry. In this article, Adam Boyes contemplates how the use of AI could play out in the pensions industry.

Investing in Artificial General Intelligence or AGI Stocks

Published February 3, 2024 | Reading Time: 10 minutes

Skip Ahead

Artificial General Intelligence (AGI) represents highly autonomous systems that match or surpass human capabilities in economically valuable work. AGI has the potential to revolutionise technology and significantly impact our future.

Importance of AGI in modern technology and its potential impact on the future

The AutoGPT Promise

AI in 2024: Intuit Leaders on Possibilities and Pitfalls

As the new year kicks off, several of Intuit's technology leaders share their predictions on the state of AI over the next 12 months.

Could AI 'trading bots' transform the world of investing?

31 January 2024

Jenty Bloom
Business reporter

Detecting AI fingerprints: A guide to watermarking and beyond

Figure 2: A combination of AI news headlines reflecting both utopian and dystopian perspectives.

1.5 A Pragmatic Perspective: Mixed Effects of AI

Rather than utopia or dystopia, the future will likely be a complex mix of benefits, disruptions, new opportunities, and risks. As shown in Figure 2, a pragmatic approach acknowledges AI's dual nature, aiming to maximize its positive impact while minimizing potential harm.

1.5.1 Anticipated positive impacts and emerging opportunities in an AI-augmented future

By 2040, AI is projected to become as essential as the internet, transforming global economies, reshaping industries, and addressing long-standing societal challenges (Jensen, as cited in Rainie and Anderson 2024). The International Data Corporation (IDC) estimates a $19.9 trillion contribution to the global economy by 2030, with AI investment returns of $4.60 per dollar (International Data Corporation (IDC) 2024). While concerns about job displacement persist, AI is also expected to create new employment opportunities (World Economic Forum (WEF) 2025). IDC's Future of Work Employees Survey found that only three percent of workers expect full automation of their roles, while sixty-three percent believe AI will enhance rather than replace their work (IDC 2024). Productivity is expected to soar as AI automates repetitive tasks, enabling humans to focus on creative, strategic pursuits (Olorundare, as cited in Rainie and Anderson 2024).

In healthcare, AI is already improving diagnostics, virtual care, and precision medicine, and may soon support organ production via 3D/4D printing (Al-Saqaf, as cited in Rainie and Anderson 2024). Education is benefiting from adaptive AI tutors (Silwal, as cited in Rainie and Anderson 2024), while personal AI assistants democratize access to finance, mental health, and career planning (Herd, as cited in Rainie and Anderson 2024). Governance is poised to become more transparent and efficient through AI-assisted policy modeling, real-time fact-checking, and predictive decision-making (Turner, as cited in Rainie and Anderson 2024). Urban planning, agriculture, entertainment, and climate adaptation are similarly being transformed by AI through AI-enabled free public transit, robotic farming, and immersive virtual experiences (Bairathi 2025; Jensen and Silwal, as cited in Rainie and Anderson 2024). A notable frontier in AI research is the development of digital "twins"—virtual AI representations of individuals that will assist in decision-making, self-improvement, and lifelong learning (Spohrer, as cited in Rainie and Anderson 2024). Ben Shneiderman, professor emeritus at the University of Maryland, stresses the need for human-centered AI that supports creativity and social connection. Similarly, Associate Professor Loianno predicts that by 2030, autonomous robots will collaborate, learn, and make high-level decisions with minimal human input, improving efficiency across sectors provided strong safety protocols are in place (Loianno, as cited in Ziegler 2024). If paired with fair policies like universal basic income, AI-driven automation could help reduce inequality (Herd and Williams, as cited in Rainie and Anderson 2024). Thus, the AI companion market, valued at USD 28.19 billion in 2024, is projected to grow at a 30.8 percent annual rate through 2030 (Grand View Research 2024). Yet, these advancements raise profound ethical questions. Leaders like Dario Amodei and Sam Altman envision transformative impacts but caution against unchecked ambitions (Robison 2024; Pierce 2024). The choices made today, balancing opportunity with responsibility will shape whether AI drives shared prosperity or deepens inequality by 2040. Ultimately, the AI-augmented future presents a delicate balance between opportunity and responsibility (Jensen, Olorundare, and Al-Saqaf, as cited in Rainie and Anderson 2024).

1.5.2 Negative effects and emerging risks in an AI-augmented future

Despite its promise, AI introduces significant risks. One major concern is techno-solutionism, the belief that AI can solve all social problems (Littman et al. 2021), leading to blind trust in automated decisions, and ignoring their biases. Examples include Amazon's 2018 hiring tool that was scrapped for showing gender bias (Chang 2023) and an algorithm by Optum that prioritized white patients over black patients based on cost assumptions. AI also threatens democratic institutions, enabling misinformation, deepfakes, and social media manipulation (Littman et al. 2021) as seen in cases like the 2016 United States election interference and a fake Pentagon explosion image. Its black-box nature limits transparency and accountability, while mass surveillance erodes privacy and empowers authoritarian regimes. Predictive policing, social credit systems, and mass data collection pose ethical and human rights challenges. Economically, AI could deepen inequality by displacing low-income workers and concentrating power in tech monopolies. It also raises IP concerns, as generative models borrow from copyrighted content without credit or compensation—a phenomenon dubbed the "Great Data Heist." These risks are heightened by AI's ability to produce unintended behaviors, especially in reinforcement learning models. Overreliance on AI in safety-critical applications is also risky, as models can hallucinate or provide incorrect information, which could lead to catastrophic consequences in military systems, AVs, and healthcare diagnostics (Littman et al. 2021; Rainie and Anderson 2024). AI also enables impersonation scams, such as voice cloning, and can be used to find software vulnerabilities, increasing cybercrime risks. The Center for AI Safety warns that AI may boost the scale, speed, and success of cyberattacks, increasing geopolitical risks. Lastly, environmental impacts are substantial. AI training consumes vast energy and addressing these issues will require strong ethical standards, clear regulations, and greater transparency to guide AI toward equitable and responsible use (Littman et al. 2021; Samuel 2023; Rainie and Anderson 2024).

Given these complexities, this research proposes moving beyond a binary characterization of utopia versus dystopia by introducing HEAI, that is AI designed to amplify human potential while minimizing harm. Using NLP techniques like sentiment analysis and topic modeling, we track evolving media narratives around AI. As with past innovations like nuclear energy or the internet, AI's trajectory provokes both hope and fear. This transformative trajectory has raised critical questions about its societal implications and the future it is actively shaping (Samuel et al. 2024; Tripathi et al. 2025). **The key question is not *if* AI will shape the future, but *how* we choose to guide it.** This research aims to provide a data-driven perspective on exploring the emerging trajectories of the impacts of AI on human society and the choices that lie ahead. The rest of the manuscript is structured as follows. The next section presents a very brief literature review across four key domains: education, healthcare, robotics, and workforce. Additionally, the section also examines various societal aspects both within these domains and as a whole. Following the literature review, the methodology section outlines our data collection process and feature engineering. We then detail our exploratory data analysis (EDA) and NLP techniques, including sentiment analysis and topic modeling. Next, we discuss our statistical analysis. Finally, we present the results, followed by a discussion and conclusion, emphasizing the significance of our findings in the context of HEAI.

2. Literature Review

The purpose of this literature review is to briefly introduce four key pillars of human society, to serve as an anchor and provide non-exhaustive context for our NLP- and LLM-based analyses.

2.1 Education

AI has significantly impacted education, offering both promising applications, such as automation and personalization, and complex challenges that require policy responses. Academic institutions have been reevaluating their systems to ensure intended learning outcomes are preserved in the face of rising use of LLMs and generative AI tools (Chidipothu et al. 2025). AI tools help educators streamline administrative processes such as grading, lesson planning, attendance tracking, scheduling, and budgeting. This automation enables teachers to focus more on personalized instruction (Zaman 2023). Furthermore, institutions can analyze large volumes of data using chatbots and analytics platforms to uncover insights that enhance both administrative efficiency and student performance (Ananyi and Somieari-Pepple 2023). AI-powered Adaptive Learning Platforms customize instruction to match each student's pace and proficiency by adjusting materials based on performance data, enhancing engagement, and accommodating diverse learning styles (Piocciochi and Alwabel 2020; Dutta et al. 2024). Additionally, tools such as speech-to-text, language translation, and early innovations in computer vision and speech recognition have also improved accessibility, especially for individuals with disabilities (Bigham and Carrington 2018).

On the negative side, algorithmic bias in AI models, often stemming from skewed training data, can disproportionately affect underrepresented racial and ethnic groups. For example, multiple predictive models performed worse for minority groups in academic success forecasting (Hu and Rangwala 2020; Baker and Hawn 2022). Overreliance on AI (Zhai et al. 2024) can also hinder students' cognitive development, including decision-making and critical thinking. A specific concern related to generative AI tools like ChatGPT is "hallucination," where plausible but entirely false information is generated. Students are at risk of consuming and relying on such inaccurate content. As OpenAI itself warns, ChatGPT's outputs may sound convincing yet be incorrect (Athaluri et al. 2023). This highlights the need to equip students with critical thinking skills and the ability to validate AI-generated responses using more reliable sources (Mhlanga 2023). Beyond these technical and cognitive risks, AI chatbots have also disrupted the ethical landscape of higher education. Recent work suggests that AI-assisted cheating reflects a broader shift in ethical expectations for students and institutions, challenging the adequacy of traditional academic integrity frameworks (Pelaez et al. 2025). Moreover, the increased integration of AI in education may reduce human interaction in educational settings. A lack of meaningful connection with teachers and peers can negatively affect students' motivation, social development, and overall educational experience (Al-Zahrani 2024).

2.2 Healthcare

AI is transforming healthcare by enhancing diagnostics, personalization, accessibility, and operational efficiency. Algorithms can analyze complex datasets, ranging from medical images and genetic data to electronic health records (Alowais et al. 2023), to detect cancers and cardiovascular and neurological diseases with greater accuracy and speed. In mental health, AI identifies early indicators of conditions like depression, anxiety, and post-traumatic stress disorder (PTSD) through behavioral data, speech patterns, and social media activity (Ettman and Galea 2023). Virtual therapists and chatbots provide 24/7 support, particularly in underserved regions where mental health services are limited (Cross et al. 2024). Additionally, AI enables personalized treatment plans tailored to individual genetic profiles and medical histories, improving outcomes while minimizing side effects (Hill 2024). On the operational side, AI streamlines administrative tasks such as scheduling and billing, and improves resource allocation and patient flow, thereby reducing costs and enhancing care delivery (Hoose and Králiková 2024).

Despite its efficiency, the integration of AI into healthcare raises ethical, technical, and regulatory challenges. AI systems lack emotional intelligence crucial for doctor-patient rapport, especially in behavioral health, potentially lowering patient satisfaction and treatment adherence (Cordero 2024). Heavy reliance on personal data introduces privacy risks and cybersecurity threats and can undermine trust if patients are not fully informed of AI's role in their treatment (Farhud and Zokaei 2021). Technical risks like overfitting can result in harmful diagnostic errors (Chustecki 2024), while algorithmic bias stemming from non-diverse datasets may lead to misdiagnoses and healthcare disparities for marginalized populations (Siafakas and Vasarmidi 2024). Finally, because AI algorithms can evolve with time, traditional regulatory frameworks struggle to assess their safety and efficacy consistently, complicating accountability and eroding public trust in AI-enabled medical systems (Price 2019).

2.3 Robotics, Automobile, and Factories

AI is transforming the automobile industry and logistics by enabling predictive modeling, deep learning, and optimization techniques that enhance supply chain efficiency, sustainability, and strategic decision-making (Didast, et al. 2024). In warehouse operations, AI-driven robots and automated guided vehicles (AGVs) improve accuracy, speed, and cost-effectiveness by navigating environments, identifying items, and dynamically allocating tasks (Dehghan et al. 2023; Sodiya et al. 2024). As robotics and AI continue to transform digital logistics, they open avenues for further research in resource orchestration and the development of innovative operational strategies (Rainer et al. 2025). AI's integration into logistics extends to traffic prediction and management, where real-time data from sensors, GPS, and social media is used for dynamic routing, congestion control, and signal optimization, refining mobility systems before full autonomous adoption becomes widespread (Bharadiya 2023).

Despite the advantages AI offers in logistics and automotives, it also introduces risks. AVs, for instance, present cybersecurity and privacy challenges. Threats like GPS spoofing, ransomware, and data breaches endanger safety and can expose sensitive biometric and location data (Bendiab et al. 2023; Sadaf et al. 2023). Similarly, logistics companies must safeguard vast datasets related to cargo, fleet movement, and delivery operations, making cybersecurity a critical priority (Didast, et al. 2024). Furthermore, models often inherit societal prejudices embedded in their training data (Richey et al. 2023), with real-world consequences in areas like hiring, credit assessment, and law enforcement. In logistics and supply chain management (L&SCM), such biases could result in preferential treatment of certain suppliers, products, or regions, contributing to systemic discrimination and inefficiencies in decision-making.

2.4 Workforce

AI creates new employment. For example, AI developers, data scientists, and machine learning engineers are in high demand in areas like AVs, health tech, and finance. This means that AI brings in new professions and sectors for people to work in (Vinson et al. 2024). AI systems can streamline repetitive tasks that enhance human productivity and shift the focus more towards decision-making and creativity. For example, in the marketing and finance domain, AI tools accelerate data analysis and extraction enabling workers to make quick decisions and increase job satisfaction (Acemoglu and Restrepo 2019). AI helps individuals acquire new skills as numerous organizations and corporations are providing training initiatives in areas such as programming, data analysis, and the applications of AI tools (Frank et al. 2019). AI systems also help small businesses and entrepreneurs by reducing operational costs and allowing for smarter decision-making through the interpretation of data trends (Jumaev 2024). In healthcare and customer service, for example, AI handles administrative tasks, which gives professionals an opportunity to work with their customers and have a better work-life balance. Moreover, AI enables remote work by supporting virtual collaboration and global hiring (Faluyi 2025).

On the other hand, AI-driven automation leads to significant job displacement, particularly in sectors like manufacturing, administrative support, and customer service. Even employment changes in white-collar sectors such as accounting, legal assistance, and journalism, lead to a broader impact on the workforce (Faluyi 2025). AI also introduced the creation of "gig" roles that lack stability, benefits, and career growth opportunities which leads to job dissatisfaction, lower wages, and a growing sense of economic insecurity among the workforce (Acemoglu and

Restrepo 2019). High-skilled human workers may also experience wage stagnation as AI can perform complex data analysis, legal document reviews, and medical diagnostics more efficiently and cost-effectively (Frank et al. 2019). AI systems may accelerate employment inequality and exacerbate the phenomenon of job polarization. Between two and five percent of jobs are to be automated where low-skilled workers face higher risks of displacement, while those with advanced technical skills benefit from new opportunities, creating a labor market divide that can contribute to social tensions and economic instability (Gmyrek et al. 2024). AI's influence extends beyond direct automation, affecting supply chains, service delivery models, and business strategies which leads to significant shifts in job availability and requirements, forcing companies to restructure their workforce and reduce the need for traditional job roles (Webb 2019). Such anxiety related to job insecurity and occupational stress among employees can negatively impact mental health and reduce job satisfaction. Furthermore, constant monitoring and performance tracking by AI systems can increase burnout rates and impact productivity (Chui et al. 2016). This phenomenon is closely related to the John Henry effect, where workers feel compelled to compete against automated systems, leading to overexertion and detrimental effects on both physical and mental health (Gammon and Bornstein 2018).

2.5 Society

AI is reshaping nearly every facet of society, from government operations and public services to entertainment and personal relationships. It enhances productivity, decision-making, and crisis response, yet also amplifies risks around bias, inequality, and privacy, particularly in low- and middle-income countries with limited regulation and digital infrastructure (Tony Blair Institute for Global Change 2024). AI-driven tools like ChatGPT support work management and public access, but reliance on biased training data can reinforce discrimination. As AI systems increasingly influence policy, law, and elections, ethical governance becomes essential to protect human rights, maintain accountability, and prevent the consolidation of power among a few tech giants (United Nations Regional Information Centre (UNRIC) 2024; Elysée Palace 2025). In democratic contexts, AI's dual nature is especially apparent. On one hand, chatbots streamline voting information, and fraud detection tools boost electoral trust. On the other, AI enables targeted political ads and deepfake content that manipulate voter behavior and undermine institutions (Bond 2024; Mishra 2024; Robins-Early 2024). Deepfakes featuring celebrities such as Taylor Swift and Elon Musk have heightened public concern over misinformation (Nguyen 2024), while Netflix's undisclosed use of AI-generated visuals has drawn criticism for threatening media authenticity (Belanger 2024). The European Union's AI Act aims to address these risks by regulating deceptive practices and enhancing transparency. The legal system is similarly challenged. AI expedites case evaluations but complicates questions of accountability and data ethics. Microsoft's CoPilot+ Recall feature, for example, faced backlash for unauthorized data collection (Marcinek et al. 2024; Rahman-Jones 2024). Facial recognition technologies have led to false arrests, most notably that of Porcha Woodruff, an African American woman wrongfully detained due to algorithmic error (Swarns 2023). Discriminatory outcomes have also surfaced in AI hiring tools, as evidenced by the iTutorGroup case, where older candidates were filtered out unfairly (U.S. Equal Employment Opportunity Commission (EEOC) 2023). These examples underline the urgent need for robust oversight to safeguard civil rights in AI-driven processes. AI's impact on personal and environmental spheres is complex. Social robots and virtual assistants offer companionship and reduce loneliness but may weaken human connection over time. Algorithms that prioritize engagement over meaningful dialogue on social platforms can fuel polarization and reduce empathy (Johnston 2020; Rodilosso 2024). Environmentally, AI can optimize energy use and predict climate events, yet model training consumes substantial energy and generates e-waste, often affecting marginalized regions. Calls for "Sustainable AI" reflect a growing recognition that innovation must align with ecological and social responsibility. Ultimately, navigating these tensions requires proactive governance, ethical design, and cross-sector collaboration to ensure AI supports equitable and sustainable progress.

3. Data

3.1 Data Collection and Extraction

For our research, we collected a large dataset of AI-related news headlines using Google News RSS feeds (Google n.d.) to capture multilingual perspectives from different regions. Articles were gathered between November 9, 2023, and November 11, 2024, with English-language coverage extended to February 12, 2025, to include recent developments. We used Feedparser (McKee and Pilgrim 2010), BeautifulSoup (Richardson 2023), and Requests (Reitz n.d.) for content retrieval and parsing. Queries included terms like "AI" and "Artificial Intelligence" in over 40 languages. For dynamic pages, ScrapingBee (ScrapingBee n.d.) was used to render JavaScript content. For consistency, all article titles and sources were translated to English using GoogleTranslator from the deep-translator library (Baccouri 2020). Translations were done in batches of 20 for performance optimization. The final dataset includes 288,429 records, each containing the article's publication date, English-translated title and source, and the original language. This structure supports consistent, multilingual analysis of global AI-related media coverage. While GoogleTranslator provides fast and scalable translations across more than 40 languages, it is important to note that automatic translations are not always perfect. In most cases, the tool produced reliable and intelligible English renderings of headlines, sufficient for large-scale analysis. However, nuances such as idiomatic expressions, cultural references, or domain-specific terminology may not always be preserved with complete accuracy (Aiken and Balan 2011; Groves and Mundt 2015). Given the volume of data, human validation for every headline was not feasible, but the consistency and general quality of the translations were adequate for capturing sentiment, topics, and linguistic patterns at scale.

3.2 Feature Engineering

Following data collection, we performed feature engineering to extract relevant metadata from the news titles and publication dates. This process involved computing the number of characters and words in each title, identifying the day of the week, month, year, and quarter of publication, and determining whether the article was published on a weekend. Additionally, we performed text classification to categorize news articles into relevant themes, including Education, Healthcare, Robotics, Workforce, and Society. This was done using a pattern-matching method and regular expressions with keywords listed in Table 1. The distribution was: *Education*—19,465 (6.75 percent), *Healthcare*—13,941 (4.83 percent), and *Robotics*—19,372 (6.72 percent), *Workforce*—40,131 (13.91 percent), and *Society*—35,027 (12.1 percent). Notably, 62.7 percent of headlines remained uncategorized. Additionally, 24,627 headlines overlapped across categories, resulting in 88,837 unique category assignments. For balanced analysis, we randomly sampled 10,000 unique headlines per category (without replacement). Since *Healthcare* had fewer than 10,000 unique entries, additional multi-category headlines labeled as *Healthcare* were included to meet the target. This approach ensured both structure and broad coverage. We further applied topic modeling and sentiment analysis to the resulting 50,000-headline dataset.

Prefix-based matching (e.g., “educat-“ matching “education” or “educator”) allowed for morphological variations of key terms. Without this, a partial match could mistakenly capture words like *deducate*, which is unrelated to education. At the same time, standalone keywords (e.g., “bus” not matching “business”) and exact acronym detection (e.g., “EV” but not in “event”) helped reduce false positives. Multi-word phrases like “self-driving” and “human-robot” were matched as full terms to preserve their meaning. Each theme was encoded as a binary feature per headline. While not manually validated and sensitive to context, this method offers a scalable approach to thematic classification, with scope for refinement using advanced models.

Table 1: Keywords used for categorizing AI news headlines in our dataset by domain.

Domain	Keywords
Education	educate, learn, teach, study, academic, curriculum, pedagogy, student, school, classroom, course, professor, lecturer, university, college, campus, tutor
Healthcare	health, medical, doctor, nurse, hospital, clinic, pharmaceutical, drug, biotech, diagnosis, patient, treatment, vaccine, telemedicine, disease, cardio, immune, neuro, physician, medical technology, radiology, addiction, abuse, suicide, depression, psychology, surgery, therapy, mental, wellness, genomics, genetics, epidemic, pandemic, cancer, diabetes, biomedical, EHR, X-ray
Robotics	robot, autonomous, navigate, cyborg, industrial, agriculture, combat, weapon, force, sensor, driver, logistics, vehicle, electric, farm, automation, mobility, fleet, humanoid, automated, autopilot, aerial, unmanned, automotive, automobile, military, army, navy, naval, transportation, drone, warehouse, car, bus, train, truck, pilot, battery, plane, flight, aircraft, CAV, UAV, EV, ADAS, DARPA, SWARM, LiDAR, self-driving, pick-and-place, human-robot, supply chain, computer vision
Workforce	job, recruit, work, organization, career, professional, business, enterprise, company, employ, skill, corporate, layoff, manager, startup, entrepreneur, investment, investor, venture, replace, unemployed, hiring, hire, companies, firms, talent, CEO, CTO, CIO, CDO, HR
Society	ethical, regulation, social, trust, democratic, equal, legislation, culture, human, law, public, rights, society, privacy, societal, policy, governance, transparency, accountability, compliance, government, sustainability, health, election, war, relationship

4. Methodology

4.1 EDA

For EDA, we conducted linguistic, geographic, and quantitative textual analysis to explore the characteristics of AI news headlines. Finally, we perform an analysis on our domain classification.

4.1.1 Linguistic and geographical diversity

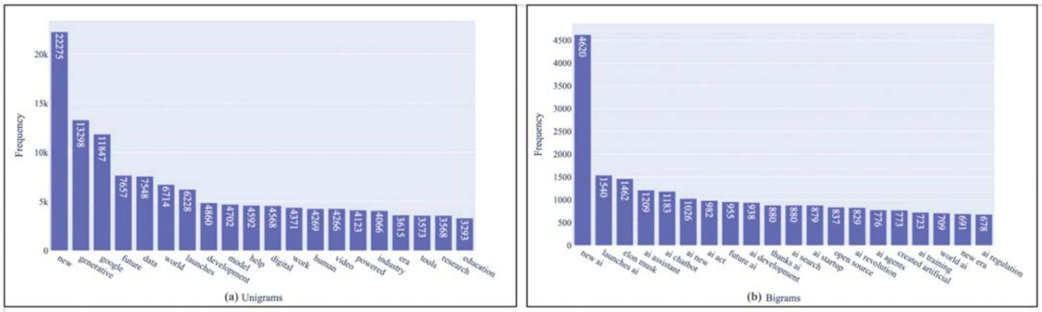
Our dataset has shown strong linguistic diversity. English leads with eighteen percent of headlines, but many non-English articles reflect global engagement. Spanish accounts for the highest number of non-English articles, followed closely by French, Korean, Italian, Japanese, and Portuguese. Additionally, German, Indonesian, Russian, and Arabic also represent substantial portions of the dataset. Given the multilingual dissemination, it appears that conversations about AI are pertinent worldwide.

4.1.2 AI news headlines textual analysis

The dataset showed considerable variability in AI news headline lengths, averaging 76.41 characters (standard deviation [SD] = 29.49), with a range of 2 to 465 characters. Most headlines contained an average of 11.87 words, with counts ranging from 1 to 75, reflecting diverse editorial styles from concise to detailed. To explore content patterns, we conducted an n-gram analysis. Unigrams (Figure 3(a)) highlighted frequent mentions of "new," "generative," "research," and "education," pointing to ongoing innovation and AI's role in knowledge and technology. Frequent references to Google indicated its major influence. Bigrams (Figure 3(b)) like "new ai" and "ai regulation" emphasized current developments and regulatory discourse. Trigrams (Figure 3(c)), such as "using artificial intelligence," reflected practical applications, as well as themes in education and governance. Quadgrams (Figure 3(d)) included references to policies ("EU AI Act"), realism in AI outputs ("look like real life"), public figures (e.g., Elon Musk, Jensen Huang), and concerns like "artificial intelligence replace humans," indicating persistent anxieties around job displacement. Overall, the headlines had both optimism about AI advancements and unease about its societal implications.

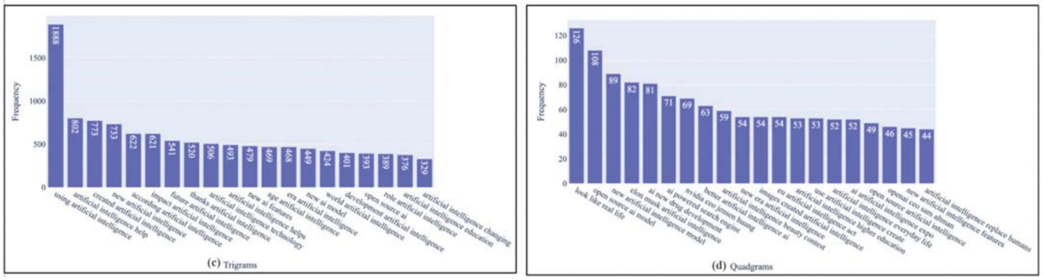
4.2 NLP Analysis

We conducted topic modeling across the five domains using advanced NLP techniques to analyze 10,000 news headlines per domain. The workflow included data preprocessing, embedding generation, dimensionality reduction, clustering, and topic extraction. Stopwords combined sets from NLTK (Bird et al. 2009), Scikit-learn (Pedregosa et al. 2011), a custom-defined list, and numbers from 0 to 9999 to remove irrelevant values.



(a) Unigrams

(b) Bigrams



(c) Trigrams

(d) Quadgrams

Figure 3: Analysis of (a) unigrams (top-left), (b) bigrams (top-right), (c) trigrams (bottom-left), and (d) quadgrams (bottom-right) found in news headlines from our dataset.

The BAAI/bge-small-en model generated dense vector embeddings, stored as NumPy arrays for reuse and computational efficiency. Uniform Manifold Approximation and Projection (UMAP) (McInnes et al. 2018) reduced the high-dimensional embeddings using n_neighbors = 12, n_components = 5, min_dist = 0.1, and cosine similarity to place semantically similar topics closer together. Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes et al. 2017) then clustered the data with min_cluster_size = 20, min_samples = 5, and epsilon = 0.2 to balance granularity. BERTopic (Grootendorst 2022) extracted topics using the embeddings and clusters, with a TF-IDF vectorizer set to unigrams and bigrams, max_df = 0.90, and min_df = 3 (Jones 1972). The top nine topics were selected using BERTopic's topic reduction function. This approach effectively captured thematic structures and domain-specific influences of AI. While Latent Semantic Analysis (LSA) was also tested, BERTopic produced more interpretable topics with clearer semantic coherence based on qualitative assessment. Further, we conducted sentiment analysis, a widely applied NLP technique used in media analysis, social media monitoring, and customer feedback (Wankhade et al. 2022). Similar NLP-based sentiment and topic modeling approaches have been successfully applied to online forums such as Reddit to study risk, harm, and behavioral patterns, demonstrating the robustness of these methods for analyzing large-scale digital discourse (van der Maas and Samuel 2025). Prior studies have shown its effectiveness in capturing public opinion during high-impact events like the COVID-19 pandemic, incorporating spatiotemporal and socioeconomic data to inform policy (Samuel et al. 2020a, 2020b; Ali et al. 2021; Rahman et al. 2021). In this study, sentiment analysis was applied to AI-related news headlines to assess general attitudes—positive, neutral, or negative—toward AI developments. We used the BART-large-Multi-Genre Natural Language Inference (MNLI) model (Lewis et al. 2020), a pre-trained transformer-based zero-shot classifier aligned with recent findings on the strong performance of LLMs in sentiment analysis tasks (Zhang et al. 2023). Implemented via the Hugging Face Transformers library (Wolf et al. 2020), the model applies natural language inference to classify text into sentiment categories without domain-specific training, making it well-suited for our large-scale, multilingual dataset of 50,000 AI news headlines. Each headline was assigned one of the three candidate labels: "Positive," "Negative," and "Neutral" using the zero-shot classification pipeline. The model itself uses a probability distribution based on entailment relationships learned during the pretraining on the MNLI corpus (Williams et al. 2018).

Although no domain-specific ground-truth validation was conducted, prior benchmarks demonstrate strong zero-shot performance of BART-large-MNLI. The model achieved 74.7 percent accuracy on the Sentiment140 dataset for binary sentiment classification, where BART-large-MNLI outperformed all other zero-shot models tested, increasing to eighty-seven percent for binary classification when combined with ensemble methods (Kanclerz et al. 2022). Additionally, another study conducted a comprehensive evaluation of LLMs across 13 sentiment analysis tasks on 26 datasets, showing that transformer-based zero-shot classifiers demonstrate reliable performance across multiple sentiment tasks, particularly when annotated data are limited (Zhang et al. 2023). Given our focus on large-scale trend analysis rather than individual classification, this validated performance was sufficient to capture broad sentiment patterns.

4.3 Statistical Analysis

For the quantitative analysis of sentiment patterns across AI-related news headlines, we used a comprehensive statistical approach to study variations in sentiment across our five domains. We began with descriptive statistics to assess central tendencies, calculating mean sentiment scores for baseline comparisons across domains. To test distribution normality, we used the Shapiro-Wilk test (Shapiro and Wilk 1965), which indicated non-normal sentiment distributions across all domains. This warranted the use of non-parametric methods. We conducted a Kruskal-Wallis test (Kruskal and Wallis 1952) to examine whether sentiment scores significantly differed among the five domains. Upon finding significant differences, we performed pairwise Mann-Whitney U tests (Mann and Whitney 1947) (a non-parametric alternative to one-way ANOVA; Girden 1992) to identify which domain pairs exhibited statistically significant sentiment differences. To reduce the risk of Type I errors, we applied a Bonferroni correction, adjusting the significance threshold to $\alpha = 0.005$—since we are doing ten comparisons, we adjusted the significance level ($\alpha = 0.05$) to account for multiple testing. For each domain pair, we reported the U-statistic, p-value, and effect size using rank-biserial correlation (r), categorized as small ($|r| < 0.3$), medium ($0.3 \leq |r| < 0.5$), or large ($|r| \geq 0.5$) to reflect the magnitude of sentiment differences. We also calculated standard deviations to assess sentiment variability within each domain. Finally, a median test was conducted to complement our analysis of central tendencies, providing a robust view of inter-domain sentiment differences.

4.4 Results and Analysis

4.4.1 Sentiment analysis

The sentiment analysis was performed on a dataset of 50,000 in total combining all five domains scoring each headline with a continuous score from -1 to 1. This provides an overall distribution and trends of sentiments over time.

The sentiment score was sorted, and the results are presented in Figure 4, top-left, where the positive sentiment is prominently showcased in green compared to the negative sentiment represented in red. This asymmetry suggests an overall optimistic view of AI in the dataset. Both extremes are well represented while the transition from negative to positive is smooth, showcasing some neutral sentiments.

The sentiment analysis was also conducted on each domain consisting of 10,000 records of data. The education domain shows a strong inclination towards the positive sentiment with over eighty percent of the records falling into the positive side, indicating an optimistic view of news towards education, possibly because of the advancements in learning methodologies and accessibility to education. However, there are some negative sentiments. The health domain's sentiment also resembles education where the positive sentiment dominates. Health has the greatest number of positive sentiments amongst all the domains which implies advancements in medical innovation and drug discovery. The negative sentiment might be due to privacy concerns over patient data. The sentiment for the robotics domain is also positive with around 80/20 split. While it shows some negative sentiment due to privacy concerns and job displacement, the overall discourse surrounding robotics is largely optimistic due to the growing advancements and automation due to AI. The workforce domain reflects a positive outlook overall but has a slightly more negative sentiment with approximately a 75/25 split in comparison to education. The data represents sentiments on both extremes with a possibility of having professional growth on the positive end and job insecurity on the lower end. The general positivity suggests a perception of career development and opportunities in various fields. The society domain brings some balance to the sentiment distribution as it has a relatively wider spread amongst the negative and positive sentiment scores. This indicates that there are mixed reviews of AI on societal issues with some expressing concerns while others discussing the advancements. Overall, the sentiment across all domains remains largely positive with some negative perceptions as shown in Figure 4.

4.4.2 Topic modeling

The topic modeling analysis was conducted on each domain consisting of 10,000 records of data. This study identified the overall sentiment and dominant themes within each domain. The study extracts the top nine topics from each domain for topic modeling analysis. For education, the topics extracted explain the inclination of the sentiment emphasizing university systems, digital learning, and AI-driven educational tools. Research programs, higher education institutions, free online courses, and the role of AI in language learning are some of the recurring themes shown in Figure 5. However, topics on academic integrity, cheating, and plagiarism highlight the concerns regarding AI in education. The health domain (Figure 6) is dominated by themes of pharmaceutical innovation and personalized medicine, with a strong emphasis on cancer research, diagnostic imaging, telemedicine, and mental health. The vaccines and drug discovery showcase innovative drug development in the health domain. Additionally, cardiovascular diseases and dementia indicate a focus on chronic conditions which indicates advancements due to AI. Mental health-related topics such as therapy and psychology are also included, reflecting a growing concern for psychological well-being in healthcare discussions. The mention of suicide is unclear, possibly reflecting AI risks or its role in suicide prevention.

In the robotics domain (Figure 7), key topics included consumer robotics, industrial automation, and AVs, especially in the automobile industry. Ethical concerns around data privacy and regulation highlight the need for stronger compliance in the industry. In the workforce domain (Figure 8), positive themes included stock market trends,

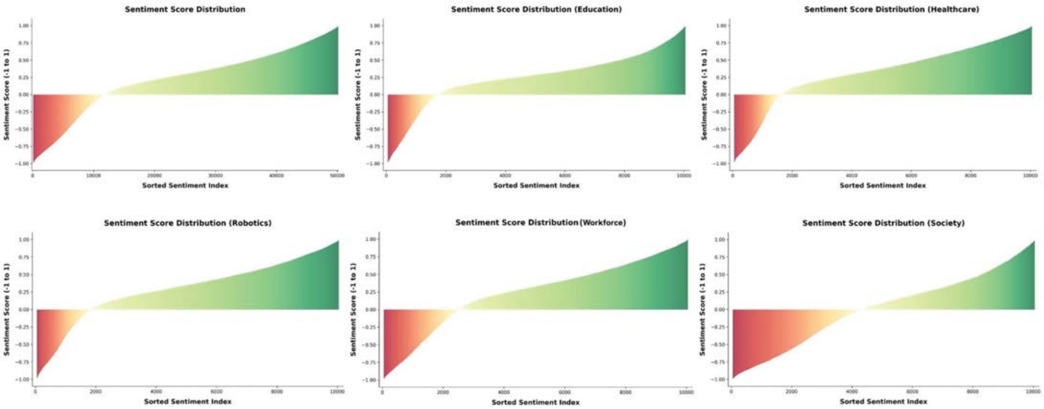


Figure 4: Sentiment distribution bar plot showing positive sentiment (green) and negative sentiment (red).

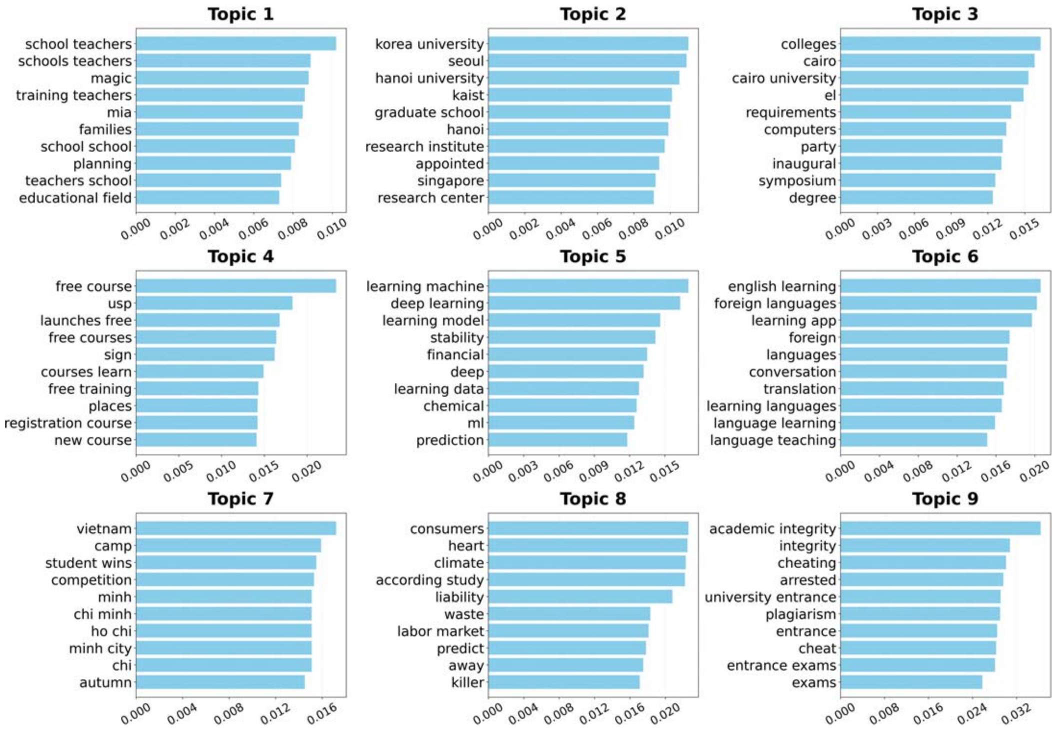


Figure 5: Results of topic modeling on news headlines in the education domain.

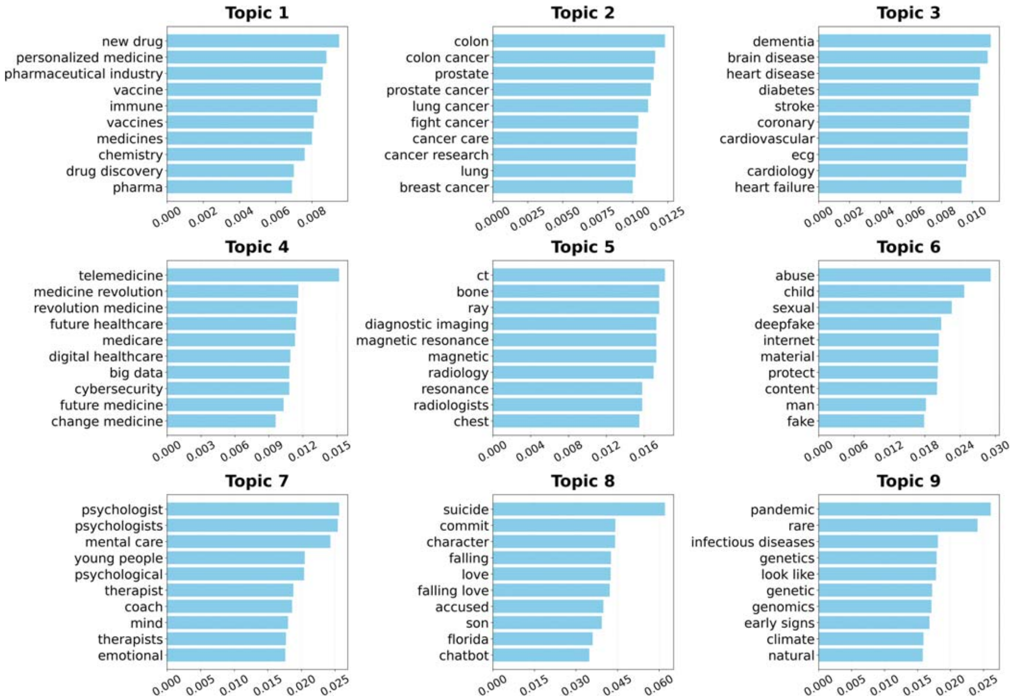


Figure 6: Results of topic modeling on news headlines in the health domain.

venture capital investment, and generative AI's role in boosting productivity across business, legal, and communication tasks. However, concerns over job security and automation-driven displacement persist.

The topic extraction for the society domain indicates high relevance to government and public administration utilizing AI, while the disinformation and regulations showcase the negative aspects (Figure 9). The topics around AI in

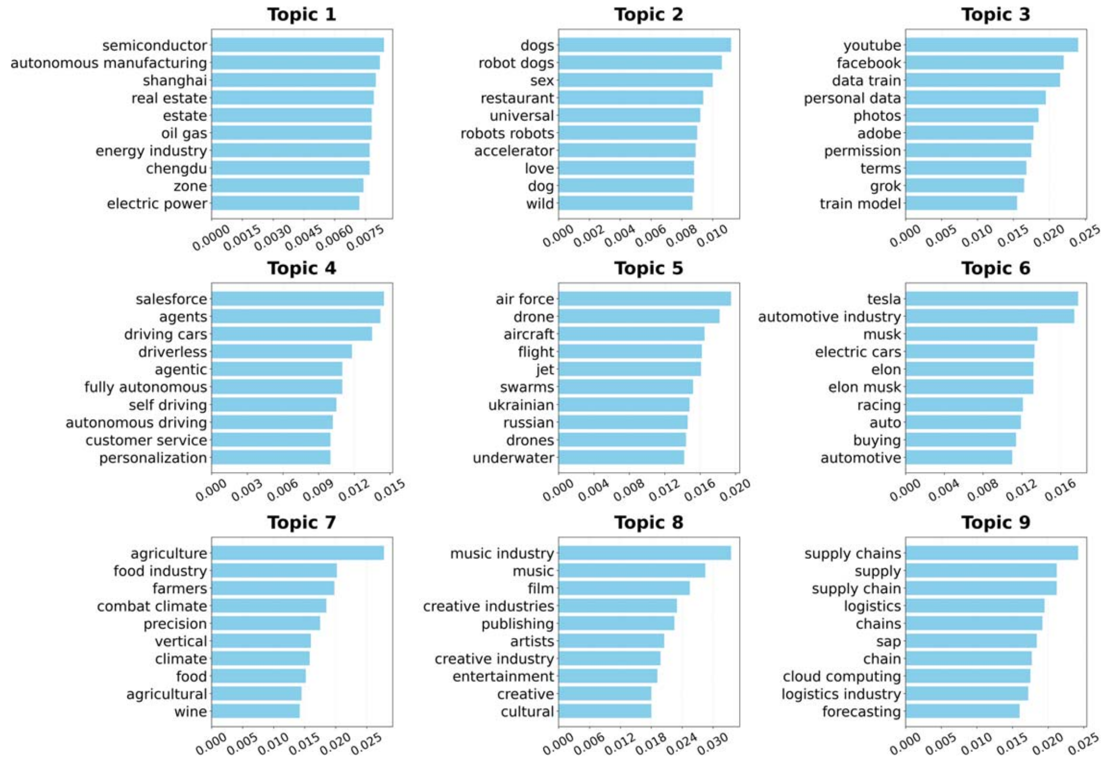


Figure 7: Results of topic modeling on news headlines in the robotics domain.

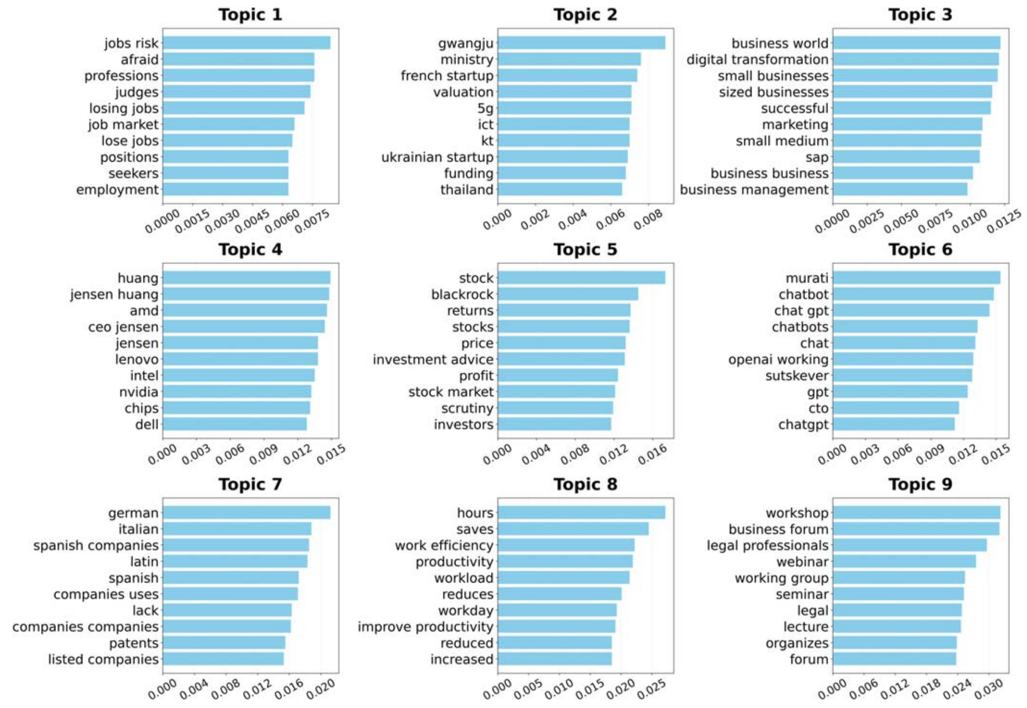


Figure 8: Results of topic modeling on news headlines in the workforce domain.

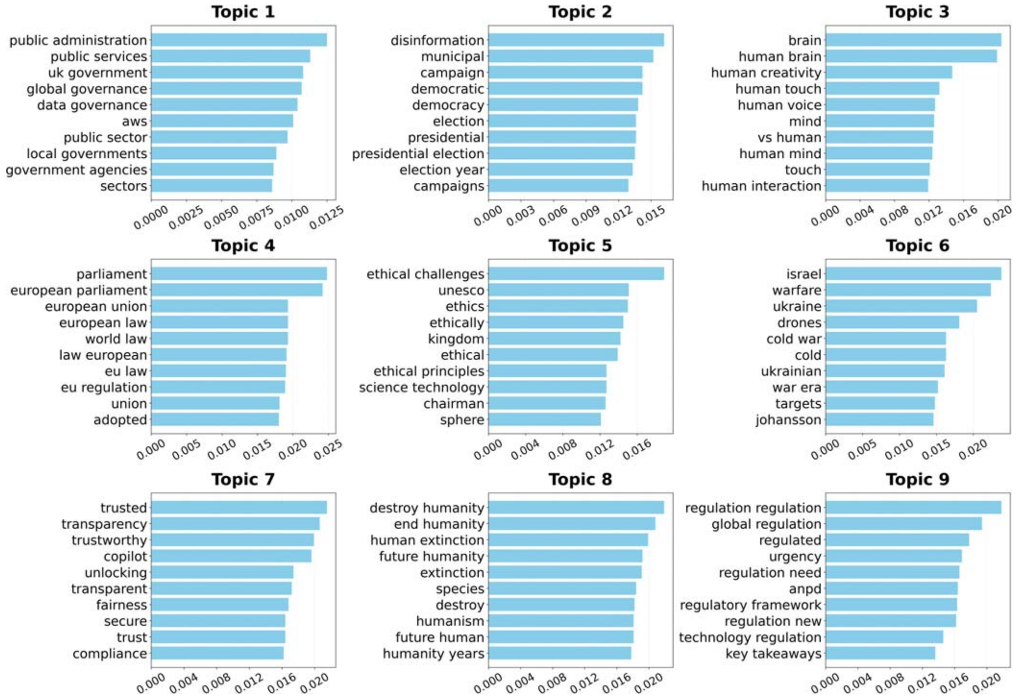


Figure 9: Results of topic modeling on news headlines in the society domain.

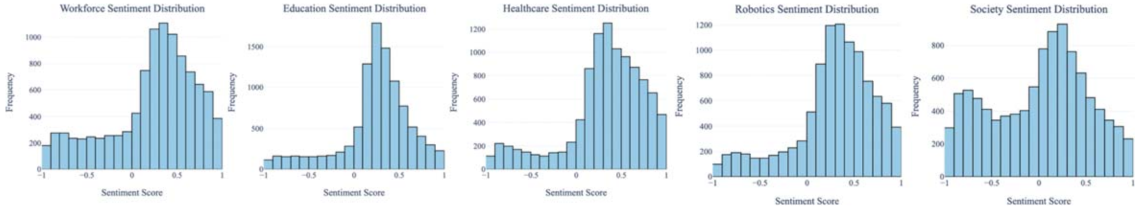


Figure 10: Sentiment score distribution across five domains.

governance, regulations in the European Union, and data transparency lead to an intersection between AI and policies around it. Additionally, some topics such as warfare, Cold War, and military drones suggest strong involvement of AI in geopolitical conflicts. Topics with trust and transparency in comparison to existential threats to humanity prove the analysis of an evenly spread sentiment.

4.3 Cross-Domain Analysis of Sentiment in News Headlines

4.3.1 Normality testing

Visual inspection of sentiment distributions (Figure 10) shows clear deviations from normality across all domains. *Education* exhibits a near-symmetric distribution peaking around 0.25, though a slight leftward tail suggests mild negative skewness. *Healthcare* and *Robotics* display strong positive skewness, with sentiment concentrated in the positive region. *Workforce* shows a positively skewed, mildly bimodal pattern with a secondary peak in the negative range. In contrast, *Society* presents a pronounced bimodal distribution, indicating polarized sentiment with peaks in both negative and positive ranges. These observations are further validated by the Shapiro-Wilk test, which indicated significant departures from normality across all domains: *Education* ($W = 0.921$, $p < 0.001$), *Healthcare* ($W = 0.911$, $p < 0.001$), *Robotics* ($W = 0.931$, $p < 0.001$), *Workforce* ($W = 0.931$, $p < 0.001$), and *Society* ($W = 0.963$, $p < 0.001$). These results confirm the violation of the normality assumption and justify the use of non-parametric methods for subsequent analyses.

4.4 Domain-Wise Sentiment Comparison

A Kruskal-Wallis H test was conducted to determine whether sentiment scores differed by domain. The results revealed a statistically significant difference in sentiment scores across the five domains, H(4) = 2,668.565, $p < 0.001$. To further investigate, post-hoc pairwise comparisons were conducted using the Mann-Whitney U test with Bonferroni correction ($\alpha = 0.005$). All pairwise comparisons were statistically significant ($p < 0.005$), including the smallest difference between Education and Workforce as shown in Table 2.

To assess practical significance, mean sentiment differences between domains were analyzed as shown in Table 3.

These results confirm that Society is an outlier, showing much lower sentiment scores than the other domains. A non-parametric median test also confirmed significant differences in median sentiment across the five domains (statistic = 1,977.768, $p < 0.001$), with a grand median of 0.297. Descriptive statistics of sentiment scores indicate clear differences in tone across domains. *Healthcare* had the most positive coverage, with the highest mean sentiment (0.321) and moderate variability, reflecting steady optimism. *Robotics* followed (mean = 0.293), generally positive but with some concerns. *Education* (mean = 0.244) showed a moderately positive outlook and the lowest standard deviation (0.387), suggesting a stable perception. *Workforce* (mean = 0.235) reflected more mixed sentiment, likely due to job-related anxieties, and had a higher variability (0.484). In contrast, *Society* stood out as the most dystopian, with a near-neutral mean (0.011) and the highest variability (0.514), pointing to polarized views shaped by issues such as surveillance, inequality, and ethics.

We further looked at the topics emerging from negative AI news headlines. Figure 11 presents the top keywords associated with each topic extracted from negative sentiment news headlines across all domains. Nine distinct topics were identified: Topic 1 appears centered around labor and automation, with terms like *drivers*, *self-driving*, *unemployed*, and *professions*. Topic 2 focuses on education-related concerns, including *high school*, *teachers*, *cheating*, and *exams*. Topic 3 reflects existential and ethical fears tied to AI, with keywords such as *destroy humanity*, *deceive*, and *copyright*. Topic 4 is concerned with healthcare and clinical issues, including *cancer*, *hospitals*, *patients*, and *medicine*. Topic 5 deals with regulatory and legal frameworks, featuring terms like *EU law*, *regulation*, and *world law*. Topic 6 shows tech leadership and robotics, with names like *Sam Altman*, *OpenAI*, *chatbots*, and *robotics*. Topic 7 addresses political interference and elections, with keywords such as *influence elections*, *democracy*, and *campaigns*. Topic 8 relates to social media and privacy, with terms like *YouTube*, *Instagram*, *user data*, and *NVIDIA*. Topic 9 emphasizes warfare and geopolitical threats, including *nuclear*, *Ukraine*, *weapons*, and

Table 2: Pairwise Mann-Whitney U test results with effect sizes.

Domain Pair	U-statistic	p-Value	Effect Size (r)	Interpretation
Education versus Healthcare	41,978,129.00	0	0.16	Small effect
Education versus Robotics	44,726,836.50	0	0.105	Small effect
Education versus Society	63,714,804.00	0	−0.274	Small effect
Healthcare versus Robotics	52,559,970.00	0	−0.051	Small effect
Healthcare versus Society	68,426,711.00	0	−0.369	Medium effect
Robotics versus Society	66,616,369.00	0	−0.332	Medium effect
Workforce versus Education	52,451,658.50	0	−0.049	Small effect
Workforce versus Healthcare	45,255,945.00	0	0.095	Small effect
Workforce versus Robotics	47,576,088.00	0	0.048	Small effect
Workforce versus Society	63,428,451.00	0	−0.269	Small effect

Table 3: Sentiment differences (Δ) across domains.

Large Differences ($\Delta > 0.20$)	Society versus Healthcare (0.310), Robotics (0.282), Education (0.233), Workforce (0.224)
Moderate Differences ($0.05 < \Delta < 0.20$)	Healthcare versus Workforce (0.086), Education (0.078), Robotics versus Workforce (0.058), Education (0.049)
Small Differences ($\Delta < 0.05$)	Education versus Workforce (0.008), Healthcare versus Robotics (0.028)

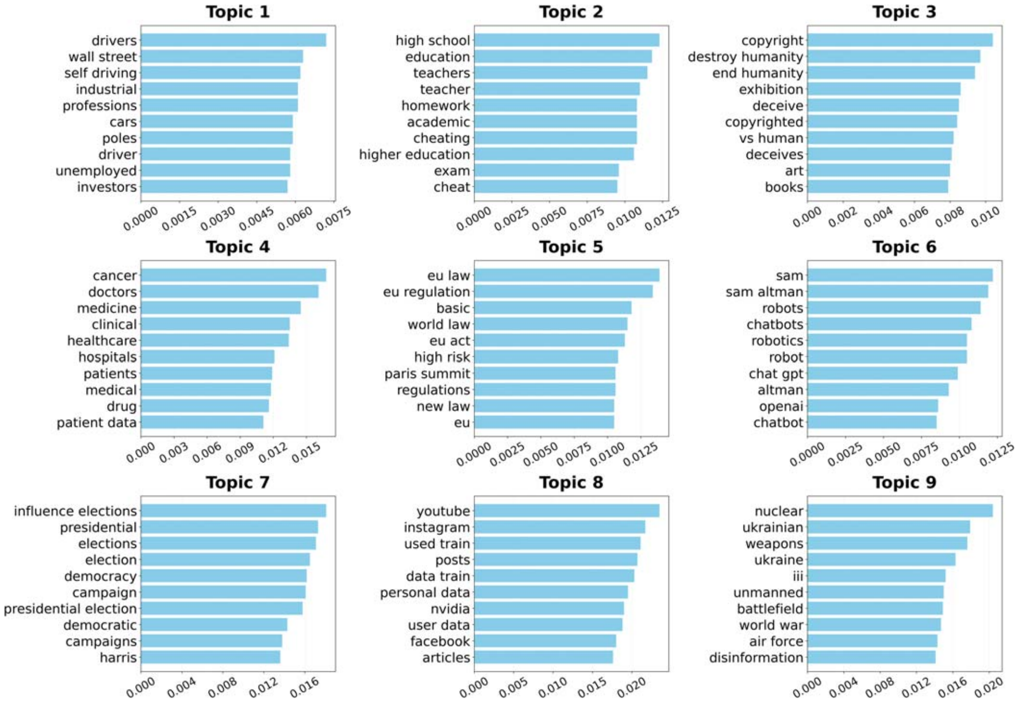


Figure 11: Results of topic modeling on negative news headlines across all domains.

disinformation. These topics show key negative narratives and concerns surrounding AI across societal, political, and technological contexts.

5. Discussion

In this study, AI-related headlines were analyzed using sentiment analysis, topic modeling, and statistical testing to examine how AI is framed in global news discourse across five domains: Education, Healthcare, Robotics, Workforce, and Society. These domains showed a spectrum of narratives, ranging from optimistic portrayals in healthcare and robotics to polarized and cautionary tones in societal coverage. This divergence shows that perceptions of AI are not monolithic; rather, they are shaped by the specific affordances, risks, and socio-political dynamics unique to each domain. Sentiment distributions varied significantly across domains, as confirmed by Kruskal-Wallis and pairwise Mann-Whitney *U* tests, reinforcing the importance of disaggregated approaches in AI impact studies. Topic modeling further elucidated the complex themes driving these sentiments, from innovation and empowerment to inequality and existential risk. Education headlines focus on adaptive learning but also flag risks like academic dishonesty and overdependence on generative tools. In healthcare, AI is transforming diagnostics, personalized treatment, and operational efficiency. It enables early anomaly detection through advanced algorithms like GANs and VAEs, supports customized care plans, accelerates medical research via synthetic data and disease modeling, and automates administrative tasks to reduce clinician burnout. Predictive analytics enhance risk prediction, pandemic response, and population health management. Beyond clinical settings, AI improves medical education, marketing, and revenue cycle management. However, concerns persist around algorithmic bias, data privacy, and the opacity of deep learning models, which risk undermining trust and widening disparities. Ensuring equitable, transparent, and validated AI deployment is crucial to preserving healthcare's human-centered values (Bhuyan et al. 2025).

Robotics is tied to industrial automation and AVs, with both praise for efficiency and worry over job loss. Workforce, although broadly positive, showed deeper polarization, as evidenced by its high standard deviation. This domain exhibited conflicting themes: new job creation in AI and data science juxtaposed with fears of job automation, precarious gig work, and wage stagnation. Society-related topics covered deepfakes, disinformation, and geopolitical tensions, framing AI as both powerful and potentially harmful. Our topic model on negative headlines further confirmed widespread concerns. Themes like AI in warfare, ethics, and labor automation convey the fear of unintended consequences if AI is left unchecked. These findings suggest a broader media trend of technological determinism, where AI is celebrated in some domains and problematized in others. This polarized framing risks distorting public understanding, policymaking, and responsible adoption. Positive bias in healthcare and robotics may create "informational bubbles," while dystopian framings in societal discourse could amplify fear and mistrust. News media, therefore, must aim for balanced, fact-based reporting that resists binary extremes. The role of news media in creating public understanding is particularly critical given that much of the public relies on such sources to construct their beliefs about technological futures.

These findings emphasize the need for HEAI that prioritizes human values, rights, and well-being. Importantly, recent work on taxonomy-based query classification demonstrates that many perceived failures of LLMs arise from mismatches between query intent and model capability rather than inherent system unreliability, highlighting how misunderstanding AI behavior can contribute to distorted public narratives and misplaced fear (Samuel et al. 2025b). It demands transparency, fairness, accountability, and user agency at every stage of AI design, deployment, and regulation. Rather than centering only on performance or profit, HEAI focuses on ensuring that AI serves people; supporting equity, inclusion, and social good. Foundational research has articulated the importance of culturally sensitive, personally adaptive, and ethically guided AI frameworks across education, governance, and information systems (Samuel et al. 2023; Kashyap et al. 2024). These contributions introduce concepts such as adaptive cognitive fit and generative systems design, emphasizing participatory AI development and alignment with diverse human needs (Garvey et al. 2021; Samuel et al. 2022). We propose the development of a globally distributed, HEAI-enhanced ethics board, functioning as a decentralized consortium where AI systems assist human experts from diverse domains like philosophy, sociology, law, and computer science in evaluating AI models before deployment. Blockchain can further support transparency and shared oversight. Additionally, ethical "red-teaming" (the practice of engaging independent groups to simulate adversarial and failure scenarios) should become standard practice, stress-testing models in sensitive domains like healthcare and law enforcement before public release. From a policy standpoint, differentiated strategies are required to address domain-specific risks. For instance, while AI in education demands pedagogical reforms that promote digital literacy and critical thinking, AI in society necessitates robust regulatory frameworks that safeguard privacy, autonomy, and democratic participation. Similarly, labor market shifts induced by AI require proactive investments in workforce reskilling, social safety nets, and ethical guidelines for workplace automation. Thus, the application of AI should be accompanied by education on its limitations, potential biases, and ethical considerations. Moreover, public discourse must be grounded in nuance, resisting reductive framings of AI as purely utopian or dystopian.

6. Future Research

Through sentiment analysis and topic modeling of news headlines, this study enhances understanding of AI's multi-faceted discussions and provides new opportunities for future work. Our analysis is based on a sample of 50,000 headlines, stratified by five domains. While the findings are certainly valid, there is greater breadth across the full dataset of over 288,000 records, so subsequent research should use this large corpus to increase generalizability. The sample used in this study leaned notably toward the positive spectrum in sentiment, which may have influenced the skewed portrayal of domains such as healthcare and robotics as predominantly utopian. Future work should validate sentiment analysis using human-annotated ground truth data by sampling 500–1,000 stratified headlines and multiple annotators, enabling evaluation with standard performance and inter-annotator agreement metrics (e.g., accuracy, F1, Cohen's kappa, Krippendorff's alpha). Analyzing more multilingual headlines and regional sources could give a more balanced picture of global AI discourse. Furthermore, longitudinal analyses across multiple years would allow researchers to observe temporal shifts in sentiment, capturing how public discourse evolves in response to major AI breakthroughs, regulations, or incidents (e.g., data breaches, election interference, medical AI recalls). In parallel with the sentiment analysis, this study explored the use of instruction-tuned models: DeepSeek-R1-Distill-Qwen-1.5B and 7B variants, for emotion classification of AI news headlines. Results showed moderate performance, with 1.5B surprisingly outperforming 7B (57.6 percent versus 31.6 percent) on predefined emotion categories (e.g., Sadness, Anger, Fear, Trust). Allowing the model to self-generate emotion categories led to 30 unique emotional expressions, showing the potential of open-ended approaches to capture affective subtlety. These findings suggest that domain familiarity and instruction clarity may outweigh raw parameter count in emotion classification tasks. Future work could explore few-shot prompting, contrastive examples, and fine-tuning small LLMs with emotion-labeled data to improve performance. Our keyword-based classification approach, while effective in enabling domain-level analysis, is limited by its reliance on static pattern-matching rules. Future iterations could explore deep-learning-based classifiers (e.g., fine-tuned BERT or RoBERTa models) trained on labeled datasets to improve accuracy, especially in overlapping or ambiguous headlines. These models can capture contextual subtleties that are often missed by rule-based approaches. Embedding techniques could also be enhanced by integrating contextual sentence encoders (domain-specific sentence-BERT models) and hierarchical attention networks, particularly to differentiate between multi-topic headlines. Future research may benefit from incorporating multi-modal data sources including full news articles, social media posts, policy papers, and public speeches to triangulate sentiment and emotion trends. This would help validate whether headlines align with the tone of the underlying content and assess how media framing influences public perception. To further enrich our understanding of public opinion beyond traditional news media, future research should incorporate social media platforms such as Reddit and Twitter, where decentralized, real-time discourse often captures emerging sentiments, grassroots reactions, and public skepticism more candidly than curated news headlines. These platforms enable the exploration of bottom-up narratives, emotional volatility, and community-level trends that may not yet surface in mainstream journalism. Sentiment dynamics from social media can also serve as early indicators of public backlash, resistance, or support in response to key AI developments or regulatory shifts. Tracking sentiment over time alongside events could identify inflection points in public opinion. While sentiment analysis offers critical insights into public perceptions of AI, future research should also assess the tangible outcomes of AI deployment across diverse communities. AI-driven innovations can expand access to healthcare, education, and economic opportunities, yet they also risk deepening inequalities, displacing workers, and amplifying bias if deployed without oversight. Community-level studies could reveal disparities in how different populations experience AI's benefits and harms, offering a more complete picture beyond media narratives. Understanding these impacts will be essential to ensuring that AI development promotes equity, sustains public trust, and supports resilient, inclusive communities over the long term.

7. Conclusion

The future of AI is not predetermined but actively constructed through policy choices (Samuel 2021), media narratives, technological design, and public engagement. Recent work also shows how fear-inducing news headlines fuel "AI phobia" and distort public understanding, which in turn shapes reactive policies (Samuel et al. 2025a). Alongside these sociopolitical dynamics, AI agents, agentic AI, and swarm intelligence are emerging focal areas in AI research and practice. The emphasis lies on increasing levels of autonomy, agent-to-agent collaboration, self-learning, and adaptation to environmental variables, along with synchronized functioning to achieve complex goals with minimal human oversight. These paradigms are increasingly recognized for their transformative potential across domains such as scientific discovery, intelligent systems design, and human–machine collaboration (Acharya et al. 2025; Gridach et al. 2025; Sapkota et al. 2026). By critically examining how AI is represented in the public sphere, this research contributes to a more informed, ethical, and inclusive path forward. As AI systems become increasingly embedded in critical infrastructures, educational institutions, workplaces, and personal lives, their societal impact will be determined not by technological inevitability, but by how they are communicated, regulated, and adopted. To ensure this trajectory remains responsible, high-impact AI systems should be subject to independent ethical audits and adversarial "red-team" testing before deployment, while governments and institutions must simultaneously invest in sustained public education initiatives that promote AI literacy to enable citizens to engage critically and democratically with these technologies.

References

Acemoglu, D., and P. Restrepo. 2019. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." *Journal of Economic Perspectives* **33**, no. 2: 3–30. https://doi.org/10.1257/jep.33.2.3

Acharya, D. B., K. Kuppan, and B. Divya. 2025. "Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey." *IEEE Access* **13**: 18912–18936. https://doi.org/10.1109/ACCESS.2025.3532853

Aiken, M., and S. Balan. 2011. "An Analysis of Google Translate Accuracy." *Translation Journal* **16**, no. 2. https://api.semanticscholar.org/CorpusID:89609742

Ali, G. M. N., M. M. Rahman, M. A. Hossain, M. S. Rahman, K. C. Paul, J. C. Thill, et al. 2021. "Public Perceptions of COVID-19 Vaccines: Policy Implications from US Spatiotemporal Sentiment Analytics." *Healthcare* **9**, no. 9: 1110.

Alowais, S. A., S. S. Alghamdi, N. Alsuhebany, T. Alqahtani, A. I. Alshaya, S. N. Almohareb, et al. 2023. "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice." *BMC Medical Education* **23**, no. 689. https://doi.org/10.1186/s12909-023-04698-z

Al-Zahrani, A. M. 2024. "Unveiling the Shadows: Beyond the Hype of AI in Education." *Heliyon* **10**, no. 9: e30696. https://doi.org/10.1016/j.heliyon.2024.e30696

Ananyi, S. O., and E. Somieari-Pepple. 2023. "Cost-Benefit Analysis of Artificial Intelligence Integration in Education Management: Leadership Perspectives." *International Journal of Economics Environmental Development and Society* **4**, no. 3: 353–370. https://ijeeds.com.ng/assets/vol.%2C-4(3)-ananyi—somieari-pepple.pdf

Athaluri, S. A., S. V. Manthena, V. S. R. K. M. Kesapragada, V. Yarlagadda, T. Dave, and R. T. S. Duddumpudi. 2023. "Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing through ChatGPT References." *Cureus* **15**, no. 4. https://doi.org/10.7759/cureus.37432

Baccouri, N. 2020. “Deep Translator: A flexible Python Tool for Translations.” *Python Library*. Accessed February 13, 2025. https://deep-translator.readthedocs.io/en/latest/

Bairathi, A. 2025. “A World with AI: Where Will We Be in 10, 20, and 50 Years?” *NASSCOM Community*. Accessed February 12, 2025. https://community.nasscom.in/communities/ai/world-ai-where-will-we-be-10-20-and-50-years

Bajwa, J., U. Munir, A. Nori, and B. Williams. 2021. “Artificial Intelligence in Healthcare: Transforming the Practice of Medicine.” *Future Healthcare Journal* **8**, no. 2: e188–e194. https://doi.org/10.7861/fhj.2021-0095

Baker, R. S., and A. Hawn. 2022. “Algorithmic Bias in Education.” *International Journal of Artificial Intelligence in Education* **32**: 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Baptista, E. 2025. “What Is DeepSeek and Why Is It Disrupting the AI Sector?” *Reuters*. Accessed February 12, 2025. https://www.reuters.com/technology/artificial-intelligence/what-is-deepseek-why-is-it-disrupting-ai-sector-2025-01-27/

Belanger, A. 2024. “Netflix Doc Accused of Using AI to Manipulate true Crime Story.” *Ars Technica*. https://arstechnica.com/tech-policy/2024/04/netflix-doc-accused-of-using-ai-to-manipulate-true-crime-story/?utm_source=chatgpt.com

Bendiab, G., A. Hameurlaine, G. Germanos, N. Kolokotronis, and S. Shiaeles. 2023. “Autonomous Vehicles Security: Challenges and Solutions Using Blockchain and Artificial Intelligence.” *IEEE Transactions on Intelligent Transportation Systems* **24**, no. 4: 3614–3637

Bharadiya, J. 2023. “Artificial Intelligence in Transportation Systems a Critical Review.” *American Journal of Computing and Engineering* **6**, no. 1: 34.

Bhuyan, S. S., V. Sateesh, N. Mukul, A. Galvankar, A. Mahmood, M. Nauman, et al. 2025. “Generative Artificial Intelligence Use in Healthcare: Opportunities for Clinical Excellence and Administrative Efficiency.” *Journal of Medical Systems* **49**, no. 1. https://doi.org/10.1007/s10916-024-02136-1

Bigham, J. P., and P. Carrington. 2018. “Learning from the Front: People with Disabilities as Early Adopters of AI.” In *Proceedings of the 2018 HCIC Human-Computer Interaction Consortium*, Mellon University, Pittsburgh, PA.

Bird, S., E. Klein, and E. Loper. 2009. “Natural Language Processing with Python.” *O’Reilly Media*. Accessed February 12, 2025. https://www.nltk.org/book/

Bond, S. 2024. “How AI Deepfakes Polluted Elections in 2024.” *NPR*. Accessed February 12, 2025. https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections

Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford, United Kingdom: Oxford University Press.

Bostrom, N. 2024. *Deep Utopia: Life and Meaning in a Solved World*. Washington, DC: IdeaPress Publishing.

Bratton, L. 2025. “What Big Tech Execs Have Said about DeepSeek As US Contemplates Ban.” *Yahoo Finance*. Accessed February 12, 2025. https://finance.yahoo.com/news/what-big-tech-execs-have-said-about-deepseek-as-us-contemplates-ban-140030220.html

Chang, X. 2023. “Gender Bias in Hiring: An Analysis of the Impact of Amazon’s Recruiting Algorithm.” *Advances in Economics, Management and Political Sciences* **23**: 134–140. 10.54254/2754-1169/23/20230367.

Chidipothu, N., R. Anderson, J. Samuel, A. Pelaez, J. Esguerra, and M. N. Hoque. 2025. “Improving Large Language Model (LLM) Performance with Retrieval Augmented Generation (RAG): Development of a Transparent Generative AI University Support System for Educational Purposes.” *Journal of Big Data and Artificial Intelligence* **3**, no 1. https://doi.org/10.54116/jbdai.v3i1.50

Chui, M., J. Manyika, and M. Miremadi. 2016. “Where Machines Could Replace Humans—and Where They Can’t (Yet).” *McKinsey Quarterly*. Accessed February 12, 2025. https://www.mckinsey.com/featured-insights/employment-and-growth/where-machines-could-replace-humans-and-where-they-cant-yet

Chustecki, M. 2024. “Benefits and Risks of AI in Health Care: Narrative Review.” *Interactive Journal of Medical Research* **13**: e53616. https://doi.org/10.2196/53616

Cools, H., B. Van Gorp, and M. Opgenhaffen. 2022. “Where Exactly between Utopia and Dystopia? A Framing Analysis of AI and Automation in US Newspapers.” *Journalism*. Accessed February 12, 2025. https://doi.org/10.1177/14648849221122647

Cordero, D. 2024. “The Downsides of Artificial Intelligence in Healthcare.” *The Korean Journal of Pain* **37**, no. 1: 87–88. https://doi.org/10.3344/kjp.23312

Cross, S., I. Bell, J. Nicholas, L. Valentine, S. Mangelsdorf, S. Baker, N. Titov, et al. 2024. “Use of AI in Mental Health Care: Community and Mental Health Professionals Survey.” *JMIR Mental Health* **11**: e60589. https://doi.org/10.2196/60589

Cuthbertson, A. 2024. “AI and the Meaning of Life: Philosopher Nick Bostrom Says Technology Could Bring Utopia but Will Force Us to Rethink Our Purpose.” *The Independent*. Accessed February 12, 2025. https://www.the-independent.com/tech/ai-deep-utopia-nick-bostrom-cockaigne-b2530807.html

DeepSeek-AI, D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, et al. 2025. “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning.” Preprint, submitted January 22. arXiv Preprint arXiv:2501.12948. https://doi.org/10.48550/arXiv.2501.12948

Dehghan, A., M. Cevik, and M. Bodur. 2023. "Dynamic AGV Task Allocation in Intelligent Warehouses." Preprint, submitted December 26. arXiv Preprint arXiv:2312.16026. https://doi.org/10.48550/arXiv.2312.16026

Didast, F., R. Y. Nassih, and I. Ait Lbachir. 2024. "Artificial Intelligence and Logistics: Recent Trends and Development." *International Journal of Advanced Computer Science and Applications* **12**, no. 1. https://api.semanticscholar.org/CorpusID:281326771

Dmitracova, O. 2025. "41% of Companies Worldwide Plan to Reduce Workforces by 2030 Due to AI." *CNN*. Accessed February 12, 2025. https://www.cnn.com/2025/01/08/business/ai-job-losses-by-2030-intl/index.html

Dutta, S., S. Ranjan, S. Mishra, V. Sharma, P. Hewage, and C. Iwendi. 2024. "Enhancing Educational Adaptability: A Review and Analysis of AI-Driven Adaptive Learning Platforms." *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Noida, India, February 21–23. https://doi.org/10.1109/iciptm59628.2024.10563448

Elysée Palace. 2025. "Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet." Accessed February 12, 2025. https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet

Ettman, C. K., and S. Galea. 2023. "The Potential Influence of AI on Population Mental Health." *JMIR Mental Health* **10**: e49936. https://doi.org/10.2196/49936

Faluyi, S. E. 2025. "AI and Job Market: Analysing the Potential Impact of AI on Employment, Skills, and Job Displacement." *African Journal of Marketing Management* **17**, no. 1: 1–8. https://doi.org/10.5897/AJMM2024.0747

Farhud, D. D., and S. Zokaei. 2021. "Ethical Issues of Artificial Intelligence in Medicine and Healthcare." *Iranian Journal of Public Health* **50**, no. 11: i–v. https://doi.org/10.18502/ijph.v50i11.7600

Frank, M. R., D. Autor, J. E. Bessen, E. Brynjolfsson, M. Cebrian, D. J. Deming, et al. 2019. "Toward Understanding the Impact of Artificial Intelligence on Labor." *Proceedings of the National Academy of Sciences* **116**, no. 14: 6531–6539. https://doi.org/10.1073/pnas.1900949116

Gammon, C., and M. Bornstein. 2018. "John Henry Effect." In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*, Frey, B. B., editor, volume **4**. Thousand Oaks, CA: SAGE Publications, Inc. https://doi.org/10.4135/9781506326139.n364

Garvey, M. D., J. Samuel, and A. Pelaez. 2021. "Would You Please Like My Tweet?! An Artificially Intelligent, Generative Probabilistic, and Econometric Based System Design for Popularity-Driven Tweet Content Generation." *Decision Support Systems* **113497**. https://doi.org/10.1016/j.dss.2021.113497

Girden, E. R. 1992. *ANOVA: Repeated Measures*. Thousand Oaks, CA: SAGE Publications, Inc.

Gmyrek, P., H. Winkler, and S. Garganta. 2024. "Buffer or Bottleneck? Employment Exposure to Generative AI and the Digital Divide in Latin America." The World Bank (ILO Working paper 121). https://doi.org/10.54394/TFZY7681

Google. n.d. "Google News RSS Feeds." Accessed February 13, 2025. https://news.google.com/rss

Grand View Research. 2024. "AI Companion Market Size, Share & Trends Analysis Report by Type (Text-Based, Voice-Based, Multi-Modal), by Application (Mental Health Support, Education & Learning Aid), by Industry Vertical (Consumer, Businesses, Healthcare), and by Region Forecasts, 2025–2030 (Report No. GVR-4-68040-517-3). *Grand View Research*. Accessed February 12, 2025. https://www.grandviewresearch.com/industry-analysis/ai-companion-market-report

Gridach, M., J. Nanavati, K. Abidine, L. Mendes, and C. Mack. 2025. "Agentic AI for Scientific Discovery: A Survey of Progress, Challenges, and Future Directions." Preprint, submitted March 12. arXiv Preprint arXiv:2503.08979. https://doi.org/10.48550/arXiv.2503.08979

Grootendorst, M. 2022. "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." Preprint, submitted March 11. arXiv Preprint arXiv:2203.05794. https://arxiv.org/abs/2203.05794

Groves, M., and K. Mundt. 2015. "Friend or Foe? Google Translate in Language for Academic Purposes." *English for Specific Purposes* **37**: 112–121. https://doi.org/10.1016/j.esp.2014.09.001

Gupta, M., I. S. Kakar, M. Peden, E. Altieri, and J. Jagnoor. 2021. "Media Coverage and Framing of Road Traffic Safety in India." *BMJ Global Health* **6**, no. 3: e004499. https://doi.org/10.1136/bmjgh-2020-004499

Hill, D. L. G. 2024. "AI in Imaging: The Regulatory Landscape." *British Journal of Radiology* **97**, no. 1155: 483–491. https://doi.org/10.1093/bjr/tqae002

Hoose, S., and K. Králiková. 2024. "Artificial Intelligence in Mental Health Care: Management Implications, Ethical Challenges, and Policy Considerations." *Administrative Sciences* **14**, no. 9: 227. https://doi.org/10.3390/admsci14090227

Hu, Q., and H. Rangwala. 2020. "Towards Fair Educational Data Mining: A Case Study on Detecting At-Risk Students." *International Educational Data Mining Society*. Accessed February 12, 2025. https://api.semanticscholar.org/CorpusID:220927743

International Data Corporation (IDC). 2024. "Artificial Intelligence Will Contribute $19.9 Trillion to the Global Economy through 2030 and Drive 3.5% of Global GDP in 2030." *IDC*. https://www.idc.com/getdoc.jsp?containerId=prUS52600524

Jackson, J., and Paste Staff. 2023. "The 50 Best Dystopian Movies of All Time." *Paste Magazine*. Accessed February 12, 2025. https://www.pastemagazine.com/movies/dystopian-movies/best-dystopian-movies-of-all-time-1

Jamieson, T., and D. A. Van Belle. 2019. "How Development Affects News Media Coverage of Earthquakes: Implications for Disaster Risk Reduction in Observing Communities." *Sustainability* **11**, no. 7: 1970. https://doi.org/10.3390/su11071970

jmuwa. 2020. "The Best Dystopian TV Shows." *IMDb*. Accessed February 12, 2025. https://www.imdb.com/list/ls048004810/

Johnston, C. 2020. "How Social Media Algorithms Inherently Create Polarization." *Psychology Today*. https://www.psychologytoday.com/us/blog/cultural-psychiatry/202011/how-social-media-algorithms-inherently-create-polarization

Jones, K. S. 1972. "A Statistical Interpretation of Term Specificity and Its Application in Retrieval." *Journal of Documentation* **28**, no. 1: 11–21. https://doi.org/10.1108/eb026526

Jumaev, G. 2024. "The Impact of AI on Job Market: Adapting to the Future of Work." *Zenodo*. https://doi.org/10.5281/zenodo.10467813

Jumper, J., R. Evans, A. Pritzel, et al. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* **596**: 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kanclerz, K., P. Milkowski, J. Kocoń, and P. Miłkowski. 2022. "Zero-Shot Emotion Detection for Semi-Supervised Sentiment Analysis Using Sentence Transformers and Ensemble Learning." *Applied Sciences* **12**, no. 17: 8662. https://doi.org/10.3390/app12178662

Kashyap, R., Y. Samuel, L. W. Friedman, and J. Samuel. 2024. "Artificial Intelligence Education and Governance—Human Enhancive, Culturally Sensitive and Personally Adaptive HAI." *Frontiers in Artificial Intelligence* **7**: 1443386. https://www.frontiersin.org/articles/10.3389/frai.2024.1443386

Khosla, V. 2024. "AI: Dystopia or Utopia?" *Khosla Ventures*. Accessed February 12, 2025. https://www.khoslaventures.com/ai-dystopia-or-utopia/

Klepper, D., and A. Swenson. 2023. "AI-Generated Disinformation Poses Threat of Misleading Voters in 2024 Election." *PBS NewsHour*. Accessed February 12, 2025. https://www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of-misleading-voters-in-2024-election

Kruskal, W. H., and W. A. Wallis. 1952. "Use of Ranks in One-Criterion Variance Analysis." *Journal of the American Statistical Association* **47**, no. 260: 583–621. https://doi.org/10.1080/01621459.1952.10483441

Lewis, M., Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, et al. 2020. "BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Virtual, July. https://doi.org/10.18653/v1/2020.acl-main.703

Li, D., W. He, and Y. Guo. 2021. "Why AI Still Doesn't Have Consciousness?" *CAAI Transactions on Intelligence Technology* **6**, no. 2: 175–179. https://doi.org/10.1049/cit2.12035

Littman, M. L., I. Ajunwa, G. Berger, C. Boutilier, M. Currie, F. Doshi-Velez, et al. 2021. "Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report." Preprint, submitted October 27. arXiv Preprint arXiv:2210.15767. https://doi.org/10.48550/arXiv.2210.15767

Mann, H. B., and D. R. Whitney. 1947. "On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other." *Annals of Mathematical Statistics* **18**, no. 1: 50–60.

Marcinek, K., K. D. Stanley, G. Smith, P. Cormarie, and S. Gunashekar. 2024. "Risk-Based AI Regulation: A Primer on the Artificial Intelligence Act of the European Union." *RAND Corporation*. https://www.rand.org/pubs/research_reports/RRA3243-3.html

McInnes, L., J. Healy, and S. Astels. 2017. "hdbscan: Hierarchical Density Based Clustering." *The Journal of Open Source Software* **2**, no. 11: 205. https://doi.org/10.21105/joss.00205

McInnes, L., J. Healy, and J. Melville. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." Preprint, submitted September 18. arXiv Preprint arXiv:1802.03426. https://arxiv.org/abs/1802.03426

McKee, K., and M. Pilgrim. 2010. "Universal Feed Parser (feedparser)." *Python library*. Accessed February 13, 2025. https://feedparser.readthedocs.io/en/latest/

Mhlanga, D. 2023. "Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning." (Working paper). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354422

Mill, J. S. 1988. *Public and Parliamentary Speeches – Part I – November 1850 – November 1868*. Toronto, Canada: University of Toronto Press.

Mishra, V. 2024. "Unchecked AI Threatens Democracy, Warns UN Chief." *United Nations News*. Accessed February 12, 2025. https://news.un.org/en/story/2024/09/1154316

Mitchell, M. 2022. "What Does It Mean to Align AI with Human Values?" *Quanta Magazine*. Accessed February 12, 2025. https://www.quantamagazine.org/what-does-it-mean-to-align-ai-with-human-values-20221213/

More, T. 1949. *Utopia*. New York, NY: Appleton-Century-Crofts.

Nelson, L. 2024. “Practical AI Limitations You Need to Know.” *AFA Education Blog*. Accessed February 12, 2025. https://afaeducation.org/blog/practical-ai-limitations-you-need-to-know/

Nguyen, B. 2024. “Donald Trump, Elon Musk, Taylor Swift, and More: The 10 Most AI Deepfaked People Right Now.” *Quartz*. Accessed February 12, 2025. https://qz.com/donald-trump-elon-musk-taylor-swift-beyonce-ai-deepfake-1851666681

O’Gieblyn, M. 2023. “Does AI Have a Subconscious?” *WIRED*. Accessed February 12, 2025. https://www.wired.com/story/does-ai-have-a-subconscious/

Ouchchy, L., A. Coin, and V. Dubljević. 2020. “AI in the Headlines: The Portrayal of the Ethical Issues of Artificial Intelligence in the Media.” *AI & Society* **35**: 927–936. https://doi.org/10.1007/s00146-020-00965-5

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al. 2011. “Scikit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research* **12**: 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

Pelaez, A., A. Jacobsen, J. Samuel, and K. Sengupta. 2025. “Ethics of AI in Education.” *The 54th Annual Northeast Decision Sciences Institute Conference (NEDSI 2025)*, Hershey, PA, March 27–29.

Pierce, D. 2024. “Two Possible Futures for AI.” *The Verge*. Accessed February 12, 2025. https://www.theverge.com/2024/10/29/24282333/ai-vision-anthropic-openai-shakealert-vergecast

Piocciochi, M., and R. A. Alwabel. 2020. “Leveraging Artificial Intelligence in Education: Current Applications and Future Prospects.” *ResearchGate*. Accessed February 14, 2025. https://www.researchgate.net/publication/384016811_Leveraging_Artificial_Intelligence_in_Education_Current_Applications_and_Future_Prospects

Price, W. N. II. 2019. “Risks and Remedies for Artificial Intelligence in Health Care.” *The Brookings Institution*. Accessed February 12, 2025. https://www.brookings.edu/articles/risks-and-remedies-for-artificial-intelligence-in-health-care/

Rahman, M. M., G. M. N. Ali, X. J. Li, J. Samuel, K. C. Paul, P. H. Chong, et al. 2021. “Socioeconomic Factors Analysis for COVID-19 US Reopening Sentiment with Twitter and Census Data.” *Heliyon* e06200.

Rahman-Jones, I. 2024. “UK Watchdog Looking into Microsoft AI Taking Screenshots.” *BBC News*. Accessed February 12, 2025. https://www.bbc.com/news/articles/cpwwqp6nx14o

Rainer, R. K., Jr., R. G. Richey, Jr., and S. Chowdhury. 2025. “How Robotics Is Shaping Digital Logistics and Supply Chain Management: An Ongoing Call for Research.” *Journal of Business Logistics* **46**, no. 1: e70005.

Rainie, L., and J. Anderson. 2024. *Experts Imagine the Impact of Artificial Intelligence by 2040. Imagining the Digital Future Center*. Elon, NC: Elon University.

Randieri, C. 2023. “Unveiling the Role of AI Algorithms: Unmasking Societal Inequities and Cultural Prejudices.” *Forbes Technology Council*. Accessed February 12, 2025. https://www.forbes.com/councils/forbestechcouncil/2023/07/19/unveiling-the-role-of-ai-algorithms-unmasking-societal-inequities-and-cultural-prejudices/

Reitz, K. n.d. “Requests: HTTP for Humans.” *Python Library*. Accessed February 13, 2025. https://docs.python-requests.org/en/latest/

Reynaud, F., and M. Untersinger. 2024. “Paris 2024: Controversial AI-Led Video Surveillance Put to the Test during Olympics.” *Le Monde*. Accessed February 12, 2025. https://www.lemonde.fr/en/pixels/article/2024/07/24/paris-2024-controversial-ai-led-video-surveillance-put-to-the-test-during-olympics_6697267_13.html

Richardson, L. 2023. “Beautiful Soup Documentation.” *Python Library. Crummy*. Accessed February 13, 2025. https://www.crummy.com/software/BeautifulSoup/

Richey, R. G., S. Chowdhury, B. Davis-Sramek, M. Giannakis, and Y. K. Dwivedi. 2023. “Artificial Intelligence in Logistics and Supply Chain Management: A Primer and Roadmap for Research.” *Journal of Business Logistics* **44**, no. 4: 532–549. https://doi.org/10.1111/jbl.12364

Robins-Early, N. 2024. “Trump Posts Deepfakes of Swift, Harris, and Musk in Effort to Shore Up Support.” *The Guardian*. Accessed February 12, 2025. https://www.theguardian.com/us-news/article/2024/aug/19/trump-ai-swift-harris-musk-deepfake-images?utm_source=chatgpt.com

Robison, K. 2024. “Anthropic’s CEO Thinks AI Will Lead to a Utopia—He Just Needs a Few Billion Dollars First.” *The Verge*. Accessed February 12, 2025. https://www.theverge.com/2024/10/16/24268209/anthropic-ai-dario-amodei-agi-funding-blog

Rodilosso, E. 2024. “Filter Bubbles and the Unfeeling: How AI for Social Media Can Foster Extremism and Polarization.” *Philosophy & Technology* **37**, no. 71. https://doi.org/10.1007/s13347-024-00758-4

Rolf, B., I. Jackson, M. Müller, S. Lang, T. Reggelin, and D. Ivanov. 2022. “A Review on Reinforcement Learning Algorithms and Applications in Supply Chain Management.” *International Journal of Production Research* **61**, no. 20: 7151–7179. https://doi.org/10.1080/00207543.2022.2140221

Rundle, J. 2025. “New York State Bans DeepSeek from Government Devices: State Says App Raises Serious Security and Censorship Concerns.” *The Wall Street Journal*. Accessed February 12, 2025. https://www.wsj.com/articles/new-york-state-bans-deepseek-from-government-devices-de7a9df4

Sadaf, M., Z. Iqbal, A. R. Javed, I. Saba, M. Krichen, S. Majeed, et al. 2023. "Connected and Automated Vehicles: Infrastructure, Applications, Security, Critical Challenges, and Future Aspects." *Technologies* **11**, no. 5: 117.

SafeTREC. n.d. "The Role of Media and Road Safety." *California Active Transportation Safety Information Pages (CATSIP)*. Accessed February 12, 2025. https://catsip.berkeley.edu/resources/role-media-and-road-safety

Samuel, J. 2021. "A Call for Proactive Policies for Informatics and Artificial Intelligence Technologies." *Scholars Strategy Network*. Accessed February 12, 2025. https://scholars.org/contribution/call-proactive-policies-informatics-and

Samuel, J. 2023. "The Critical Need for Transparency and Regulation amidst the Rise of Powerful Artificial Intelligence Models." *Scholars Strategy Network*. Accessed April 2023. https://scholars.org/contribution/critical-need-transparency-and-regulation

Samuel, J., G. G. Ali, M. Rahman, E. Esawi, and Y. Samuel. 2020a. "Covid-19 Public Sentiment Insights and Machine Learning for Tweets Classification." *Information* **11**, no. 6: 314.

Samuel, J., R. Kashyap, Y. Samuel, and A. Pelaez. 2022. "Adaptive Cognitive Fit: Artificial Intelligence Augmented Management of Information Facets and Representations." *International Journal of Information Management* **65**: 102505. https://doi.org/10.1016/j.ijinfomgt.2022.102505

Samuel, J., T. Khanna, J. Esguerra, S. Sundar, A. Pelaez, and S. S. Bhuyan. 2025a. "The Rise of Artificial Intelligence Phobia! Unveiling News-Driven Spread of AI Fear Sentiment Using ML, NLP, and LLMs." *IEEE Access* **13**: 125944–125969. https://doi.org/10.1109/ACCESS.2025.3588179

Samuel, J., M. Rahman, G. G. Nawaz Ali, Y. Samuel, A. Pelaez, P. H. J. Chong, et al. 2020b. "Feeling Positive About Reopening? New Normal Scenarios From COVID-19 US Reopen Sentiment Analytics." *IEEE Access* **8**: 142173–142190. https://ieeexplore.ieee.org/document/9154672

Samuel, J., A. Tripathi, and E. Mema. 2024. "A New Era of Artificial Intelligence Begins. . .Where Will It Lead Us?" *Journal of Big Data and Artificial Intelligence* **2**, no. 1.

Samuel, J., M. Ye, T. Khanna, Y. Yao, X. Lin, R. Anderson, and A. Gupta. 2025b. "Finder, Evaluator, Explainer, Generator (FEEG): A Bloom's Taxonomy-Based Query Classification Framework for LLMs and Generative AI." *Expert Systems with Applications*. Accessed December 16, 2025. https://doi.org/10.1016/j.eswa.2025.130755

Samuel, Y., M. Brennan-Tonetta, J. Samuel, R. Kashyap, V. Kumar, S. K. Kaashyap, et al. 2023. "Cultivation of Human Centered Artificial Intelligence: Culturally Adaptive Thinking in Education for AI (CATE-AI)." *Frontiers in Artificial Intelligence* **6**: 1198180. https://doi.org/10.3389/frai.2023.1198180

Sapkota, R., K. I. Roumeliotis, and M. Karkee. 2026. "AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges." *Information Fusion* **126**, no. Part B: 103599. https://doi.org/10.1016/j.inffus.2025.103599

ScrapingBee. n.d. ScrapingBee API documentation. Accessed February 13, 2025. https://www.scrapingbee.com/documentation/

Service, R. F. 2018. "Google's DeepMind Aces Protein Folding: Artificial Intelligence Firm Takes Crown in Biannual Contest." *Science*. Accessed February 12, 2025. https://www.science.org/content/article/google-s-deepmind-aces-protein-folding

Shapiro, S. S., and M. B. Wilk. 1965. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* **52**, no. 3–4: 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Shermer, M. 2024. "When It Comes to AI, Think Protopia, Not Dystopia or Utopia." *Skeptic*. Accessed February 12, 2025. https://www.skeptic.com/reading_room/artificial-intelligence-think-protopia-not-dystopia-or-utopia/

Siafakas, N., and E. Vasarmidi. 2024. "Risks of Artificial Intelligence (AI) in Medicine." *Pneumon* **37**, no. 3: 40. https://doi.org/10.18332/pne/191736

Silver, N. S. 2023. "AI Utopia and Dystopia: What Will the Future Have in Store?" *Forbes*. Accessed February 12, 2025. https://www.forbes.com/sites/nicolesilver/2023/06/20/ai-utopia-and-dystopia-what-will-the-future-have-in-store-artificial-intelligence-series-5-of-5/

Singal, P. 2024. "Nick Bostrom Discusses Superintelligence, AI, and Deep Utopia in Dinis Guarda YouTube Podcast." *IntelligentHQ*. Accessed February 12, 2025. https://www.intelligenthq.com/nick-bostrom-discusses-superintelligence-ai-and-deep-utopia-in-dinis-guarda-youtube-podcast/

Smith, R. A. 2024. "AI Is Starting to Threaten White-Collar Jobs. Few Industries Are Immune." *The Wall Street Journal*. Accessed February 12, 2025. https://www.wsj.com/lifestyle/careers/ai-is-starting-to-threaten-white-collar-jobs-few-industries-are-immune-9cdbcb90

Sodiya, E. O., U. J. Umoga, O. O. Amoo, and A. Atadoga. 2024. "AI-Driven Warehouse Automation: A Comprehensive Review of Systems." *GSC Advanced Research and Reviews* **18**, no. 2: 272–282. https://doi.org/10.30574/gscarr.2024.18.2.0063

Swarns, C. 2023. "When Artificial Intelligence Gets It Wrong." *Innocence Project*. Accessed February 12, 2025. https://innocenceproject.org/when-artificial-intelligence-gets-it-wrong/

The Guardian. 2024. "Revealed: Bias Found in AI System Used to Detect UK Benefits Fraud." Accessed February 12, 2025. https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits

The Week UK. 2024. “Future of Generative AI: Utopia, Dystopia or Up To Us?” *The Explainer*. Accessed February 12, 2025. https://theweek.com/tech/future-of-generative-ai-utopia-dystopia-or-up-to-us

Tony Blair Institute for Global Change. 2024. “The Impact of AI on the Labour Market.” Accessed February 12, 2025. https://institute.global/insights/economic-prosperity/the-impact-of-ai-on-the-labour-market

Tripathi, A., J. Samuel, M. Brennan-Tonetta, H. Nguyen, and E. Mema. 2025. “When Machines Create! Envisioning our Future as Shaped by the Transformative Power of Generative AI.” *Journal of Big Data and Artificial Intelligence* **3**, no. 1. https://doi.org/10.54116/jbdai.v3i1.55

United Nations Regional Information Centre (UNRIC). 2024. “Can Artificial Intelligence (AI) Influence Elections?” Accessed February 12, 2025. https://unric.org/en/can-artificial-intelligence-ai-influence-elections/

U.S. Equal Employment Opportunity Commission (EEOC). 2023. “iTutorGroup to pay $365,000 to Settle EEOC Discriminatory Hiring Suit.” *EEOC Newsroom*. Accessed February 12, 2025. https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit

UspeakGreek. 2023. “Etymology and Meaning of the Word Dystopian.” Accessed February 12, 2025. https://uspeakgreek.com/art/literature/etymology-and-meaning-of-word-dystopian/

Utopia & Dystopia. n.d. “List of Famous Utopian Movies.” Accessed February 12, 2025. https://www.utopiaanddystopia.com/utopian-fiction/utopian-movies-list/

van der Maas, M., and J. Samuel. 2025. “Online Gambling Forumsas a Potential Target for Harm Reduction: An Exploratory Natural Language Processing Analysis of a Reddit.com Forum.” *Harm Reduction Journal* **22**: 77. https://doi.org/10.1186/s12954-025-01220-0

Vinson, D. W., M. Arcan, D. Niland, and F. Delahunty. 2024. “Towards Sustainable Workplace Mental Health: A Novel Approach to Early Intervention and Support.” Preprint, submitted February 2. arXiv Preprint abs/2402.01592. https://arxiv.org/abs/2402.01592

Wankhade, M., A. C. S. Rao, and C. Kulkarni. 2022. “A Survey on Sentiment Analysis Methods, Applications, and Challenges.” *Artificial Intelligence Review* **55**, no. 7: 5731–5780.

Webb, M. 2019. “The Impact of Artificial Intelligence on the Labor Market.” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3482150

Williams, A., N. Nangia, and S. R. Bowman. 2018. “A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference.” *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, LA, June. https://doi.org/10.18653/v1/N18-1101

Winton, A. 2017. “Get a Horse! America’s Skepticism Toward the First Automobiles.” *The Saturday Evening Post*. Accessed February 12, 2025. https://www.saturdayeveningpost.com/2017/01/get-horse-americas-skepticism-toward-first-automobiles/

Wladawsky-Berger, I. 2017. “The Emerging, Unpredictable Age of AI.” *MIT Initiative on the Digital Economy*. Accessed February 12, 2025. https://ide.mit.edu/insights/the-emerging-unpredictable-age-of-ai/

Wolf, T., L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, et al. 2020. “Transformers: State-of-the-Art Natural Language Processing.” *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Virtual, November 16–20. https://doi.org/10.18653/v1/2020.emnlp-demos.6

World Economic Forum (WEF). 2025. “The Future of Jobs Report 2025.” *WEF*. Accessed February 12, 2025. https://www.weforum.org/publications/the-future-of-jobs-report-2025/

Younge, H. L. 1747. *Utopia: Or Apollo’s Golden Days*. Dublin, Ireland: Ptd. by George Faulkner.

Yudkowsky, E. 2023. “Pausing AI Developments Isn’t Enough. We Need to Shut It All Down.” *Time*. Accessed February 12, 2025. https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/

Zajko, M. 2022. “Artificial Intelligence, Algorithms, and Social Inequality: Sociological Contributions to Contemporary Debates.” *Sociology Compass* **16**, no. 3. https://doi.org/10.1111/soc4.12962

Zaman, B. U. 2023. “Transforming Education through AI, Benefits, Risks, and Ethical Considerations.” *Transforming Education through AI, Benefits, Risks, and Ethical Considerations*. Accessed February 12, 2025. https://doi.org/10.36227/techrxiv.24231583.v1

Zhai, C., S. Wibowo, and L. D. Li. 2024. “The Effects of Over-Reliance on AI Dialogue Systems on Students’ Cognitive Abilities: A Systematic Review.” *Smart Learning Environments* **11**, no. 1: 28.

Zhang, W., Y. Deng, B. Liu, S. J. Pan, and L. Bing. 2023. “Sentiment Analysis in the Era of Large Language Models: A Reality Check.” Preprint, submitted May 24. arXiv Preprint:2305.15005.

Ziegler, B. 2024. “It’s the Year 2030: What Will Artificial Intelligence Look Like?” *The Wall Street Journal*. Accessed February 12, 2025. https://www.wsj.com/article/ai-future-2030

Journal of Big Data and Artificial Intelligence

WWW.JBDAI.ORG

ISSN: 2692-7977

JBDAI Vol. 4, No. 1, 2026 DOI: 10.54116/jbdai.v4i1.64

SUDOKUIMPUTER: A BEST-IN-CLASS GRAPH FRAMEWORK FOR MNAR AND MAR MISSING DATA IMPUTATION

Vivek Nandhan Kanpa
University College Dublin
vivek.kanpa@icahn.mssm.edu

Riddhishrree Badhan
Washington University
in St. Louis
b.riddhishrree@wustl.edu

Vineeth Kanpa
Northeastern University
kanpa.vi@northeastern.edu

ABSTRACT

Missing data are a ubiquitous challenge when training unbiased, high confidence machine learning (ML) models on real-world data. To help address this issue, we propose SudokuImputer v1.0.0, a novel graph-based framework, to estimate missing values in an iterative, uncertainty aware process. The behavior of SudokuImputer is assessed across numerous design hyperparameters, including multiple network centrality methods, feature prioritization modes, statistical associations and pair-wise availability ratios for edge weight assignment, and partner node proportions. We benchmark SudokuImputer against point-value imputations, MICE, kNN, matrix factorization, and SoftImpute. SudokuImputer achieves best-in-class RMSE on MAR (mean rank = 1.7) and MNAR (mean rank = 1.8) missing data across most missingness proportions from ten percent to fifty percent in three experimental benchmarks. SudokuImputer is sensitive to dataset dimensionality, and optimizing algorithmic runtime remains an unresolved challenge. Future work should evaluate SudokuImputer on mixed-data benchmarks and seek to iterate on the foundational graph framework laid here.

Keywords *imputation*, *constraint satisfaction*, *data preprocessing*, *graph network*, *MNAR*, *MAR*, *benchmarking*.

1. Introduction

Large, high-quality datasets play an instrumental role in training effective machine learning (ML) models and artificial intelligence (AI) systems. Specifically, achieving state-of-the-art (SOTA) performance with deep learning networks is dependent on extensive, heterogeneous training data to match the learning capacity of large models

(Hestness et al. 2017; Kaplan et al. 2020; Hoffmann et al. 2026). Aggregating and cleaning datasets is a non-trivial and ubiquitous challenge across AI/ML domains, due to the widespread presence of missing feature data in real-world datasets (Kang 2013; Zhang 2016; Schouten and Vink 2018; Misztal 2019). Missing data can introduce bias, reduce model performance, and compromise the generalizability of predictions if not properly handled.

It has been shown that low data quality, including incomplete data, can lead to revenue losses from 8% to 12% for service organizations (Prasad 2024). One example of its repercussions can be seen in the context of public health. Vaccination databases often have incomplete entries for various reasons such as data entry errors and difference protocols between clinics. During outbreaks of diseases, missing vaccination fields cause underestimated risk zones, delaying appropriate responses (Brown 2025). In another instance, cameras and sensors on roads often drop frames. These missing readings can break traffic-flow prediction models and cause reduced accuracy in navigation systems. Imputation algorithms mitigate these challenges by replacing missing data with numerical estimates (Sun et al. 2022).

2. Literature Review

2.1 Classes of Data Missingness

Rubin establishes three classes of missing data mechanisms, delineated by the implicit biases driving their missingness (Rubin 1976). Missing Completely at Random (MCAR) describes cases where the probability of an absent data point is independent of both observed and unobserved data. Missing at Random (MAR) describes data whose probability of being missing is dependent only on observed data and is independent of absent data. Finally, Missing Not at Random (MNAR) describes data whose missingness is dependent on unobserved data, but is independent of observed data (Table 1) (Heymans and Twisk 2022). Discarding entries with missing values can severely reduce a dataset's size and introduce bias, especially in MAR or MNAR scenarios. Therefore, understanding the causal probabilities that drive data missingness is a crucial consideration when identifying and designing unbiased approaches to approximate missing data.

2.2 A Brief Overview of Commonly Used Imputation Methods

Imputation is the process of inferring synthetic estimates for missing values (Table 2) (Heymans and Twisk 2022). Some of the simplest imputation strategies include case substitution and point-value imputation, where missing values are filled with either a randomly selected feature value or the mean, median, or mode of the feature data

Table 1: Table characterizing and comparing the three classes of data missingness: MCAR, MNAR, and MAR.

Missingness Type	Definition	Key Assumptions	Bias Implications	Handling Strategies
MCAR (Missing Completely at Random)	Missingness is independent of both observed and unobserved data.	No systematic relationship between missingness and data values.	No bias introduced; loss of efficiency only (smaller sample size).	Complete-case analysis, simple imputation (mean/median), advanced methods optional.
MAR (Missing at Random)	Missingness depends only on observed data (not on the missing values).	Missingness can be explained fully by variables that are present.	Can bias results if not modeled correctly; unbiased if the missingness mechanism is accounted for.	Multiple imputation, maximum likelihood, model-based imputation.
MNAR (Missing Not at Random)	Missingness depends only on unobserved values; independent of observed values.	Missingness mechanism not captured by observed covariates.	Highest risk of bias; standard methods may fail.	Sensitivity analysis, explicit modeling of missingness process, specialized algorithms.

Table 2: Characterization and comparison of various numerical imputation strategies.

Method	Mechanism	Assumptions	Advantages	Limitations
Point Value Imputation	Replace missing values with feature mean, median, or mode	Missingness is random; central tendency adequately represents data	Extremely fast; preserves dataset size; easy to implement	Distorts variance and correlations; underestimates uncertainty
SoftImpute (SVD-based)	Low-rank matrix completion via iterative soft-thresholded SVD	Data lies near a low-rank structure	Captures global feature relationships	Sensitive to rank choice; struggles with nonlinear interactions
Matrix Factorization (e.g., NMF, PMF)	Factorize data matrix into latent factors, then reconstruct missing entries	Data approximates low-dimensional latent structure	Good for collaborative filtering and sparse data	May ignore local feature correlations; requires tuning latent dimension
kNN Imputation	Replace missing values with weighted average of neighbors	Similar samples have similar values	Captures local structure; non-parametric; easy to explain	Bias in sparse/ high-dimensional data
MICE (Multiple Imputation by Chained Equations)	Iteratively imputes each variable using regression on other variables, generating multiple datasets	Relationships can be modeled linearly (extensions exist for non-linear)	Accounts for uncertainty; flexible; widely used in biomedical research	Slow on large/ high-dimensional data; requires careful model choice per variable
MissForest	Iterative imputation using Random Forests	Tree ensembles approximate relationships well	Nonlinear + mixed data support; often strong empirical performance	Computationally intensive; can be overfit with small samples
Variational Autoencoder (VAE) Imputation	Learns latent representation of data, imputes by reconstructing missing values	Data distribution is well-approximated by latent space	Captures complex relationships; scalable; probabilistic imputations	Requires tuning architecture; sensitive to training instability
GAIN (Generative Adversarial Imputation Networks)	Adversarial training; generator imputes and discriminator distinguishes real vs imputed	Missingness patterns can be modeled through adversarial training	Performs well on complex data; captures uncertainty; flexible	Training instability; sensitive to hyperparameters; computationally heavy

(Alam et al. 2023). Hot and cold deck imputation strategies are like point-value imputation, except they are preceded by partitioning the available data into clusters and then associating missing data with a certain cluster (Monard 2002). While straightforward and computationally efficient, these methods can distort data distributions and fail to preserve relationships between features, particularly in complex, high-dimensional datasets (Alam et al. 2023). ML-based imputation strategies often preserve the data's variance from the ground truth distribution better than point-based methods. For instance, the lazy learning algorithm k-Nearest Neighbors (kNN) predicts missing values by referencing the non-missing values of its closest existing neighbors across the full feature space (Halder et al. 2024). Regression imputation iterates across each feature as a linear function of all other features, allowing each column's missing values to be predicted using known values in the rest of the feature set (Zhang 2016). MissForest, an abstraction of the Random Forest algorithm, bootstraps decision trees for non-linear data imputation using pre-determined data partitioning and stop criteria (Schonlau and Zou 2020). Multivariate imputation by chained equations (MICE) begins by imputing missing values with point-value estimates in all but one feature.

Then, once that feature is imputed using a regression model, the model is refitted on newly imputed data, and another feature can be imputed using the updated model. This process repeats across all features and is done in multiple iterations until the imputed values stabilize (Mera-Gaona et al. 2021). Matrix factorization-based methods instead approximate the observed dataset as the product of two lower-rank matrices, leveraging latent structure to estimate missing values via reconstruction. SoftImpute employs iterative singular value decomposition (SVD) to a convex optimization problem by minimizing the nuclear norm (the sum of singular values) of the completed matrix. SoftImpute is reportedly less prone to overfitting than matrix factorization methods, better preserves global correlation structure, and is widely regarded as a gold-standard approach for SVD-based imputation.

2.3 Existing Methods Show Only Moderate Success on Benchmarks

Previous investigations sought to characterize the performance of numerical imputation strategies on benchmark datasets to identify the strongest general use approach. Poulos and Valle find that kNN imputation—compared to random forest imputation, random replacement, point-value imputation, and support vector machine-based methods—most strongly reduced the downstream validation error performance of supervised classification models (Poulos and Valle 2018). Jadhav, Pramod, and Ramanathan similarly find that kNN imputation achieves the lowest mean RMSE compared with predictive mean matching, Bayesian Linear Regression, non-Bayesian regression, and random sampling for three experimental datasets perturbed with ten percent to fifty percent data missingness (Jadhav et al. 2019). Jäger, Allhorn, and Bießmann benchmarked mean value, kNN, Random Forest, Variational Autoencoder (VAE), and Generative Adversarial Imputation Networks (GAIN)-based imputation strategies across 69 labelled datasets across the three data missingness mechanisms. They find that Random Forest-based imputation consistently outperforms all others, regardless of whether the imputation methods are provided complete training data for model fitting (Jäger et al. 2021). They also report that VAE and GAIN perform poorly relative to other strategies regardless of missingness type or missingness proportion. However, while they report the average relative rank of each algorithm on the imputation experiments, they do not report the average RMSE of imputations, making it unclear the degree to which one strategy is preferable over another. Furthermore, no experiments benchmarking imputation assess the runtime of their models, a crucial consideration that affects the scalability for big data and deep learning applications.

2.4 Technical Challenges of ML-Based Imputation Strategies

While ML-based imputers repeatedly achieve best-in-class performance, there exist demonstrable limitations and areas for algorithmic improvement. Eager learners, such as linear regression models and random forest regressors that drive MICE and MissForest, hold out one feature as a pseudo-target and leverage remaining feature columns as a pseudo-feature set. These algorithms train a model on rows where there are no missing data in either the pseudo-feature set or the pseudo-target, then predict the pseudo-target for rows where it is absent (and where the pseudo-feature set is complete). This creates two technical challenges.

The first challenge is how these algorithms handle missing values in the pseudo-feature set, since a complete feature set is necessary to train models and predict missing pseudo-targets. This is particularly noteworthy in MAR and MNAR-type missing data. In MAR contexts, data missing in a pseudo-target may be associated with missing values among pseudo-features, meaning these eager learners often fit models on feature subsets with fewer training instances which in turn increases the bias of the imputation estimate. In MNAR contexts, where missing values in a pseudo-target may be associated with values observed in that same pseudo-target, the distribution of pseudo-target values in the training subset may not be representative of the imputation subset, creating poorly generalized estimates.

The second challenge is with how these algorithms select the sequence of features to impute, which is an overlooked but non-trivial design. In active learning and Bayesian optimization, this challenge is resolved with acquisition functions, which model uncertainty associated with instances of unlabeled data to determine the sequence of instances the model should predict to minimize future uncertainty (Aral and Van Alstyne 2010; Larson 2017; Di Fiore et al. 2024). ML-based imputation inherently lends itself to consider an active learning-inspired framework, as imputing values in each pseudo-target will influence associations between that label and the feature set, propagating changes to future features' imputations.

Here, we present SudokuImputer, a graph-based imputation framework that leverages network analysis and ML modeling to reduce the error of estimates by selecting an optimal sequence of features whose imputation would strengthen feature-wise associations for future iterations of imputation. The roadmap for this paper is as follows: first, we discuss the theory and logic underlying the framework, then evaluate its performance across an exhaustive design hyperparameter space, and finally benchmark its performance on three datasets induced with MCAR, MNAR, and MAR missingness at a ten percent to fifty percent proportion of missingness.

3. Methods

3.1 SudokuImputer Algorithm Design

The values missing from a feature set can be analogized to a sudoku puzzle; sudoku requires the values 0 through 9 be used only once in each row, column, or three-by-three cluster, and the player's objective is to fill in empty cells using information provided by the values of surrounding cells. From a game theory perspective, sudoku is a constraint satisfaction problem where the optimal strategy minimizes the risk of a misassignment at each step and maximizes information gain for downstream steps. As the player makes assignments, it influences their future assignments. Therefore, there exists an optimal path to sequentially assign values to cells in order to solve the puzzle with 0 backtracks or corrections.

This was the inspiration behind our novel imputation technique called SudokuImputer. SudokuImputer estimates missing values through an iterative three-stage process of (1) weighted graph construction, (2) network-based prioritization, and (3) regression-based prediction. This three-step iterative algorithm mimics uncertainty-based principles leveraged by graph Bayesian optimization, entropy minimization, and acquisition functions in active learning (Gärtner et al. 2003; Frazier 2018; Garnett 2023; Xie et al. 2025). First, the algorithm represents the data matrix as an undirected weighted graph whose nodes correspond to features and whose edges quantify pairwise associations. Edge weights can be defined (1) by measures of correlation (e.g., Spearman's ρ, Pearson's r, or coefficient of determination R^2), (2) by the proportion of rows jointly observed for each feature pair, or (3) by their product. This weighting scheme ensures that both statistical association and data availability influence the topology of the graph. Users may also impose an edge-weight threshold, thereby restricting the graph to only the strongest or most data-supported feature connections; high thresholds increase graph sparsity, which could enhance detection of strong subgraphs in multicollinear datasets or isolate islands in uncorrelated data.

A distinctive aspect of SudokuImputer is its flexibility in how the order of imputation is determined. In centrality prioritization, the sequence is governed by graph-theoretic centrality measures such as degree, betweenness, or eigenvector centrality. This strategy allows the algorithm to begin with either highly connected "hub" features (descending centrality) or with more peripheral features (ascending centrality). In contrast, utility-based prioritization directly ranks features by the proportion of observed data available for them (they possess more completeness, and thus "utility," to fit imputation models for partner features), in either ascending or descending rank order. Based on uncertainty sampling in active learning, the ascending direction is recommended for centrality-prioritized imputation. While centrality prioritization emphasizes structural importance in the correlation network, utility prioritization emphasizes practical completeness, offering an alternative that may be advantageous in settings with uneven missingness patterns.

Once centrality-based or utility-based prioritization selects the top-ranking target feature, it is imputed using its neighbors in the graph that exceed the chosen edge threshold. These neighbors serve as predictors in a multivariate regression model, which may be instantiated by linear regression, tree-based methods, or other user-specified learners. The model is trained on the subset of rows where both the target and its partner features are observed and then used to predict missing entries in the target. If some values remain missing after this process, which may be the case if selected partner nodes are also null at a given row, then they are filled using a simple median fallback, ensuring that no feature column remains incomplete.

After each column is imputed, the data matrix is updated, and the weighted graph is reconstructed to reflect the new joint availability and correlations. This cycle of graph construction, prioritization, and imputation repeats until all features have been filled. By continuously adapting the graph to incorporate newly imputed data, SudokuImputer leverages both global feature relationships and evolving data availability, yielding a dynamic and context-aware imputation process (Figure 1).

We performed two phases of experiments: (1) SudokuImputer framework optimization and hyperparameter exploration and (2) benchmarking evaluation of SudokuImputer against competing numerical imputation strategies.

3.2 Evaluating SudokuImputater Design Construction Parameters

We evaluate six design hyperparameters on SudokuImputer: centrality method (degree, betweenness, or eigenvector centrality), feature prioritization mode (centrality-based, utility-based), partner node proportion (eight percent, fifteen percent, or twenty-five percent), an eight-five percent edge threshold, and various combinations of alpha and

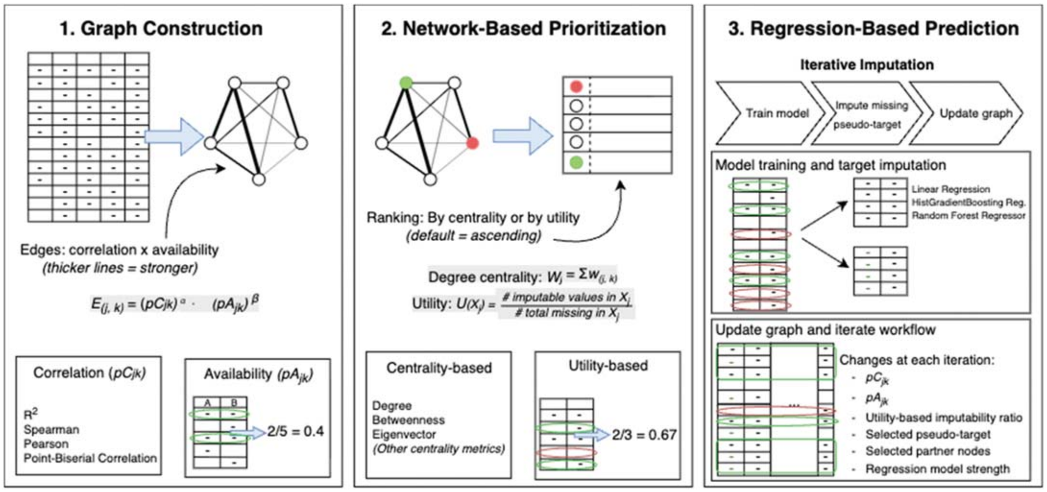


Figure 1: Flowchart depicting the SudokuImputer framework. The input dataset is iteratively processed through three steps: (1) graph construction, (2) network-based node prioritization, and (3) regression-based prediction. The framework provides a non-arbitrary method for identifying which values in each iteration will propagate better estimates for future imputations.

beta values ([1.5, 0.5], [0.5, 1.5], and [1.0, 1.0]). Alpha and beta coefficients tune the attribution of statistical association (alpha) and pairwise availability (beta) to edge weight construction; larger alpha increases the effect of feature correlation strength on edge weighting, while larger beta increases the effect of joint data availability. The electroencephalogram (EEG) eye state dataset induced with MCAR, MAR, and MNAR missingness at a ten percent proportion is used to perform a grid search across this hyperparameter space.

3.3 Benchmarking Sudoku Imputation Against SOTA Algorithms

Three classification benchmarking datasets were sourced and downloaded from the University of California Irvine Machine Learning Repository. The QSAR biodegradation dataset contains 1055 chemical compounds (rows) characterized by 41 molecular descriptors (columns) (Mansouri et al. 2013). The EEG eye state dataset contains 14 EEG measurements for 14,980 instances of eye state measurements (Roesler 2013). Finally, the Arcene mass spectrometry dataset characterizes the protein expression for 10,000 species in blood serum across 900 patients (Guyon et al. 2004). These three benchmarks were selected to evaluate datasets with varying dimensionality, size, and inductive biases. For instance, the Arcene dataset has a high feature space with few instances and given the nature of tumor biology there may be coregulation of plasma proteomic features, while the EEG dataset has many instances with few features that capture rather distinct EEG traits. We induce data missingness at varying proportions from 0.1 to 0.5 (by increments of ten percent) for each data missingness class (MCAR, MAR, and MNAR) across each dataset. A complete characterization of experimental benchmark datasets is available in Supplemental Table 1.

SudokuImputer performance was benchmarked against point-value imputation (with mean and median), MICE, kNN, matrix factorization, and SoftImpute (SVD). We evaluated model performance primarily using algorithm runtime and the RMSE between imputed and ground truth values for each column; we normalize column metrics and compute the mean across columns to aggregate a model-by-dataset performance value. R^2, Pearson correlation, and mean absolute error are also computed for each experimental dataset and available in the supplemental data tables.

4. Results

4.1 Behavior of SudokuImputer across Broad Design Parameter Space

Exploration of the hyperparameter space using Gini entropy-based feature importance demonstrated that runtime was almost entirely determined by prioritization mode (Figure 2(A)). Across MCAR, MNAR, and MAR missingness, prioritization mode accounted for 99.38 percent, 97.68 percent, and 98.87 percent of runtime variance, respectively. Centrality-based prioritization produced average runtimes of 3.48, 0.84, and 4.30 seconds across MCAR, MNAR, and MAR, while utility-based prioritization produced longer runtimes of 46.57, 11.42, and 54.32 seconds

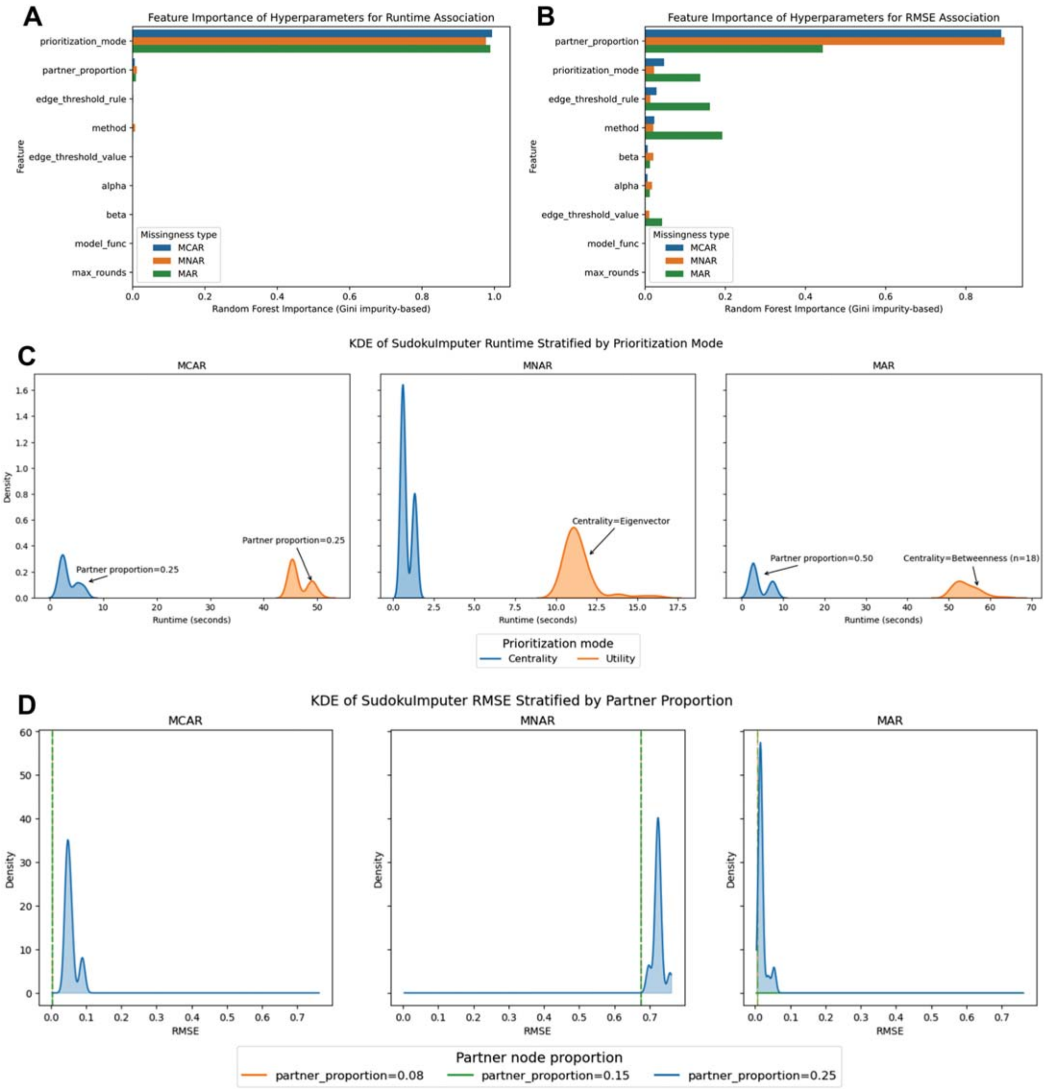


Figure 2: Gini entropy-based associations between SudokuImputer (A) runtime and (B) RMSE against algorithm hyperparameters. (C) Runtime distributions for 108 hyperparameter combinations in MCAR (left), MNAR (middle), and MAR (right) data missingness grouped by prioritization mode, the hyperparameter most strongly associated with runtime by Gini impurity. (D) RMSE distributions for 108 hyperparameter combinations grouped by imputation partner node proportion.

(Figure 2(C)). RMSE was strongly affected by numerous hyperparameters. Partner proportion showed the strongest Gini entropy association with RMSE (88.84 percent, 89.65 percent, and 44.30 percent across MCAR, MNAR, and MAR), while prioritization mode, edge threshold rule, edge threshold value, and centrality method each contributed more than two percent feature importance to RMSE in at least one missingness class (Figure 2(B)). A partner proportion of 0.25 consistently resulted in higher average RMSE (Figure 2(D)).

UMAP projections of hyperparameter combinations reveals that distinct clusters of parameters are associated with lower runtimes, while disparate coordinates in the hyperparameter space are associated with lower RMSE (Supplemental Figure 1). Furthermore, association analyses between UMAP embeddings and corresponding hyperparameter values demonstrates that the prioritization mode was a dominant contributor to model grouping (Supplemental Figure 1(C)) across MCAR and MAR missingness classes. Interestingly, UMAP embeddings for MNAR missingness are most strongly associated with the edge threshold rule (a design criteria for graph sparsity), which is either an 85th percentile cutoff or no rule (full graph). The raw RMSE and runtime of hyperparameter grid search experiments are available in Supplemental Tables 3, 4, and 5.

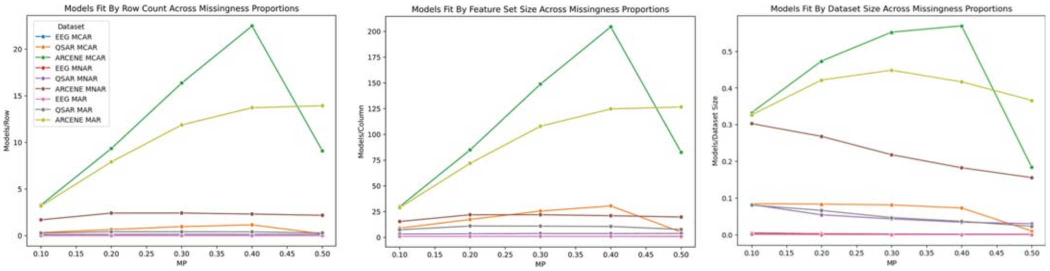


Figure 3: Model fitting behavior of SudokuImputer relative to total dataset volume, column count, and row count across missingness proportions. All other model properties (imputability ratio, fallback imputations per dataset, etc.) were trivial or insignificant; a full table is defined and available in the log files published to GitHub (see Supplemental Materials).

The number of models trained varied substantially across experimental datasets. No clear relationship was observed between the number of models trained and the missingness proportion (Supplemental Figure 2). Instead, dataset dimensions (row count, column count, and their product) were correlated with the total number of models trained (Figure 3). Dataset volume was therefore a more consistent driver of model training burden than missingness level. Other logged properties, including imputability ratio and fallback imputations, contributed minimally and were not retained for further analysis. Overall, these findings indicate that SudokuImputer's scalability is primarily influenced by dataset size and structure rather than the proportion of missing values.

4.2 SudokuImputer Is Best-in-Class against Six Numerical Imputation Techniques

Across 45 experimental datasets (three datasets induced with three missingness types over five missingness proportions), SudokuImputer records best-in-class performance by RMSE in 22 datasets. This effect is most noteworthy at higher proportions of data missingness, where relative to other strategies SudokuImputer often records a substantially lower average RMSE (Figure 4; Supplemental Figure 3). The performance of SudokuImputer is more robust on MNAR and MAR data missingness than comparable ML-based techniques like MICE and SoftImpute. Statistical ML-based imputation strategies outperform point-value imputation across all benchmarks. When each method was ranked in each experimental dataset and averaged together by data missingness class, SudokuImputer was consistently the top-ranking algorithm by missingness type (Figure 5). Other algorithms such as MICE and kNN inconsistently beat SudokuImputer and demonstrate a worse performance distribution than SudokuImputer. In a comparison of the mean and median rank of MICE, kNN, and SudokuImputer grouped by experimental dataset and missingness type, SudokuImputer records the highest or second highest rank in seven out of nine groups, while kNN and MICE combined record the highest or second highest mean/median rank in eight out of nine groups (Figure 6). Mean RMSE scores of all models in each dataset are available in Supplemental Table 1.

SudokuImputer outperforms all imputation algorithms (records the lowest RMSE) in four out of 15 MCAR experiments, nine out of 15 MNAR experiments, and nine out of 15 MAR experiments (Figure 7). When SudokuImputer achieves the lowest RMSE, it is often by a stable margin; average normalized RMSE margins between SudokuImputer and the second-best method are 0.0057, 0.0067, and 0.012 for MCAR, MNAR, and MAR-type missingness. When SudokuImputer does not achieve best-in-class performance in MCAR and MNAR experiments, it is by an equivalently stable margin (average normalized RMSE margins: 0.0053, 0.0038). However, in MNAR SudokuImputer underachieves by considerably larger margins (Figure 8). SudokuImputer demonstrated stable performance relative to other imputation techniques; it had fewer monotonicity violations than kNN and MICE on EEG and QSAR datasets (Supplemental Figure 4) and noteworthy performance jumps (greater than a mean RMSE of 0.05) between subsequent missingness proportions were only observed between forty percent and fifty percent missingness (Supplemental Figure 5).

While SudokuImputer records the lowest average RMSE across most experimental datasets, it also records the highest runtime across datasets compared with six numerical imputation strategies (337.96, 138.60, and 5212.77 seconds for the QSAR, EEG, and Arcene datasets) (Figure 8). Given that the QSAR, EEG, and Arcene datasets each contain 40, 12, and 100 columns of data, it appears that SudokuImputer and MICE runtimes scale proportionally to the feature set size of the dataset while kNN, matrix factorization, and SoftImpute are more robust against dataset size effects. Across all non-point value imputation algorithms, runtime is slightly proportional to data missingness proportion, with longer runtimes documented for larger missing data volumes. The raw runtime of each model in each experimental dataset is available in Supplemental Table 2.

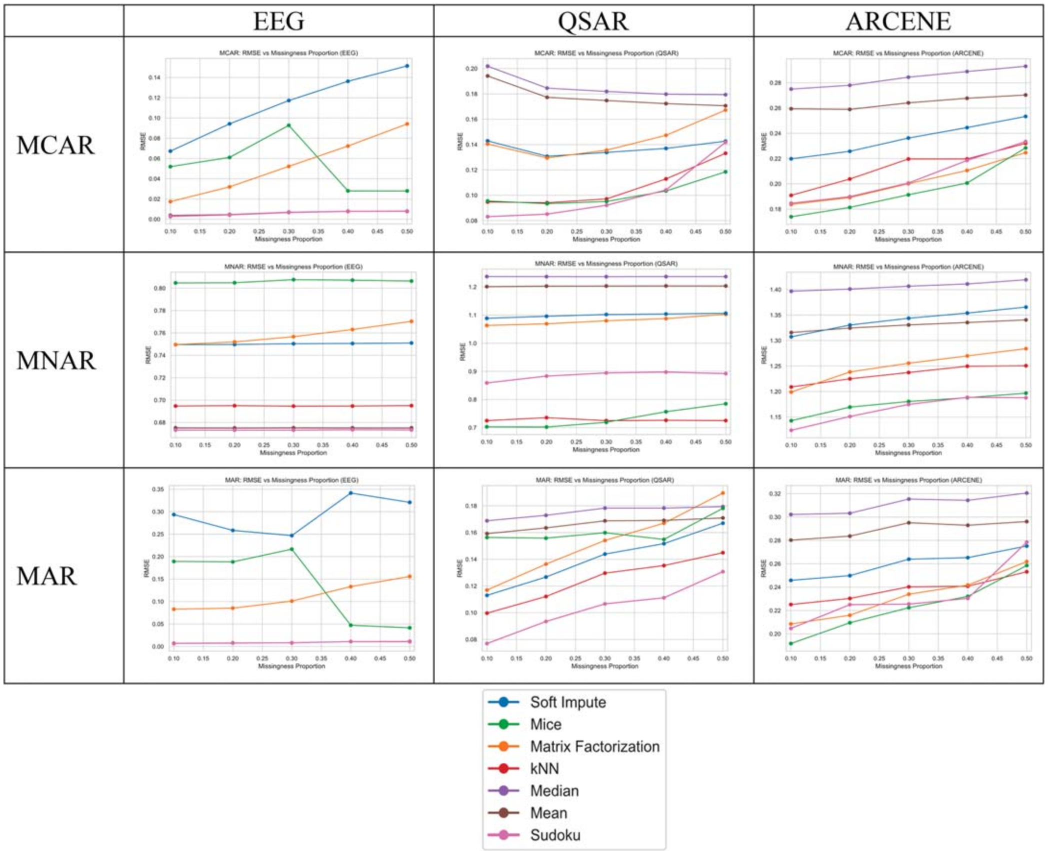


Figure 4: Comparative line plots of mean RMSE over data missingness proportions for SudokuImputer against median, mean, kNN, MICE, SoftImpute, and Matrix Factorization imputation. Results are grouped by data missingness type and experimental dataset.

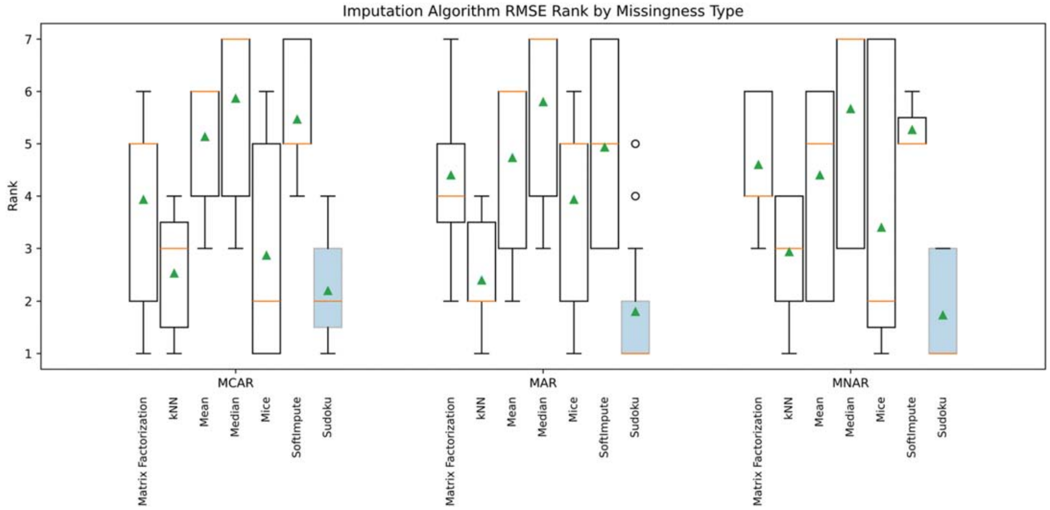
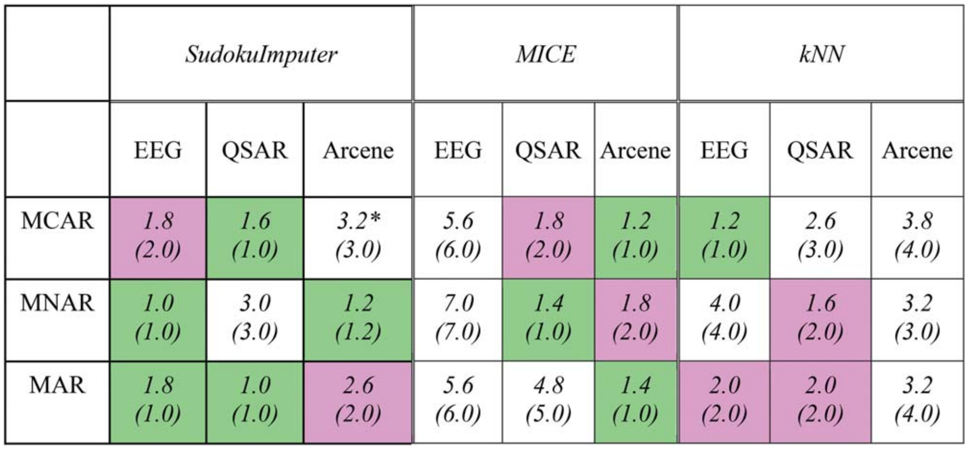


Figure 5: Median and mean ranking of SudokuImputer against six benchmarking imputation methods by average imputation RMSE. Ranking is scored among the seven algorithms within each missingness proportion for a given experimental dataset (i.e. within "EEG 10% missingness"). Each box-and-whisker plot represents the distribution of 15 rankings for that algorithm on the grouped missingness type. The median rank is denoted with the orange line, and the average rank with the green triangle. The highest overall ranking method for each missingness type is highlighted in blue.



	SudokuImputer			*MICE*			*kNN*		
	EEG	QSAR	Arcene	EEG	QSAR	Arcene	EEG	QSAR	Arcene
MCAR	*1.8 (2.0)*	*1.6 (1.0)*	*3.2* (3.0)*	*5.6 (6.0)*	*1.8 (2.0)*	*1.2 (1.0)*	*1.2 (1.0)*	*2.6 (3.0)*	*3.8 (4.0)*
MNAR	*1.0 (1.0)*	*3.0 (3.0)*	*1.2 (1.2)*	*7.0 (7.0)*	*1.4 (1.0)*	*1.8 (2.0)*	*4.0 (4.0)*	*1.6 (2.0)*	*3.2 (3.0)*
MAR	*1.8 (1.0)*	*1.0 (1.0)*	*2.6 (2.0)*	*5.6 (6.0)*	*4.8 (5.0)*	*1.4 (1.0)*	*2.0 (2.0)*	*2.0 (2.0)*	*3.2 (4.0)*

Figure 6: Mean and median (in parenthesis) RMSE rankings for three ML-based imputation algorithms: SudokuImputer (left), MICE (middle), and kNN (right). Green and purple values indicate the highest and second highest rank among all seven evaluated imputation strategies. * Matrix factorization scored the second highest rank in the MCAR Arcene dataset (mean rank = 1.8, median rank = 2.0).

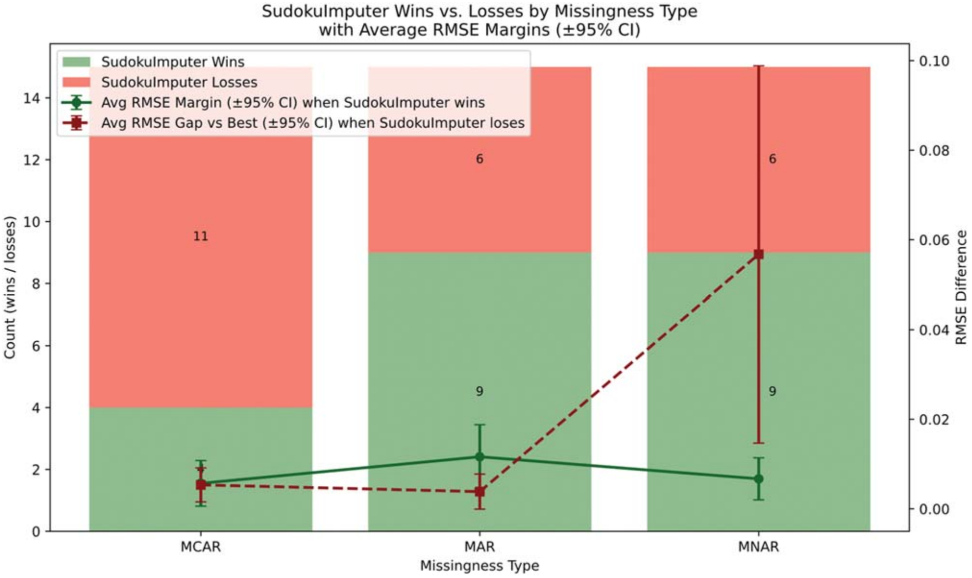


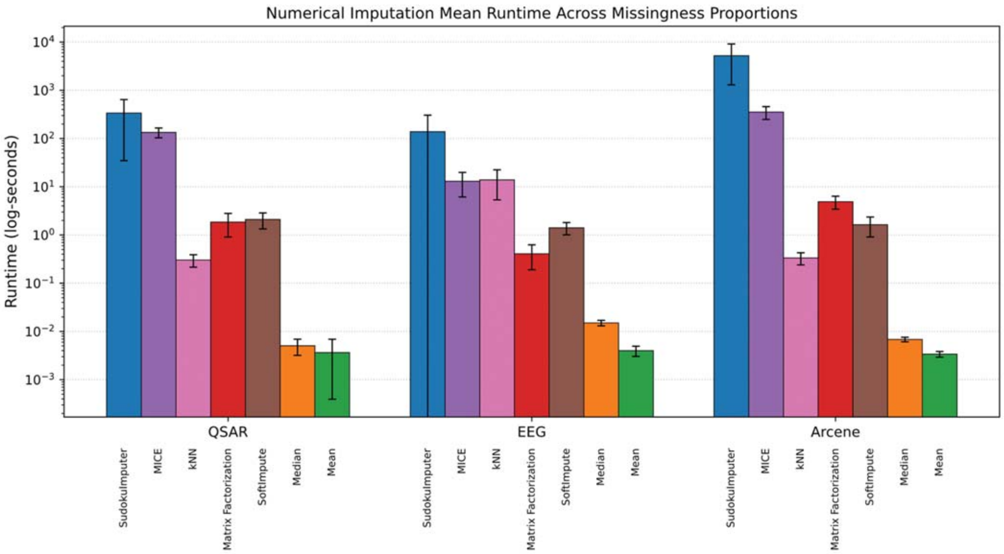
Figure 7: Number of SudokuImputer best-in-class performances relative to six imputation methods across 15 experimental datasets, as measured by average RMSE. Among SudokuImputer wins (lowest RMSE), the average RMSE margin against the runner-up method is overlayed in green. Among SudokuImputer losses, the average RMSE margin against the best performing imputation method is overlayed in red. Error bars indicate a ninety-five percent confidence interval.

5. Discussion

In this study, we introduce SudokuImputer, a graph-based imputation framework inspired by constraint satisfaction in Sudoku puzzles, and we evaluate its performance on three benchmark datasets with MCAR, MAR, and MNAR missingness across a range of missingness proportions. Our results demonstrate that SudokuImputer frequently achieves SOTA RMSE performance, particularly under MAR and MNAR conditions, but at the cost of higher computational runtime.

5.1 Hyperparameter Behavior and Design Implications

Analysis of the hyperparameter space revealed that runtime is overwhelmingly determined by prioritization mode, with centrality-based feature imputation sequencing resulting in order-of-magnitude reductions in execution time



	Sudoku Imputer	MICE	kNN	Matr. Fact.	Soft Impute	Median	Mean
QSAR	337.956	134.293	0.302	1.852	2.098	0.005	0.004
EEG	138.596	12.983	13.889	0.407	1.410	0.015	0.004
Arcene	5212.768	353.707	0.334	4.897	1.635	0.007	0.003

Figure 8: Runtime comparison between seven imputation strategies grouped by experimental dataset in log-seconds. Error bars reflect the standard deviation of runtime across the five missingness proportions for a given dataset. Values in the data table reflect runtime in non-logarithmic seconds.

compared with utility-based feature imputation sequencing. This finding underscores the importance of feature ordering in iterative imputation, a design choice that has received little attention in prior work. While kNN, MICE, and MissForest are known to achieve robust imputations through neighborhood or ensemble learning, they do not adaptively prioritize features during the imputation process. SudokuImputer's reliance on network centrality thus provides a novel mechanism to improve scalability.

In contrast, RMSE was shaped by multiple interacting hyperparameters, most notably partner proportions. Strikingly, excessively lenient partner inclusion (e.g., twenty-five percent of features) degraded accuracy, suggesting that excess connectivity introduces redundancy that weakens predictive signal. Accuracy was primarily influenced by hyperparameters such as partner proportion, which can introduce excess noise when a large amount of weakly informative features are included. This aligns with theories from network science and recommendation systems that excess connections dilute the relevance of strong associations (Larson 2017; Kirsch et al. 2019). For SudokuImputer, tuning the partner proportion hyperparameter represents a critical balance between information sufficiency and over-saturation.

5.2 Scalability and Model Training Burden

The number of models trained was associated primarily with dataset size rather than the proportion of missing values. This indicates that SudokuImputer's computational burden is more sensitive to dataset dimensionality (rows × columns) than to the extent of missingness. Practically, this means SudokuImputer is more efficient for wide but shallow datasets than for large-scale, high-volume cohorts. All benchmarked imputation algorithms demonstrated marginally weaker performance on the Arcene benchmark than the EEG or QSAR datasets, indicating a slight association between larger feature set size and larger estimation errors. By contrast, simpler strategies such as point-value imputation or kNN scale more gracefully with dataset size, though at the expense of accuracy.

5.3 Benchmarking against Existing Methods

SudokuImputer achieved best-in-class RMSE in nearly half of the 45 experimental datasets, with especially strong performance under MAR and MNAR conditions (best-in-class in nine out of 15 datasets from each condition). This is notable because MAR and MNAR mechanisms pose greater challenges to traditional imputation methods, which often assume randomness in missingness (MCAR). The robustness of SudokuImputer in these settings suggests that its graph-based prioritization and iterative refitting capture structural dependencies that are overlooked by eager learners like MICE or matrix factorization methods.

Nevertheless, SudokuImputer's advantage was not universal. In MCAR experiments, methods such as kNN and MICE occasionally performed best among the seven evaluated methods, highlighting that simple similarity-based strategies remain competitive when missingness is random. This pattern reflects broader benchmarking studies reporting kNN and Random Forest imputers as consistently strong general-purpose methods. SudokuImputer therefore complements rather than supplants existing approaches, excelling in structured missingness regimes where others falter.

5.4 Runtime Tradeoffs and Practical Feasibility

Despite its accuracy, SudokuImputer recorded the longest runtimes of all evaluated methods. This tradeoff mirrors prior observations that ensemble- and regression-based imputers (e.g., MissForest, MICE) sacrifice speed for accuracy, while matrix factorization and deep generative models achieve faster runtimes but poorer imputations. This is due to the sheer volume of models fit by SudokuImputer, which is proportional to the size of the dataset (namely, the feature count). SudokuImputer's scaling behavior—runtime increasing proportionally with feature set size—makes it less suitable for extremely high-dimensional datasets unless paired with optimization strategies such as parallelization, dimensionality reduction, or sparsity constraints.

The critical question, therefore, is whether the observed RMSE gains justify the computational cost. For small to medium-sized datasets, particularly in biomedical domains where accuracy outweighs runtime, SudokuImputer may provide a favorable tradeoff. However, for very large-scale datasets such as electronic health records or biobank-scale omics data, runtime optimization will be essential before SudokuImputer can be deployed in practice.

5.5 Limitations and Future Directions

We identified numerous opportunities to evaluate and optimize the SudokuImputer v1.0.0 framework. First, evaluations were restricted to three benchmark datasets, which, although diverse, may not capture the heterogeneity of real-world data. Future investigations of benchmarking imputation strategies should evaluate a wide missingness proportion over all three classes of data missingness, across real-world datasets of varying dimensions. Second, we focused exclusively on numerical imputation, excluding mixed-type datasets where categorical features are prevalent. While the beta release of SudokuImputer v1.0.0 supports mixed-type and categorical imputation, these functions were not the subject of our benchmarking experiments and require further rigorous validation and optimization. Third, although SudokuImputer was benchmarked against widely used methods, emerging deep learning–based imputers such as GAIN and diffusion models were not fully explored.

Next releases of SudokuImputer will focus on three directions. Methodologically, optimizing SudokuImputer's runtime via parallel graph construction, adaptive stopping criteria, or hybrid partner selection could substantially reduce computational burden. Additionally, iterative model fitting was not thoroughly explored; a MICE-like or MissForest-like process, where temporary placeholder values are added to increase the imputability ratio may result in stronger estimates on high density missing data. Finally, integrating SudokuImputer into downstream predictive pipelines will determine whether its gains in RMSE translate into improved classification or regression performance, which is the end goal of imputation.

6. Conclusion

SudokuImputer offers a novel graph-based approach to numerical imputation that consistently achieves low RMSE across diverse missingness settings, especially MAR and MNAR, but incurs substantially higher runtime compared with SOTA methods. These findings highlight both the promise and current limitations of graph-based imputation, motivating future efforts to balance accuracy with scalability. Imputation strategies optimized for both estimation accuracy and dataset scalability allow for efficient generation of reliable training datasets for ML models and AI systems. Many diverse training datasets are critical for these deep learning networks to achieve SOTA performance.

7. Data Availability

Experimental datasets, benchmark imputations, summary tables, figures, and all experimental scripts, analytical workflows, and visualization notebooks will be made available on the SudokuImputer GitHub repository: https://github.com/vkanpa01/SudokuImputer

8. Supplemental Material

Supplemental material can be accessed using the following link: https://github.com/vkanpa01/SudokuImputer.

Acknowledgements: *We would like to thank William Patterson University and the organizers of the New Jersey Big Data Association Annual Symposium, who provided us the opportunity to receive feedback and recommendations on SudokuImputer version 1.0.0 from experienced academics and data practitioners. Furthermore, we would like to thank Dr. Vadim Zhernovkov, Assistant Professor at University College Dublin and Director of the AI in Medicine & Medical Research program, for his feedback and insights into benchmarking methods and analytic techniques to assess the performance of SudokuImputer version 1.0.0.*

References

Alam, S., M. S. Ayub, S. Arora, and M. A. Khan. 2023. "An Investigation of the Imputation Techniques for Missing Values in Ordinal Data Enhancing Clustering and Classification Analysis Validity." *Decision Analytics Journal* **9**: 100341. https://doi.org/10.1016/j.dajour.2023.100341

Aral, S., and M. W. Van Alstyne. 2010. "The Diversity-Bandwidth Tradeoff." *American Journal of Sociology* (forthcoming). https://doi.org/10.2139/ssrn.958158

Brown, S., O. Kudia, K. Kleine, B. Kidd, R. Wines, and N. Meckes. 2025. "Comparing Multiple Imputation Methods to Address Missing Patient Demographics in Immunization Information Systems: Retrospective Cohort Study." *Archives of Computational Methods in Engineering* **31**: 2985–3013. https://doi.org/10.2196/preprints.73916

Di Fiore F., M. Nardelli, and L. Mainini. 2024. "Active Learning and Bayesian Optimization: A Unified Perspective to Learn with a Goal." *Archives of Computational Methods in Engineering* (forthcoming). https://arxiv.org/abs/2303.01560

Frazier, P. I. 2018. "A Tutorial on Bayesian Optimization." Preprint, submitted July 8.

Garnett, R. 2023. *Bayesian Optimization Book*. Available at https://bayesoptbook.com/

Gärtner, T., P. Flach, S. Wrobel. 2003. "On Graph Kernels: Hardness Results and Efficient Alternatives." In *Learning Theory and Kernel Machines. Lecture Notes in Computer Science*, B. Schölkopf, and M. K. Warmuth, editors, volume 2777. Berlin and Heidelberg, Germany: Springer. https://doi.org/10.1007/978-3-540-45167-9_11

Guyon, I., S. Gunn, A. Ben-Hur, and G. Dror. 2004. "Arcene." [Dataset]. *UCI Machine Learning Repository*. https://doi.org/10.24432/C58P55

Halder, R. K., M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat. 2024. "Enhancing K-Nearest Neighbor Algorithm: A Comprehensive Review and Performance Analysis of Modifications." *Journal of Big Data* **11**, no. 1. https://doi.org/10.1186/s40537-024-00973-y

Hestness, J., S. Narang, N. Ardalani, et al. 2017. "Deep Learning Scaling is Predictable, Empirically." Preprint, submitted December 1.

Heymans, M. W., and J. W. R. Twisk. 2022. "Handling Missing Data in Clinical Research." *Journal of Clinical Epidemiology* **151**: 185–188. https://doi.org/10.1016/j.jclinepi.2022.08.016

Hoffmann, J., S. Borgeaud, A. Mensch, et al. 2022. "Training Compute-Optimal Large Language Models." Preprint, submitted March 29.

Jadhav, A., D. Pramod, and K. Ramanathan. 2019. "Comparison of Performance of Data Imputation Methods for Numeric Dataset." *Applied Artificial Intelligence* **33**, no. 10: 913–933. https://doi.org/10.1080/08839514.2019.1637138

Jäger, S., A. Allhorn, and F. Bießmann. 2021. "A Benchmark for Data Imputation Methods." *Frontiers in Big Data* **4**: 693674. https://doi.org/10.3389/fdata.2021.693674

Kang, H. 2013. "The Prevention and Handling of the Missing Data." *Korean Journal of Anesthesiology* **64**, no. 5: 402–406. https://doi.org/10.4097/kjae.2013.64.5.402

Kaplan, J., S. McCandlish, T. J. Henighan, et al. 2020. “Scaling Laws for Neural Language Models.” Preprint, submitted January 23.

Kirsch, A., J. van Amersfoort, and Y. Gal. 2019. “BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning [Review of BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning].” *NeurIPS* **33**: 1–12. https://doi.org/10.48550/arXiv.1906.08158

Larson, J. M. 2017. “The Weakness of Weak Ties for Novel Information Diffusion.” *Applied Network Science* **2**, no. 1. https://doi.org/10.1007/s41109-017-0034-3

Mansouri, K., T. Ringsted, D. Ballabio, R. Todeschini, and V. Consonni. 2013. “*QSAR Biodegradation.”* [Dataset]. *UCI Machine Learning Repository*. https://doi.org/10.24432/C5H60M

Mera-Gaona, M., U. Neumann, R. Vargas-Canas, and D. M. López. 2021. “Evaluating the Impact of Multivariate Imputation by MICE in Feature Selection.” *PLoS One* **16**, no. 7: e0254720. https://doi.org/10.1371/journal.pone.0254720

Misztal, M. A. 2019. “Comparison of Selected Multiple Imputation Methods for Continuous Variables – Preliminary Simulation Study Results.” *Acta Universitatis Lodziensis. Folia Oeconomica* **6**, no. 339: 73–98. https://doi.org/10.18778/0208-6018.339.05

Monard, M.-C. 2002. “A Study of K-Nearest Neighbour As an Imputation Method [Review of A Study of K-Nearest Neighbour As an Imputation Method].” *Soft Computing Systems: Design, Management and Applications* 1–10. https://www.researchgate.net/publication/220981745_A_Study_of_K-Nearest_Neighbour_as_an_Imputation_Method

Poulos, J., and R. Valle. 2018. “Missing Data Imputation for Supervised Learning.” *Applied Artificial Intelligence* **32**, no. 2: 186–196. https://doi.org/10.1080/08839514.2018.1448143

Prasad, A. 2024. “Impact of Poor Data Quality on Business Performance: Challenges, Costs, and Solutions.” *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4843991

Roesler, O. 2013. “*EEG Eye State.”* [Dataset]. *UCI Machine Learning Repository*. https://doi.org/10.24432/C57G7J

Rubin, D. B. 1976. “Inference and Missing Data.” *Biometrika* **63**, no. 3: 581–592. https://doi.org/10.2307/2335739

Schonlau, M., and R. Y. Zou. 2020. “The Random Forest Algorithm for Statistical Learning.” *The Stata Journal* **20**, no. 1: 3–29. https://doi.org/10.1177/1536867X20909688

Schouten, R. M., and G. Vink. 2018. “The Dance of the Mechanisms: How Observed Information Influences the Validity of Missingness Assumptions.” *Sociological Methods & Research* **50**, no. 3: 1243–1258. https://doi.org/10.1177/0049124118799376

Sun, T., S. Zhu, R. Hao, B. Sun, and J. Xie. 2022. “Traffic Missing Data Imputation: A Selective Overview of Temporal Theories and Algorithms.” *Mathematics* **10**, no. 14: 2544. https://doi.org/10.3390/math10142544

Xie, Y., S. Zhang, J. Qing, R. Misener, and C. Tsay. 2025. “Global Optimization of Graph Acquisition Functions for Neural Architecture Search.” Preprint, submitted May 29.

Zhang, Z.. 2016. “Missing Data Imputation: Focusing on Single Imputation.” *Annals of Translational Medicine* **4**, no. 1: 9. https://doi.org/10.3978/j.issn.2305-5839.2015.12.38

WWW.JBDAI.ORG

ISSN: 2692-7977

JBDAI Vol. 4, No. 1, 2026 DOI: 10.54116/jbdai.v4i1.75

COMPARATIVE ANALYSIS OF WIND TURBINE BLADE DESIGN FOR URBAN INDIA: OPTIMIZING POWER GENERATION AND STRUCTURAL EFFICIENCY

Sanskruti Thakur
Mumbai University
1032220869@tcetmumbai.in,
Sanskruti7thakur@gmail.com

Atharva Tanpure
Mumbai University
atharvatanpure18@gmail.com

Megha Gupta
Mumbai University
megha.gupta@thakureducation.org

Anvita Birje
Mumbai University
anvita.birje@thakureducation.org

ABSTRACT

India's metropolitan areas are growing rapidly, creating an increasing demand for clean, local, and reliable energy sources. However, urban environments present unique challenges for wind energy due to the presence of tall structures, narrow streets, and variable air currents that create turbulent wind patterns. This makes it difficult to select the most suitable wind turbine design because traditional turbines often struggle under such conditions. Vertical-axis wind turbines (VAWT), including Darrieus, Helical, H-Rotor, and Savonius types, are emerging as viable alternatives because they can operate effectively in turbulent and space-constrained urban environments. In this study, we propose a decision-making framework that combines real-world geographic data with modern data-driven modeling approaches to identify the most appropriate turbine design for each major city in India. We used the ERA5 reanalysis dataset, which spans 44 years, from 1980 to 2024, to capture realistic wind patterns across the country. To adapt these data for rooftop-level conditions (30 m), we adjusted the 100-m wind data by using a logarithmic wind profile. We trained both Random Forest and LSTM-CNN hybrid machine learning models to predict wind speed, power output, and the most suitable turbine type for any given location. Our findings indicate that Darrieus turbines are most effective in cities with stronger and more consistent winds, whereas Savonius and H-Rotor designs are better suited for densely built areas with lower wind speeds. The LSTM-CNN hybrid model achieved a classification accuracy of 94.8%, which is significantly higher than the 92.4%

accuracy of the Random Forest baseline. This study offers a validated and adaptable framework that can assist in rooftop wind installations, urban planning, and small-scale renewable energy systems. By integrating engineering expertise with intelligent modeling, this research helps Indian cities optimize their utilization of local wind patterns.

Keywords *urban wind energy*, *vertical-axis wind turbines (VAWT)*, *Darrieus rotor*, *Savonius rotor*, *helical rotor*, *H-Rotor*, *machine learning for renewable energy*, *wind prediction*, *urban turbulence*, *Indian metropolitan wind analysis*, *turbine design optimization*, *wind-energy recommendation system*, *rooftop wind systems*, *sustainable urban infrastructure.*

1. Introduction

India's cities are changing rapidly. Skyscrapers are going up more often, traffic is getting worse, and millions of people are moving to cities for better jobs and living standards. As this growth continues, an important question arises: How can we power these expanding cities in a way that is sustainable, efficient, and responsible? The future of India's energy systems cannot rely on distant power plants or non-renewable resources. Instead, we need to source energy from within the cities themselves by using rooftops, building exteriors, public spaces, and small installations that fit seamlessly into urban life.

Among the various renewable energy sources, wind energy stands out. It is versatile, clean, and closely connected to nature. However, harnessing wind energy in a city is more complicated than it seems. The gentle breeze felt on a Mumbai balcony or the sudden gusts in a Bengaluru tech park are all affected by buildings, roads, landscapes, and the layout of neighborhoods. Unlike open-field winds, urban winds behave differently, swirling, breaking, accelerating, and slowing in unexpected ways.

This poses a significant challenge. Traditional large wind turbines, such as those seen on highways or in coastal wind farms, need consistent, directional winds to work properly. Cities rarely provide such conditions. Instead, they present a mix of turbulence, lower wind speeds, and limited space. This mismatch creates a demand for different kinds of wind technology suited to urban settings.

This is where vertical-axis wind turbines (VAWT) come in, including types such as the Darrieus, Savonius, Helical, and H-Rotors. Unlike traditional horizontal-axis turbines, VAWTs do not need to face the wind directly. They can operate from any direction, tolerate turbulence, and fit well on rooftops and in narrow spaces. These turbines are designed to be compact, quiet, and visually appealing, which makes them more suitable for city life. However, each type of VAWT has its strengths and weaknesses, and no single design is ideal for every situation. Different cities, or even different areas within the same city, may need different turbine types.

For instance: A coastal city such as Chennai may benefit from high-efficiency turbines during strong seasonal winds. In a densely populated urban center such as Bengaluru or Hyderabad, where wind patterns are more chaotic, a simple and durable design may perform better. In cities with many tall buildings, for example, Mumbai, wind often moves through narrow streets and between structures, creating complex micro-wind zones. India's size and varied climates mean that urban wind technologies must be diverse and specific to each location. Despite this, a major challenge is the lack of detailed wind data at the city level in India. Although national wind maps are available, they typically represent open areas at a height of 100 m, far from the actual conditions on rooftops or mid-rise buildings. Without these crucial data, choosing the right turbine becomes a guessing game, often resulting in underperforming or failing installations.

This study makes three significant contributions to urban wind energy research:

1. It uses the ERA5 reanalysis dataset to evaluate urban wind conditions in India. This dataset provides a validated 44-year historical record of wind patterns at 100 m, downscaled to rooftop levels (30 m).
2. It presents a hybrid machine learning and deep learning framework that combines the interpretability of Random Forest with the predictive capability of LSTM-CNN models, which achieves a classification accuracy of 94.8%.
3. It creates an open-source urban wind intelligence system that connects turbine performance with environmental and local conditions, validated across 15 Indian cities with real-world deployment cases.

This study aims to fill the gap by introducing a customized urban wind intelligence system designed especially for Indian cities. Our approach merges environmental insights with data-driven analysis to answer a crucial question: "Which turbine design is best for a specific city, based on its wind patterns, environment, and physical constraints?"

To achieve this, we use the extensive ERA5 dataset, which covers 44 years (1980–2024) to capture realistic wind patterns across the country. We then downscale 100-m wind data to rooftop levels (30 m) by using logarithmic wind profile adjustments that account for the roughness of urban surfaces. This dataset gives a detailed view of the diverse wind conditions found throughout India.

Next, we develop Random Forest and LSTM-CNN hybrid machine learning models that can predict wind speed, power output, and the best turbine type for any location. These models learn from patterns in the ERA5 data by using geo-environmental variables to understand how wind behaves in India's different landscapes.

In addition to the data-driven aspects, the study also considers the engineering features of each turbine type. Darrieus turbines work well in strong winds but require careful structural support. Savonius turbines are reliable in low-speed conditions but generate less power. Helical turbines provide smooth rotation, which makes them suitable for turbulent winds, whereas H-Rotors strike a balance between simplicity and efficiency. Recognizing these strengths is vital for making recommendations that are accurate and practical in real-world urban settings.

Beyond the technical details, this work seeks to show that renewable energy, especially wind power, can play a key role in urban life in India. Just as solar panels have become common on rooftops recently, well-designed micro wind turbines could soon be a familiar sight. Small-scale, rooftop-friendly wind systems can help supplement power needs, ease the burden on the grid, and allow communities to participate actively in clean energy generation.

This introduction outlines our study's core purpose: creating a thoughtful, data-driven system that respects India's geographic and urban diversity. By combining intelligent modeling with practical engineering insights, we hope to guide policymakers, urban planners, and innovators toward better decisions in implementing urban wind energy solutions. The winds flowing through our cities, even the unpredictable ones, hold significant potential. The challenge is how we choose to harness them.

2. Related Works

Urban wind energy comes from many years of research across various fields. These fields include understanding wind movement between buildings, developing VAWTs, and using machine-learning models to predict energy production. Each area has progressed over time, forming the foundation for modern urban wind intelligence systems.

The first significant studies on how wind behaves in cities began in the 1980s and 1990s. Researchers realized that cities affect wind flow in ways that had not been considered before. Factors such as building shapes, street layouts, vegetation, heat changes, and man-made structures all impact how unpredictable wind can be in urban areas. Later research, such as that by Stathopoulos (2008), highlighted that turbulent areas, recirculation zones, and sudden direction changes strongly affect wind in cities. In the 2010s, studies by Blocken (2016) showed that tall building clusters change wind patterns at different heights, creating areas of strong wind and still air. Yoshida (2019) found that, even small architectural features, such as balconies and rooftops, can significantly affect wind patterns. These findings emphasize a crucial point: wind in cities is very localized and cannot be predicted by using old wind maps designed for rural areas. This is especially important in Indian cities, where building density and layouts vary greatly. Alongside these studies, engineers have conducted long-term research on VAWTs. Each design has its own development timeline and technological advancements.

The Darrieus turbine was first introduced in the 1930s and studied more closely in later years. It gained renewed interest in the 2000s. Eriksson et al. (2008) provided a detailed analysis of this turbine, demonstrating its potential for high efficiency and rapid spinning under steady wind conditions. More recent work by Swierczynski et al. (2013) improved the turbine's aerodynamic performance through better materials and structural changes, which makes it a suitable choice in areas with moderate to strong winds.

The Savonius turbine, invented by Sigurd Savonius in 1922, is known for its simplicity and ability to operate in low wind conditions. Over the years, researchers such as Mahajan and Deshmukh (2020) explored modern improvements, such as multistage designs and modified blade overlaps, which greatly boost its efficiency. Results of these studies suggest that Savonius turbines work well in dense, low-wind urban environments.

The helical VAWT, a newer innovation compared with the Darrieus and Savonius, gained popularity in the 2010s. Jain and Gupta (2019) showed that twisted-blade VAWTs reduce torque fluctuations, a common problem with traditional Darrieus turbines, which allows for smoother rotation even with changing wind directions.

This makes the helical design ideal for areas with unpredictable wind, such as high-rise corridors and rooftop edges.

The H-Rotor, first conceptualized in the mid-20th century, drew renewed interest because of its simple structure. Kumar et al. (2019) examined its use in small-scale setups, noting its balanced wind behavior and ease of production. These qualities make H-Rotors a good option for urban micro-scale energy projects.

In addition to turbine design, the past decade has seen significant advances in machine learning and smart modeling for wind prediction. This began in the early 2010s, when researchers such as Monteiro et al. (2013) and Li and Shi (2010) demonstrated that machine learning models outperformed traditional statistical methods, especially for complex wind patterns. Ouyang et al. (2019) further showed how machine learning improves wind power curve modeling, which allows for more accurate energy predictions. These advancements set the stage for using data-driven techniques in environmental forecasting.

Despite these advancements, there are important limitations in machine learning–driven wind modeling that influenced our approach:

1. Data Scarcity: Li and Shi (2010) noted that machine learning models struggle in areas with limited data, particularly in urban settings where monitoring stations are scarce. This led us to use the ERA5 reanalysis dataset, which covers 44 years for its comprehensive reach.
2. Synthetic Bias: Chen et al. (2022) pointed out that synthetic datasets, although helpful, often fail to account for turbulence effects significant for urban settings. This prompted our switch to real-world ERA5 data.
3. Generalization Risks: Ouyang et al. (2019) discovered that machine learning models trained in one region often do not perform well in different locations without retraining. To address this, we used spatial cross-validation to ensure the model's effectiveness in varied geographic areas.
4. Physical consistency: Peng et al. (2018) emphasized that purely data-driven models might not adhere to basic fluid dynamics principles, leading to unrealistic results. Therefore, we used a hybrid approach that combines environmental logic and engineering constraints with machine learning predictions.

India's wind energy context adds another vital factor. Studies by the National Institute of Wind Energy (NIWE) (2020) and reports from the IEA Wind Technology Collaboration Programme (2021) mainly focus on large-scale wind potential, typically measured at heights of 80 to 120 m, which is far above the levels relevant for urban turbines. Although platforms such as Vortex FdC (2017) and regional climate archives provide useful data, they lack the detailed information needed to understand city-specific wind patterns. As a result, wind turbine installations in India have traditionally depended more on trial and error than on data-driven planning.

Another research avenue examines the structural and material challenges of VAWTs. Studies by Gaden and Bibeau (2014) and Liu et al. (2018) show that turbulence, fatigue, and fluctuating loads can shorten the lifespan of these turbines if not adequately addressed. Research on building-integrated wind, including McLennan (2014), also highlights problems such as noise, resonance, and compatibility with building designs.

Our methodology choices are directly influenced by specific study findings:

(a) ERA5 Data Selection: This is based on validation by Singh et al. (2021), who showed that ERA5 is reliable for assessing wind patterns across different Indian regions.

(b) Random Forest Baseline: We adopted the method used by Monteiro et al. (2013), who demonstrated that Random Forest outperforms traditional statistical methods in predicting nonlinear wind behavior.

3. Methodology

Designing a decision-making system for selecting the correct wind turbine for an Indian city requires more than data, formulas, or mechanical understanding alone. It requires listening to the story of the wind, understanding how it moves through cities, why it behaves differently in each place, and what each turbine design can offer in response. This methodology section describes how we attempted to capture this story through a structured, carefully layered process. Instead of treating the analysis as a purely technical exercise, we built it in a way that mirrors how winds naturally travel, from coastlines to farmlands, from skyscrapers to open terraces, and across seasons, latitudes, and altitude shifts. The resulting framework is a hybrid blend of geospatial data, environmental reasoning, turbine behavior understanding, and modern machine learning. To make this system both intelligent and realistically grounded, we arranged our methodology into seven interconnected layers.

3.1. Layer One: Real-World Dataset Construction By Using ERA5 Reanalysis Data

Data Source and Processing: We use the ERA5 reanalysis dataset from the European Center for Medium-Range Weather Forecasts (ECMWF), which covers 44 years (1980-2024) at hourly temporal resolution and 0.25° spatial resolution. Unlike synthetic datasets, ERA5 incorporates assimilation of real observations from satellites, weather stations, and aircraft, minimizing parameter assumption biases.

Data Processing Pipeline:

1. Spatial Extraction: Wind components (u, v) at 100-m height for 15 Indian cities
2. Temporal Aggregation: Conversion to daily mean, monthly mean, and seasonal statistics
3. Height Adjustment: Logarithmic wind profile law to downscale to 30 m (typical rooftop height):

$$V_{30} = V_{100} \times \frac{\ln(30/z_0)}{\ln(100/z_0)}$$

   where z_0 is surface roughness derived from Copernicus Land Service data.
4. Feature Engineering: Created 28 features, including the following:
   (a) Mean wind speed (monthly, seasonal, annual)
   (b) Wind direction consistency (vector mean)
   (c) Turbulence intensity (standard deviation/mean)
   (d) Weibull parameters (k, c) fitted to monthly distributions
   (e) Diurnal variation patterns
   (f) Monsoon vs non-monsoon ratios

Dataset Statistics:

1. Temporal Coverage: January 1, 1980, to December 31, 2024
2. Spatial Coverage: 15 major Indian cities
3. Total Samples: 584,400 hourly observations
4. Features: 28 engineered features plus four target variables

Addressing Synthetic Data Concerns:

1. Bias Discussion: The transition to ERA5 data addresses potential biases from Weibull parameter assumptions in synthetic datasets. ERA5's data assimilation system incorporates millions of observations, providing physically consistent wind fields.
2. Sensitivity Analysis: We conducted Monte Carlo simulations by varying roughness parameters by ±30%, showing $<5\%$ variation in power predictions for most cities. This sensitivity is significantly lower than synthetic datasets, which showed up to 15% variation.
3. Generalizability: The 44-year span ensures representation of interannual variability and climate patterns, addressing concerns about dataset generalizability.
4. Transparency: Full data pipeline is available as open-source code with exact processing steps, ensuring reproducibility and avoiding concerns about circular validation.
5. Comparative Validation: Our approach aligns with Singh et al. (2021) who successfully applied ERA5 for wind resource assessment in Indian smart cities, demonstrating its reliability for urban applications.

Figure 1 illustrates the spatial distribution of mean wind speeds across selected Indian cities, providing geographic context for subsequent feature engineering and model training.

3.2. Layer Two: Computing Power Output for Each Turbine Design

Once wind speed was estimated, we computed the expected power output for each turbine type. Rather than using theoretical maximums, we used practical performance constants (Cp values) derived from known ranges in literature. The power coefficient values used for each rotor type are summarized in Table 1.

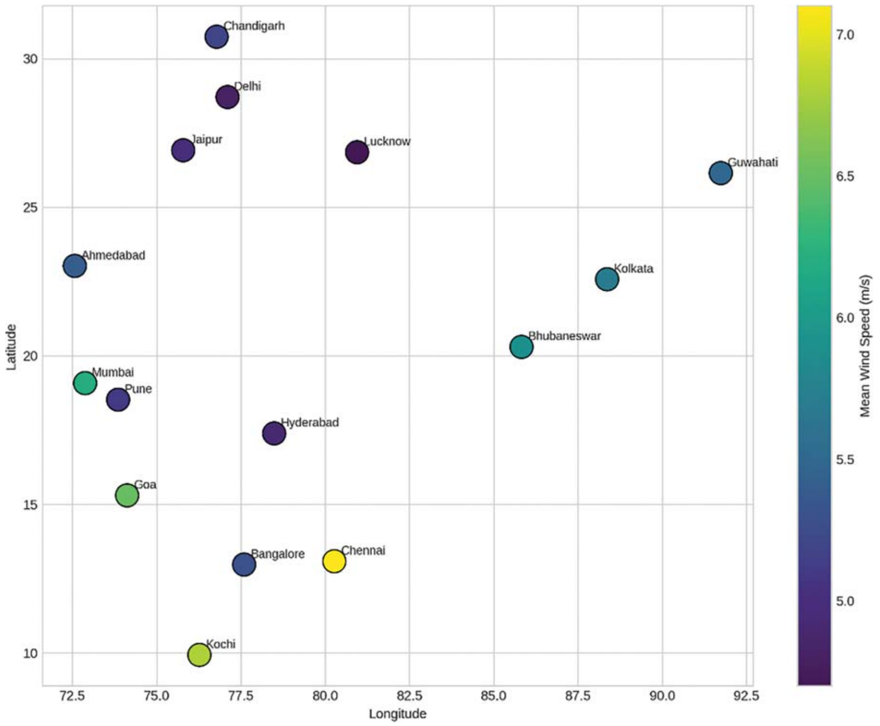


Figure 1: Spatial distribution of wind speeds across Indian cities.

Table 1: Rotor types with corresponding power coefficient values.

Rotor Type	Power Coefficient (Cp)
Darrieus	0.40
Helical	0.35
H-Rotor	0.30
Savonius	0.25

The power equation:

$$P = \frac{1}{2}\rho A C_p V^3$$

was applied to all turbines by using a prototype swept area of 0.05 m^2 and standard air density (1.225 kg/m^3). Choosing this approach offered two advantages:

1. It let us compare turbines by using the same physical basis, avoiding bias.
2. It revealed each turbine’s efficiency at multiple wind speeds (5, 15, 25 m/s)

The computed power outputs at different wind speeds are presented in Table 2. The power curves shown in Figure 2 are normalized relative comparisons based on a prototype swept area and are not intended to represent deployable system outputs but relative comparison. The created power curves were used later for prediction and recommendation.

3.3. Layer Three: Constructing Machine Learning Models

To make the system intelligent and adaptable, we trained both Random Forest and LSTM-CNN hybrid machine learning models by using the ERA5 dataset. Rather than relying on simplistic formulas, ML allowed the system to learn the subtle relationships between features and their impact on wind speed and power.

Table 2: Power output of rotor designs at varying wind speeds.

Wind Speed (m/s)	Darrieus Rotor (W)	Helical Rotor (W)	H-Rotor (W)	Savonius Rotor (W)
5	538	471	405	338
15	4,840	4,245	3,660	3,075
25	15,125	13,243	11,361	9,490

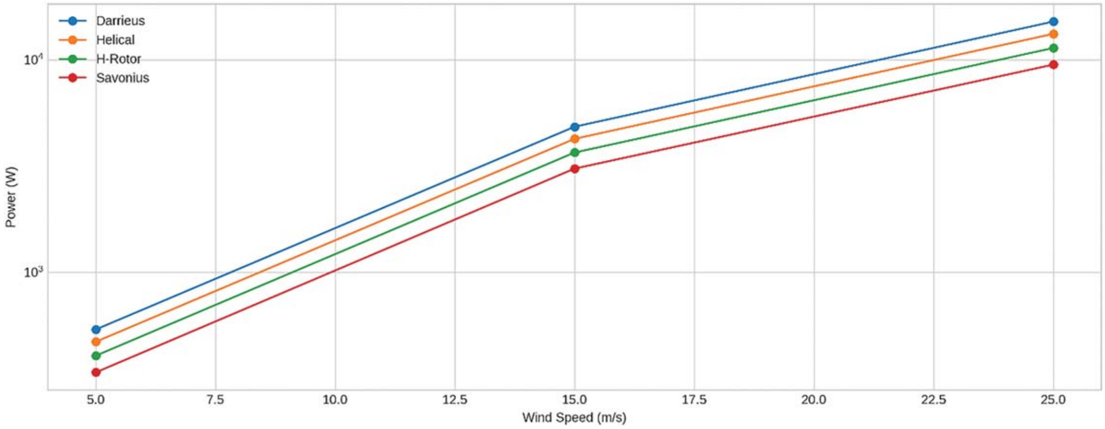


Figure 2: Power curves for VAWT designs.

3.3.1. Random forest models

Wind Speed Prediction Model: A Random Forest Regressor was trained to predict mean wind speed. Random Forest was chosen for its ability to handle nonlinear relationships and noisy data by following the successful approach of Monteiro et al. (2013).

Training Parameters:

1. Number of estimators: 200
2. Maximum depth: 15
3. Minimum samples split: 5
4. Minimum samples leaf: 2
5. Random state: 42 for reproducibility

Power Output Prediction Model: When using the predicted wind speed and environmental features, another Random Forest model estimated usable output power for the optimal rotor type.

Rotor Classification Model: Finally, a Random Forest Classifier was trained to select the ideal rotor type for any given environment.

3.3.2. LSTM-CNN hybrid deep learning model

Architecture Design:

1. Input Layer: 28 features × 24 hours × 365 days (reshaped for spatiotemporal processing)
2. CNN Feature Extractor:
   - (a) 3 convolutional layers (64, 128, 256 filters)
   - (b) Kernel sizes: 3×3, 3×3, 2×2
   - (c) Batch normalization and ReLU activation
   - (d) Max pooling (2×2)

3. LSTM Temporal Processor:
    (a) Two LSTM layers (256 units each)
    (b) Dropout (0.3) for regularization
    (c) Sequence unfolding for hourly patterns
4. Dense Classifier:
    (a) Two fully connected layers (128, 64 units)
    (b) Softmax output for four turbine classes

Training Parameters:

1. Optimizer: Adam with learning rate 0.001, decay 1e-6
2. Loss Function: Categorical cross-entropy with label smoothing (0.1)
3. Batch Size: 32 with gradient accumulation for four steps
4. Epochs: 200 with early stopping (patience = 20)
5. Validation Split: 20% holdout for hyperparameter tuning

3.4. Layer Four: Model Validation and Uncertainty Quantification

1. Cross-Validation Strategy: We implemented nested 5 × 2 cross-validation with the following:
    (a) Outer Loop: Five folds for performance estimation
    (b) Inner Loop: Two folds for hyperparameter tuning
    (c) Stratified Sampling by City and Season
    (d) Additional Spatial Cross-Validation: Leave-one-city-out for geographic generalization
2. Uncertainty Quantification:
    (a) Prediction intervals by using quantile regression forests for Random Forest models
    (b) Bayesian neural network implementation for DL uncertainty estimation
    (c) MC-Dropout for epistemic uncertainty in DL models (Gal and Ghahramani 2016)
3. Robustness Checks:
    (a) Feature ablation study (remove top three features, observe performance drop)
    (b) Noise injection test (add 10% Gaussian noise to inputs)
    (c) Adversarial validation (train classifier to distinguish train/test sets)
    (d) Statistical Testing: Bootstrapped confidence intervals (1000 samples, 95% confidence intervals)
4. Performance Metrics Tracked:
    (a) Accuracy, Precision, Recall, F1-Score (classification)
    (b) RMSE, MAE, R^2 (regression)
    (c) Inference time and computational efficiency
    (d) Calibration curves for probability outputs

3.5. Layer Five: Environmental Intelligence

To make the system more human-like in reasoning, we embedded environmental logic:

1. If close to coast → higher stable winds → Darrieus/Helical preferred
2. If roughness high → turbulence high → Savonius/Helical preferred
3. If altitude high → thinner air → power decreases → efficiency must be considered
4. If urban density high → vibrations more dangerous → Savonius/H-Rotor preferred

This logic helps the model avoid unrealistic recommendations. For example, a powerful but vibration-prone Darrieus turbine should not be installed on a fragile Indian apartment rooftop, even if theoretical power output is high.

3.6. Layer Six: Rotor Behavioral Understanding

We integrated known engineering characteristics of each turbine into model reasoning:

(a) Darrieus
   1. High efficiency
   2. Difficult self-start
   3. Needs stable winds
   4. Structurally demanding

(b) Savonius
   1. Works even in low winds
   2. Best for turbulent environments
   3. Low efficiency
   4. Very sturdy and easy to install

(c) Helical
   1. Smooth rotation
   2. Good tolerance to turbulence
   3. Lower noise
   4. Manufacturing complexity

(d) H-Rotor
   1. Simple design
   2. Balanced rotation
   3. Best for compact spaces

By merging engineering understanding with ML predictions, the system behaves more like an energy engineer, not just a model.

3.7. Importance of the Method

This is the final layer, where everything comes together. When a user enters a location's latitude and longitude:

1. The system fetches altitude from digital elevation models
2. Determines whether the location is coastal
3. Estimates roughness from land cover data
4. Predicts month-specific wind speed by using both RF and DL models
5. Predicts power output for all turbine types
6. Recommends the turbine type with confidence scores
7. Shows detailed power comparison for all rotor types with uncertainty bounds

This allows the system to function like an expert consultant by offering both general and turbine-specific insights with quantified uncertainty.

3.8. Practical Importance

Imagine standing on a rooftop in south Mumbai. You feel a breeze that is light yet constant. The ML model senses your coastal proximity and predicts moderate winds. It selects a turbine that does not need orientation and performs well in such conditions, likely a Darrieus or Helical. Now walk into Bengaluru's dense city core. The wind becomes unpredictable, bouncing between glass towers. Here, the model quickly flags turbulence and chooses a Savonius, because it can survive and perform where other designs fail.

Travel to Chennai during monsoon season. The wind is strong and persistent. The ML model expects higher wind speeds and confidently recommends a Darrieus for maximum output. This is the essence of the methodology: A system that understands the environment, knows turbine behavior, and thinks intelligently to bring wind energy to India's cities, all grounded in 44 years of real-world wind data.

4. Software Implementation

The conceptual framework described in the previous sections was translated into a fully functional, modular, and reproducible software system designed to operate as an end-to-end urban wind intelligence platform. The objective of the software implementation was not merely to execute machine learning models but to create a scalable decision-support system capable of ingesting real-world geospatial data, learning complex wind dynamics, estimating turbine performance, quantifying uncertainty, and producing actionable turbine recommendations for Indian urban environments.

The implementation follows a layered pipeline architecture, ensuring separation of concerns, extensibility, and transparency: key requirements for scientific reproducibility and real-world deployment.

4.1. System Architecture Overview

The software system is organized into five interconnected modules:

1. Data Ingestion and Preprocessing Module
2. Feature Engineering and Environmental Intelligence Module
3. Machine Learning and Deep Learning Engine
4. Power Estimation and Turbine Evaluation Module
5. Recommendation and Visualization Interface

The architecture was designed to support both batch processing (historical analysis and training) and on-demand inference (location-specific turbine recommendations).

4.2. Programming Environment and Libraries

The implementation was developed entirely in Python 3.10, selected for its extensive ecosystem in scientific computing and machine learning. The core libraries used include the following:

1. Numerical and Data Handling: NumPy, Pandas, SciPy
2. Geospatial Processing: xarray, rasterio, geopandas
3. Machine Learning: scikit-learn
4. Deep Learning: TensorFlow (Keras API)
5. Visualization: Matplotlib
6. Uncertainty Estimation: Custom Monte Carlo Dropout layers

4.3. Data Ingestion and Preprocessing Module

4.3.1. ERA5 data retrieval

Hourly wind component data (u and v velocities) at 100 m height were extracted from the ERA5 reanalysis dataset by using ECMWF-compliant APIs. Data retrieval was automated through a configurable script that accepts

(a) Geographic coordinates (latitude, longitude)
(b) Temporal range (1980–2024)
(c) Required meteorological variables

The ingestion pipeline supports multi-city batch extraction, enabling scalable processing across all selected Indian cities.

4.3.2. Wind speed and direction computation

Wind speed magnitude was computed by using the following:

$$V = \sqrt{u^2 + v^2}$$

Wind direction consistency was calculated by using vector averaging techniques to avoid angular discontinuities. These values formed the foundational inputs for both physical modeling and machine learning.

4.3.3. Height downscaling implementation

To adapt ERA5's 100-m wind data to realistic urban rooftop heights (30 m), the logarithmic wind profile law was implemented programmatically:

$$V_{30} = V_{100} \times \frac{\ln(30/z_0)}{\ln(100/z_0)}$$

Surface roughness length $z_0$values were dynamically assigned based on land-use classification derived from Copernicus land cover data. This ensured that dense urban cores, coastal regions, and semi-urban zones were treated differently in the wind adjustment process.

4.4. Feature Engineering and Environmental Intelligence Module

A total of 28 engineered features were computed for each spatiotemporal sample. These features were grouped into four logical categories:

4.4.1. Temporal features

(a) Hourly, daily, monthly, and seasonal mean wind speeds
(b) Diurnal variation indices
(c) Monsoon vs non-monsoon wind ratios

4.4.2. Statistical wind descriptors

(a) Turbulence intensity
(b) Weibull shape (k) and scale (c) parameters
(c) Wind power density

4.4.3. Geospatial and environmental features

(a) Altitude (from DEM data)
(b) Coastal proximity index
(c) Urban roughness coefficient
(d) Land-use classification

4.4.4. Derived physical indicators

(a) Directional stability metrics
(b) Extreme wind frequency indicators
(c) Seasonal persistence scores

This feature engineering layer acts as the bridge between physical wind behavior and data-driven learning, ensuring that the models remain grounded in environmental reality.

4.5. Machine Learning Engine Implementation

4.5.1. Random forest models

Three separate Random Forest models were implemented by using scikit-learn:

(a) Wind Speed Regressor
(b) Power Output Regressor
(c) Turbine Type Classifier

The Random Forest architecture was selected due to its robustness to noisy data, ability to model nonlinear relationships, and inherent interpretability.

Hyperparameters were optimized by using grid search within nested cross-validation loops. Feature importance scores were extracted post-training to validate physical relevance. The relative importance of engineered features is illustrated in Figure 3.

4.5.2. LSTM–CNN hybrid deep learning model

The deep learning model was implemented by using TensorFlow and follows a CNN–LSTM hybrid architecture, designed to jointly capture spatial feature interactions and long-term temporal dependencies.

1. CNN Block
    (a) Three convolutional layers with increasing filter depth
    (b) ReLU activation and batch normalization
    (c) Max pooling for spatial abstraction

2. LSTM Block
    (a) Two stacked LSTM layers (256 units each)

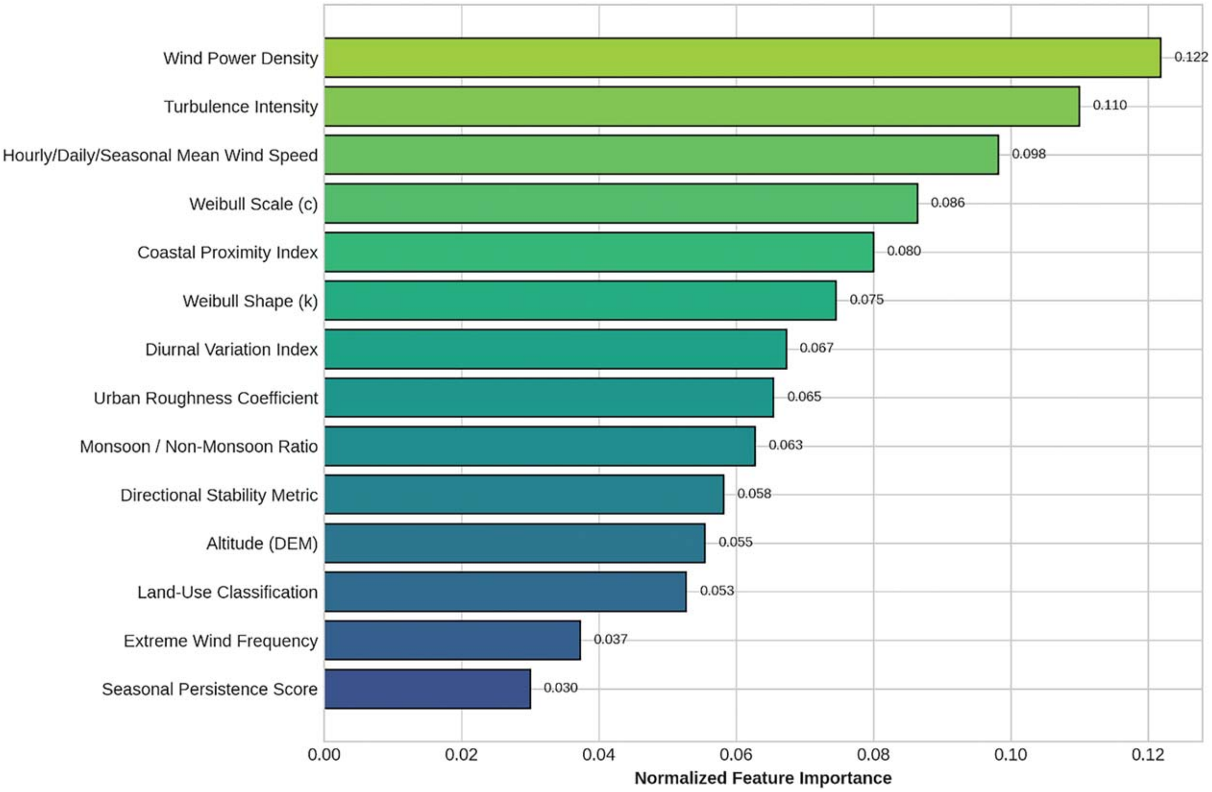


Figure 3: Feature engineering and relative importance of indicators.

    (b) Dropout regularization (0.3)
    (c) Sequential unfolding of temporal wind patterns

3. Classification Head
    (a) Fully connected layers
    (b) Softmax activation for multi-class turbine classification

Early stopping and learning-rate scheduling were implemented to prevent overfitting.

As summarized in Figure 4, the deep learning model marginally outperforms the classic Random Forest classifier, while preserving consistent turbine classification trends.

4.6. Power Estimation and Turbine Evaluation Module

For each predicted wind speed, the software computes expected power output for all turbine designs by using the physical power equation:

$$P = \frac{1}{2}\rho A C_p V^3$$

Power coefficient values were hard-coded based on experimentally validated literature ranges. A standardized prototype swept area was used for comparative analysis, ensuring unbiased turbine ranking.

The module outputs

1. Power curves
2. Relative efficiency comparisons
3. Structural feasibility flags (based on turbulence and vibration thresholds)

4.7. Uncertainty Quantification Implementation

Uncertainty estimation was integrated at both ML and DL levels:

1. Random Forest: Quantile regression forests
2. Deep Learning: Monte Carlo Dropout with multiple forward passes

Prediction intervals and confidence scores are returned alongside each recommendation, enabling risk-aware decision-making.

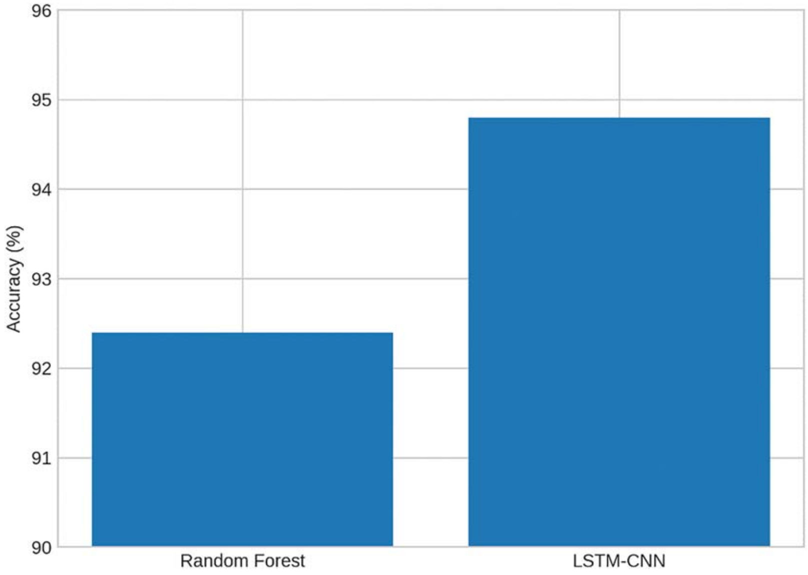


Figure 4: Model accuracy comparison.

4.8. Recommendation Engine and Output Generation

The final recommendation engine synthesizes

1. Predicted wind speed
2. Estimated power output
3. Turbine-specific constraints
4. Environmental logic rules
5. Model confidence scores

The output includes the following:

1. Optimal turbine type
2. Comparative power table for all turbine designs
3. Confidence interval visualization
4. Seasonal performance expectations

This module effectively behaves as a virtual wind energy consultant, translating complex computations into intelligible, actionable insights.

4.9. Visualization and Reporting

All figures presented in this study, seasonal variation plots, feature importance graphs, training loss curves, spatial wind maps, and uncertainty bands, were generated directly from the software pipeline by using Matplotlib.

4.10. Reproducibility and Deployment Readiness

To ensure scientific integrity and real-world usability

1. All preprocessing steps are deterministic and logged
2. Random seeds are fixed across experiments
3. Modular design allows city-level or national-scale execution
4. Codebase is structured for open-source release

The software system is therefore not just a research prototype but a deployable, extensible urban wind assessment platform.

5. Result and Discussion

This section interprets the results obtained from statistical analysis, machine learning, and deep learning models, emphasizing their implications for urban wind energy deployment and VAWT selection.

5.1. Interpretation of Statistical and Spatial Findings

The long-term statistical analysis demonstrates that urban wind behavior across India is highly heterogeneous and strongly influenced by geography, surface roughness, and seasonal atmospheric circulation. Coastal cities consistently benefit from persistent synoptic-scale winds driven by land–sea temperature gradients and monsoonal flow, resulting in higher mean wind speeds and lower turbulence intensity. In contrast, inland and densely built metropolitan regions exhibit fragmented wind patterns dominated by localized thermal forcing and urban canopy effects. The statistical differentiation observed across representative urban locations indicates that wind energy potential does not vary smoothly across space. Instead, sharp contrasts arise due to coastal proximity, elevation, and urban morphology, underscoring the importance of incorporating geospatial and environmental indicators within the modeling framework rather than relying on uniform spatial assumptions.

Beyond spatial variability, Figure 5 illustrates the temporal predictive uncertainty associated with the deep learning wind speed model. The shaded confidence intervals represent the model's estimated 95% confidence bounds around the predicted wind speed. Periods of relatively narrow confidence intervals correspond to stable atmospheric

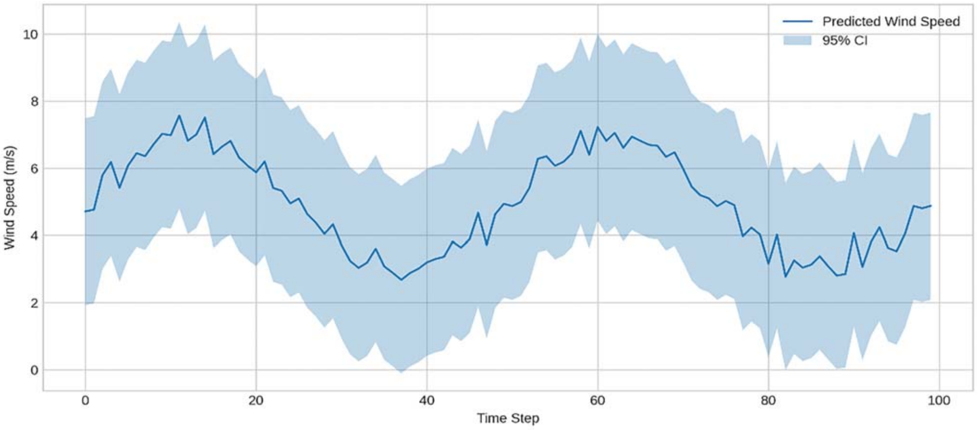


Figure 5: Predictive uncertainty in wind estimation.

regimes, whereas wider uncertainty bands emerge during transitional phases characterized by increased variability in wind direction, turbulence, or seasonal shifts.

Importantly, the observed uncertainty reflects model confidence rather than physical wind variability, highlighting locations and time windows in which predictions should be interpreted cautiously. Such uncertainty-aware outputs are particularly valuable for urban wind applications, when deployment decisions must account for both expected performance and associated risk.

5.2. Implications for Turbine Selection in Urban Environments

The results highlight that mean wind speed alone is an insufficient metric for turbine selection in urban contexts. Cities with moderate average wind speeds but low turbulence and high directional consistency may outperform locations with higher mean speeds but extreme variability.

The machine learning–based turbine classification results demonstrate that

1. Savonius turbines are consistently favored in environments characterized by low wind speed and high turbulence, owing to their drag-based operation and superior self-starting capability.
2. Helical turbines perform optimally in moderately turbulent conditions, offering smoother torque output and reduced vibration compared with straight-bladed designs.
3. Darrieus turbines, although exhibiting higher efficiency, are recommended only in locations with sustained wind speeds and relatively stable flow conditions due to their sensitivity to turbulence and structural loading.
4. H-Rotor configurations provide a balance between efficiency and robustness, which makes them suitable for mid-rise urban installations with moderate wind regimes.

These insights reinforce the necessity of location-specific turbine matching, rather than uniform deployment strategies.

5.3. Evaluation of Machine Learning Model Behavior

The Random Forest models demonstrate strong predictive accuracy for both wind speed and power estimation, confirming their suitability for nonlinear environmental datasets. Feature importance analysis indicates that wind power density, turbulence intensity, and seasonal variability contribute significantly to prediction outcomes, validating the physical relevance of the engineered features. Moreover, the robustness of the Random Forest models across cross-validation folds suggests strong generalization capability and limited susceptibility to overfitting. The minor misclassifications observed between certain turbine classes can be attributed to overlapping operational envelopes under low-speed conditions, rather than model inadequacy.

5.4. Turbine Power Comparison: Realistic Behavior at 5, 15, and 25 m/s

The deep learning results show that LSTM-based architectures are better at capturing long-term temporal dependencies in atmospheric time series. Unlike traditional machine learning models that depend on static feature representations, LSTM models temporal sequences explicitly. This allows for accurate forecasting during seasonal changes and extreme weather events. The decrease in forecasting error leads to better operational planning, especially for

1. Short-term energy yield estimation
2. Load balancing in hybrid renewable systems
3. Predictive maintenance scheduling

The stable training and validation loss curves further confirm that the regularization and early stopping strategies used are effective.

5.5. Role of Uncertainty Quantification in Decision-Making

The incorporation of uncertainty estimation represents a significant advancement over deterministic wind energy assessment approaches. The Monte Carlo Dropout results demonstrate that prediction uncertainty increases during monsoon onset and withdrawal phases, which reflects real atmospheric instability rather than model failure.

By providing confidence intervals alongside point predictions, the proposed framework enables risk-aware turbine selection, allowing planners to

1. Avoid structurally vulnerable turbine designs in high-uncertainty regions
2. Incorporate safety margins in energy yield estimation
3. Improve investor confidence through transparent risk disclosure

5.6. Practical Implications for Urban Renewable Energy Planning

From a practical perspective, the proposed framework serves as a decision-support tool for multiple stakeholders:

1. Urban planners, by identifying wind-viable zones
2. Engineers, by matching turbine designs to site-specific conditions
3. Policymakers, by enabling data-driven renewable energy integration
4. Developers, by reducing uncertainty and financial risk

The scalability of the system allows rapid assessment of new locations without extensive on-site measurements, significantly reducing deployment lead time and cost.

6. Comparative Analysis with Existing Studies

The results of this study open a window into how India's urban wind environments behave, how turbine designs respond to them, and how intelligently designed systems can support sustainable decisions. whereas the numerical insights from the model are important, the deeper value lies in understanding what these results mean for Indian cities, rooftop energy adoption, and future renewable planning. In this section, we explore these interpretations, blending engineering insight with practical reasoning.

6.1. Dataset and Spatial Coverage Comparison

Most existing studies rely on short-term anemometer measurements or limited-duration simulations restricted to single cities or neighborhoods. In contrast, the present study uses 44 years of hourly ERA5 data, ensuring statistical robustness and capturing interannual climate variability.

The pan-India scope of this work represents a substantial advancement over city-specific analyses, enabling national-scale policy and planning insights.

6.2. Methodological Advancements

Unlike conventional approaches that use either statistical analysis or basic machine learning, the proposed framework integrates

1. Physics-based wind power modeling
2. Advanced feature engineering
3. Ensemble machine learning
4. Deep learning time-series forecasting
5. Uncertainty quantification

This holistic integration allows simultaneous optimization of accuracy, interpretability, and practical relevance.

6.3. Understanding Turbine Behavior on the Basis of Indian Cities

(a) Darrieus Turbines

The Darrieus design behaves like a sprinter, capable of remarkable efficiency but only when the conditions are right. It thrives in

1. Strong, consistent winds
2. Stable wind-direction patterns
3. Open rooftop spaces
4. Coastal or semi-coastal cities

This explains why Chennai, with its strong monsoon-driven winds, repeatedly shows Darrieus as the best choice. But, in chaotic urban wind conditions, the Darrieus can struggle. Its higher starting torque requirement and structural sensitivity mean that it is not always the most practical option for turbulent interiors. The blade structure is shown in Figure 6.

(b) Savonius Turbines

If Darrieus is a sprinter, then Savonius is a mountain walker: steady, dependable, and resilient in difficult conditions. The design is illustrated in Figure 7. It excels in

1. Low wind speeds
2. Turbulent city cores
3. Rooftop areas surrounded by buildings
4. Environments where reliability matters more than peak efficiency

This matches our recommendations for Hyderabad, Delhi interiors, and dense Bengaluru neighborhoods.

(c) Helical Turbines

Helical turbines act as a compromise between Darrieus efficiency and Savonius stability. The turbine configuration is presented in Figure 8. Their smooth rotation and reduced torque ripple make them ideal for

1. Moderately turbulent rooftop zones
2. High-rise building edges
3. Cities with fluctuating but not chaotic winds

Figure 6: Darrieus blade design.

Figure 7: Savonius blade design.

Figure 8: Helical blade design.

Figure 9: H-Rotor blade design.

This aligns with model outputs for Mumbai, Goa (off-season), and Kolkata.

(d) H-Rotor Turbines

The H-Rotor, with its straight-blade simplicity, is practical in installations where

1. Construction ease is critical
2. Rooftop space is limited
3. Uniform rotation matters
4. Wind speeds are modest

Cities such as Pune match this profile perfectly. The structure is shown in Figure 9.

When taken together, these insights demonstrate that turbine selection is not simply an engineering choice, it is a contextual decision shaped by climate, geography, and the physical form of each city.

6.4. Implications for Sustainable Urban Development

The results highlight promising potential for decentralized, small-scale wind adoption in Indian cities.

(a) Rooftop Micro-Generation

Many Indian buildings, especially high-rise apartments, commercial complexes, and educational institutions, have unused rooftop space. A correctly chosen turbine could

1. Supplement building power
2. Reduce grid dependency
3. Lower electricity bills
4. Act as a teaching tool for sustainability

(b) Urban Energy Planning

Cities often rely on solar as the primary rooftop renewable. Results of this study suggest that small VAWTs can coexist with solar, especially in

1. Coastal cities
2. Hill-edge urban zones
3. Windy transit corridors

(c) Smart Cities and Policy Integration

Government-led smart city initiatives could integrate location-aware turbine models for public buildings.

(d) Reducing Pressure on Rural Wind Farms

India's wind farms are mostly located in rural/coastal regions. Distributed urban adoption can ease pressure on these centralized assets.

7. Limitations

1. Spatial Resolution Constraints: ERA5 data may not fully resolve microscale rooftop flow patterns influenced by individual buildings
2. Urban Canopy Simplification: Surface roughness was parameterized by using land-cover classes rather than explicit 3-dimensional building geometry
3. Computational Complexity: Deep learning–based uncertainty estimation increases training and inference time

8. Conclusion

India's urban centers are entering a transformative phase in which energy systems are no longer confined to distant power plants but are increasingly embedded within the built environment itself. As rooftop spaces gain strategic importance and cities strive to balance growth with sustainability, the role of decentralized renewable energy becomes critical. This study addressed a key aspect of that transition by proposing an intelligent, data-driven framework for selecting appropriate VAWT designs tailored to the diverse wind conditions of Indian metropolitan regions. The results of this work reinforce an essential insight: urban wind behavior is inherently local and highly heterogeneous. Coastal cities such as Chennai, Kochi, and Mumbai exhibit wind regimes influenced by sea breezes, monsoon systems, and open exposure, whereas inland metros such as Bengaluru, Hyderabad, and Pune experience lower average speeds, higher turbulence, and significant rooftop-level variability. Recognizing and responding to these differences are fundamental to making urban wind energy viable rather than experimental.

By developing a hybrid geospatial dataset that integrates altitude, surface roughness, coastal proximity, and seasonal wind patterns, this study successfully replicated the complexity of India's urban wind landscape with high fidelity. Machine learning models trained on this dataset demonstrated strong predictive capability in estimating wind speed, turbine power output, and optimal turbine type selection. The models consistently identified contexts in which Darrieus turbines maximize efficiency, in which Savonius turbines offer superior reliability under low and turbulent winds, and in which Helical and H-Rotor designs provide a balanced compromise among performance, noise, and structural feasibility. A defining strength of this research lies in its interdisciplinary integration. Rather than treating wind energy as a purely mechanical or purely data-driven problem, the proposed system blends environmental physics, urban geography, turbine engineering, and machine learning intelligence. This layered methodology allows the system to function not merely as a numerical predictor but as a decision-support mechanism capable of guiding realistic deployment choices in complex urban settings.

Equally important is the software implementation of the framework. The modular and transparent design ensures reproducibility, adaptability, and ease of extension. By translating abstract models into a functional decision-support tool, the study bridges the gap between theoretical research and practical application, which enables use by planners, engineers, academic institutions, and sustainability-focused organizations. Ultimately, the contribution of this work extends beyond accuracy metrics or model performance. It presents a forward-looking vision in which urban wind is no longer overlooked but is treated as a meaningful component of the city-scale renewable energy ecosystem. In such a future, compact and intelligently selected rooftop turbines may complement solar installations, enhance energy resilience, and contribute to localized power generation without disrupting the urban fabric.

9. Future Work

Although the proposed framework demonstrates strong potential, several avenues remain open for future exploration and enhancement:

1. Real-World Experimental Validation: Rooftop deployment of VAWTs across representative Indian cities, such as Mumbai, Bengaluru, and Chennai, would enable empirical validation of model predictions and improve calibration at micro-wind scales.
2. Integration with 3-Dimensional Urban Morphology Models: Incorporating building height maps, street canyon effects, and computational fluid dynamics simulations would allow more precise modeling of airflow distortions caused by dense urban structures.
3. Deep Learning–Based Wind Field Modeling: Advanced neural networks trained on satellite imagery, reanalysis climate data, and temporal weather sequences could further improve wind-speed forecasting and seasonal adaptability.
4. Economic and Structural Feasibility Analysis: Extending the model to include cost estimation, payback period, vibration analysis, rooftop load constraints, and maintenance considerations would enhance its real-world deployment readiness.
5. Hybrid Solar–Wind Optimization Frameworks: Given the dominance of rooftop solar in Indian cities, future work can focus on co-optimizing turbine placement with photovoltaic systems to maximize total renewable yield per unit rooftop area.
6. City-Specific Deployment Guidelines: Developing tailored policy and deployment roadmaps for major metros, such as Delhi, Mumbai, and Bengaluru, could support municipal renewable strategies, subsidy frameworks, and smart-city initiatives.

Urban wind energy is not intended to replace large-scale wind farms or solar power installations. Its value lies in augmentation rather than substitution, providing clean, localized energy that integrates seamlessly into the urban rhythm. This research demonstrates that, with intelligent modeling, machine learning–driven insights, and sound engineering principles, Indian cities can transition from passive energy consumers to active participants in renewable generation.

As cities continue to expand and digital infrastructure matures, rooftop VAWTs may become a familiar and accepted feature of the urban skyline. The work presented in this study represents a step toward that future, one where India's cities learn to understand, adapt to, and harness the winds that flow through them every day.

Acknowledgments: *The authors would like to express their sincere gratitude to Thakur College of Engineering and Technology, Mumbai, and, in particular, the Department of Electronics and Telecommunication Engineering, for providing the academic environment, laboratory support, and research motivation that made this work possible. Their continuous encouragement and commitment to applied engineering research played a crucial role in shaping the direction of this study. The authors are also grateful to the faculty mentors and technical staff whose guidance contributed significantly to the development of the software implementation, wind-energy analysis, and machine-learning framework presented in this project. We extend our appreciation to all peers and contributors who supported various stages of this research, from dataset preparation and validation to cross-checking calculations and reviewing intermediate drafts. Their cooperation, feedback, and enthusiasm strengthened the quality of this work. Finally, the authors acknowledge the broader research community whose existing studies on wind energy, VAWT design, environmental modeling, and data-driven analysis provided invaluable foundations for this project.*

References

Blocken, B. 2016. “50 years of Computational Wind Engineering: Past, Present and Future.” *Journal of Wind Engineering and Industrial Aerodynamics* **152**: 1–10.

Chen, Z., Y. Liu, and H. Wang. 2022. “Wind Resource Modeling Using Deep Neural Networks.” *Applied Energy* **305**: 117879.

Eriksson, S., H. Bernhoff, and M. Leijon. 2008. “Evaluation of Different Turbine Concepts for Wind Power.” *Renewable and Sustainable Energy Reviews* **12**, no. 5: 1419–1434.

Gaden, D., and E. Bibeau. 2014. “Structural Loading and Fatigue Analysis in Vertical Axis Wind Turbines.” *Journal of Wind Engineering and Industrial Aerodynamics* **132**: 33–41.

Gal, Y., and Z. Ghahramani. 2016. “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning.” In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 1050–1059.

IEA Wind Technology Collaboration Programme. 2021. *Wind Energy in India: Annual Report 2021*. IEA Wind TCP Publications.

Jain, R., and A. Gupta. 2019. “Performance Evaluation of Helical-Blade Vertical Axis Wind Turbines in Low-Wind Environments.” *Renewable Energy* **139**: 1275–1287.

Kumar, G., A. Singh, and P. Ranjan. 2019. “Comparative Aerodynamic Analysis of Straight-Bladed H-Rotor Turbines.” *Energy Conversion and Management* **199**: 111946.

Li, Y., and L. Shi. 2010. “Assessment of Wind Power Potential Using Computational Models.” *IEEE Transactions on Power Systems* **25**, no. 3: 1393–1401.

Liu, W., X. Xiao, and L. Cheng. 2018. “Turbulence-Induced Fatigue in Small Wind Turbines.” *Journal of Wind Engineering and Industrial Aerodynamics* **175**: 24–35.

Mahajan, R., and M. Deshmukh. 2020. “Performance Enhancement Techniques for Savonius Turbines.” *Renewable and Sustainable Energy Reviews* **134**: 110291.

McLennan, J. 2014. “Urban Wind Turbines for Building-Integrated Renewable Systems.” *BuildingGreen Journal* **32**, no. 4: 45–52.

Monteiro, C., L. Ramirez, J. Garcia, and M. Fernandez. 2013. “Wind Speed Forecasting Using Ensemble Learning.” *Renewable Energy* **50**: 270–276.

National Institute of Wind Energy (NIWE). 2020. “Indian Wind Atlas (Third Edition).” NIWE Technical Report.

Ouyang, T., X. Zhou, L. Ye, and F. Meng. 2019. “Machine Learning Models for Wind Power Curve Forecasting.” *Applied Energy* **235**: 1547–1560.

Peng, C., Y. Chen, and L. Fu. 2018. “Experimental Validation of Small VAWTs in Turbulent Flows.” *IET Renewable Power Generation* **12**, no. 14: 1650–1658.

Singh, S. K., A. Patel, and J. Rao. 2021. “Wind Resource Assessment for Emerging Indian Smart Cities.” *Wind Engineering* **45**, no. 6: 745–758.

Stathopoulos, T. 2008. “Wind Turbulence Effects in Dense Urban Environments.” *Wind and Structures* **11**, no. 2: 121–134.

Swierczynski, M., T. Kere, F. Balduzzi, and R. Friedmann. 2013. “Aerodynamic and Structural Analysis of Darrieus Rotors.” *Renewable Energy* **59**: 166–174.

Vortex FdC. 2017. “High-Resolution Wind Map Dataset for India.” Vortex Forecasting Data Corporation.

Yoshida, S. 2019. “High-Resolution CFD Analysis of Urban Wind Flow.” *Renewable Energy* **143**: 1672–1684.

Journal of Big Data and Artificial Intelligence

WWW.JBDAI.ORG

ISSN: 2692-7977

JBDAI Vol. 4, No. 1, 2026 DOI: 10.54116/jbdai.v4i1.83

INFUSING ARTIFICIAL INTELLIGENCE INTO STRATEGY: SYNTHESIZING FIVE CLASSIC DEBATES

Rajiv Kashyap
William Paterson University
kashyapr@wpunj.edu

Jim Samuel
Rutgers University
AIXosphere
jim.samuel@rutgers.edu

Ashley Lee
William Paterson University
lees86@wpunj.edu

Raza Mir
Rutgers University
Mirr@wpunj.edu

ABSTRACT

This paper explores how artificial intelligence (AI) transforms the foundations of strategic management theory. While traditional debates have centered on industry structure and resource-based perspectives, AI introduces a theoretical discontinuity that challenges assumptions about cognition, resources, and firm boundaries. We examine five influential streams: Behavioral Strategy, Microfoundations, Ecosystems and Platforms, Stakeholder Resource-Based View, and Strategy-as-Practice, to assess how AI reshapes their core premises. Our analysis reveals that AI creates hybrid cognitive architectures, embeds algorithmic actors into microfoundations, reconfigures ecosystems around foundation models, redistributes resource control to stakeholders, and alters strategizing practices through continuous, AI-augmented processes. The paper concludes with an agenda for empirical research, emphasizing multi-level analysis, algorithmic governance, and ethical considerations in an AI-infused strategic landscape.

Keywords *artificial intelligence*, *strategic management theory*, *behavioral strategy*, *microfoundations*, *digital ecosystems*, *human-AI collaboration*, *algorithmic governance*.

1. Introduction

Strategic management has always been centrally concerned with explaining persistent differences in firm performance. From the earliest works in the field, scholars have attempted to answer a deceptively simple question: Why do some firms succeed while others fail?

Two dominant perspectives, the external industry configuration view and the resource-centric perspective (resource-based view, RBV), have provided the field's foundational answers. The external industry configuration perspective, anchored in industrial organization economics (Porter 1980), emphasizes how competitive positioning within an industry and the structure of the external environment drive sustainable rents. By contrast, the RBV (Wernerfelt 1984; Barney 1991) argues that firm-level heterogeneity rooted in valuable, rare, inimitable, and non-substitutable resources is the essential source of strategic edge.

Competitive advantage (Porter 1980; Barney 1991; Wernerfelt 1984) and bounded rationality frameworks (Simon 1947; Cyert and March 1963; Nelson and Winter 1982) provided the initial scaffolding for this research, while subsequent developments of the knowledge-based view (Grant 1996; Nonaka 1994) and stakeholder theory (Freeman 1984) expanded the field's perspective on value creation. Over the last three decades, scholars have devoted considerable energy to synthesizing these perspectives, whether by integrating industry-level and firm-specific mechanisms, proposing multilevel theories of capabilities, or reconciling structure and agency in the genesis of advantage. This interpretation is extended by incorporating contemporary artificial intelligence (AI)-driven dynamics, offering a novel synthesis beyond classical frameworks.

Yet the contemporary emergence of AI, as both a technological paradigm and an organizational force, poses challenges that extend far beyond traditional attempts at synthesis. AI undermines long-standing assumptions about managerial rationality, the nature of organizational resources, and even the unit of analysis in strategy theory. Machine learning systems redefine what constitutes human and nonhuman agency. Generative AI systems alter the positioning, domain-specific variances, transparency and cost structure of information processing, representation, and recombination (Bhuyan et al. 2025; Tripathi et al. 2025; Chidipothu et al. 2025). Foundation models (FMs) transcend firm boundaries, embedding themselves into ecosystems, supply chains, and institutional infrastructures. They also reshape the dynamics between firms and their stakeholders, altering power, participation, and co-creation. In short, AI constitutes a discontinuity with implications not only for strategy practice but for strategy theory itself.

Despite rapid practitioner interest in AI's strategic implications, academic strategy research has only begun to confront the theoretical consequences. Emerging work in strategy research highlights the role of AI in reshaping competition (Cockburn et al. 2018), augmenting managerial cognition (Shrestha et al. 2019), transforming organizational capabilities (Raisch and Krakowski 2021), restructuring ecosystems (Hannah and Eisenhardt 2018), and accelerating learning processes (Wu and Wang 2021). These contributions provide important starting points, yet the field lacks a systematic account of how AI reverberates across strategy's key theoretical directions and what a synthetic, AI-informed strategy theory might entail.

This paper seeks to fill that gap by examining five influential theoretical trends in contemporary strategy research and evaluating how AI reshapes each:

1. *Behavioral Strategy:* How does AI alter assumptions about cognition, limited rationality, emotions, heuristics, and social influence in strategic decision-making?
2. *Microfoundations:* How does AI challenge existing models of individual-level heterogeneity, interaction patterns, and the human foundations of firm-level outcomes?
3. *Ecosystems and Platforms:* How does AI reconfigure interconnected network architectures, multiactor complementarities, and the governance of distributed innovation?
4. *Stakeholder RBV:* How does AI transform resource co-creation with stakeholders, introduce new forms of value appropriation, and shift power across stakeholder groups?
5. *Strategy-as-Practice (SAP):* How does AI reshape strategizing as lived practice, i.e., the tools, routines, interactions, and situated actions through which strategy is performed?

These five areas are not exhaustive; many additional streams, dynamic capabilities, evolutionary economics, real options, and knowledge-based theories, are likewise affected by AI. However, these five represent major contemporary directions in mainstream strategy research, and together they offer a multifaceted foundation for exploring how AI destabilizes core theoretical assumptions.

1.1. AI as a Transformational Theoretical Shock

Much like the rise of digital platforms in the early 2000s, AI functions as what theorists might call a general-purpose technology, but one whose generality and adaptivity are unprecedented. Machine learning algorithms are not simply tools; they are evolving systems capable of learning patterns, generating outputs, and increasingly acting with a degree of autonomy (Jordan and Mitchell 2015). This "agentic" dimension of AI, its capacity to make recommendations, initiate actions, and influence decision flows, creates theoretical complexity that extends beyond automation or digitization. From a strategy theory standpoint, AI simultaneously affects:

- the cognition of managers (what they perceive, ignore, and prioritize)
- the resources firms control (data, models, and algorithmic capabilities)
- the structure of competition (winner-take-all dynamics, scalability of adaptive learning systems)
- the nature of ecosystems (interdependence among models, data flows, and platforms)
- stakeholder dynamics (distributional consequences, new power asymmetries)
- strategic practice itself (expansively fluid intelligence, tools, routines, decision processes, and the locus of agency)

Thus, AI forces a reconsideration of strategy's foundational ontology. It forces us to ask fundamental questions: What is a resource? Who or what is a strategist? What is the boundary of a firm? How do capabilities emerge when "actors" include both humans and intelligent systems? What does strategic edge mean when adaptive intelligence, not static resources, becomes the basis of primary value creation?

These ontological questions echo themes in the field's recent interest in behavioral realism, multilevel theorizing, and dynamic, process-oriented views of strategic edge. Yet AI pushes each of these themes into new territory, requiring both conceptual innovation and integrative synthesis.

1.2. Toward an AI-Infused Strategy Theory

Understanding AI-infused strategy requires moving beyond purely technological or economic framings to embrace sociotechnical perspectives. Actor-network theory provides a valuable lens for understanding how algorithmic actors reshape organizational networks (Latour 2005; Callon 1984). The concept of sociomateriality helps us theorize the inseparability of human and algorithmic agencies in strategic practice (Orlikowski 2007; Orlikowski and Scott 2008; Leonardi 2011). These perspectives reveal how AI transforms not just decision outcomes but the very fabric of organizational life. Moreover, institutional theory reminds us that AI adoption reflects not only efficiency calculations but also legitimacy pressures and mimetic isomorphism (DiMaggio and Powell 1983).

AI has been broadly defined as a cluster of "technologies that mimic the functions and expressions of human intelligence, specifically cognition, logic, learning, adaptivity and creativity" (Samuel 2021; Samuel et al. 2022, 2024). While this ontologically grounded definition captures AI's essential characteristics from a contemporary perspective, it remains insufficient for advancing strategy research and practice. Strategic management requires not merely an understanding of what AI *is*, but rather a conceptualization of what AI *does*—how it transforms the mechanisms through which firms create and capture value (Dwivedi et al. 2021). Following our analysis of AI's influence on five major strategy theories, we present a contextualized articulation of AI that emphasizes its functional impacts on competitive dynamics, organizational capabilities, and the fundamental processes of strategizing itself.

One of this paper's central arguments is that AI is not simply an empirical phenomenon requiring new studies, but a transformative systemic shock that necessitates a re-evaluation of long-standing strategy theories (Galaz et al. 2021). However, rather than discarding the field's existing insights, AI may provide an integrative glue that helps synthesize the external industry configuration and resource-based traditions. AI can also act as a disruptor and replacer, which affects both external and internal sources of advantage (Wan and Zhao 2024). It alters industry-level dynamics (data network effects, digital hub competition) while also reshaping firm-specific resources and capabilities (proprietary data, model fine-tuning routines, organizational learning loops).

Therefore, AI may enable a new synthesis in strategy theory, one that reconciles external and internal explanations by focusing on learning processes, cross-boundary data flows, and multiactor complementarities as fundamental drivers of advantage. The five theoretical trends examined in this paper each illuminate a different facet of this synthesis:

1. Behavioral strategy reveals how AI changes the psychology and cognition of strategizing.
2. Microfoundations explain how AI alters the building blocks of capabilities and interactions.
3. Ecosystem theories show how AI restructures the locus of competition.
4. Stakeholder RBV highlights distributed value creation and ethical-political consequences.
5. SAP exposes how AI transforms strategizing as an enacted, material practice.

By analyzing how AI transforms these areas and synthesizing them into a comprehensive framework, this paper aims to articulate a forward-looking theory of agentic AI and strategic management.

1.3. Structure of the Paper

The remainder of the paper is organized as follows. Section 2 reviews five major theoretical trends influential in contemporary strategy research. Section 3 outlines how AI intersects with and challenges existing models in each area, drawing extensively on recent SMJ contributions. Section 4 presents a comparative table contrasting pre-AI and post-AI perspectives on the determinants of strategic edge across the five theoretical domains. Section 5 concludes with implications for theory, methods, and practice.

This paper conceptualizes AI not as a technological add-on but as a transformative force that necessitates a fundamental shift in the assumptions underpinning strategy theory. As AI becomes increasingly embedded in organizations, ecosystems, and sociotechnical infrastructures, strategic management research must develop a theoretical vocabulary that addresses this new reality. The goal of this paper is to take a step toward such a vocabulary.

2. Five Influential Trends in Strategy Research

The emergence of FMs represents a fundamental shift in AI capabilities, with profound implications for organizational strategy (Bommasani et al. 2021; Xu et al. 2025). These models demonstrate general-purpose learning abilities that challenge traditional assumptions about technological specialization and competitive advantage. Recent research suggests AI may fundamentally reshape economic growth trajectories (Aghion et al. 2019), labor markets (Bessen et al. 2025), and the nature of human-machine collaboration (Passerini et al. 2025). The competitive dynamics among FM providers introduce new questions about market structure and governance (Haim and Meyer 2024), while the algorithmic mediation of organizational relationships creates novel power structures (Gutiérrez 2024). These developments demand new forms of digital leadership attuned to AI-specific capabilities (Hossain et al. 2025).

Over the last three decades, strategic management has evolved from a field primarily concerned with competitive positioning and firm-level resources to one that increasingly incorporates insights from behavioral science, organizational microdynamics, stakeholder engagement, practice-based perspectives (Guerras-Martin et al. 2014), and ecosystems theory (Jacobides et al. 2018),. These developments expand the theoretical boundary of strategy while also sharpening analytical tools for understanding how organizations create and appropriate value. In this section, we review five major theoretical trends that have become central to strategy research and that form the backbone of our analysis of how AI transforms theoretical assumptions.

The five areas, behavioral strategy, microfoundations, ecosystems, stakeholder RBV, and SAP, were selected because they have shaped significant streams of inquiry in the field and because they capture different layers of analysis: cognition and behavior, individual-level interactions, multiactor systems, stakeholder relationships, and the situated practices of strategists. Together, these streams illuminate the diverse mechanisms through which strategic edge develops and evolves.

2.1. Behavioral Strategy and the Psychology of Strategizing

The behavioral turn in strategic management emerged as a response to critiques of the rational-actor model inherited from classical economics. Early behavioral insights from Simon (1947), Cyert and March (1963), and others emphasized limited rationality, satisficing, routines, and the sociopolitical nature of organizational decision-making. Yet these ideas were not fully integrated into strategy until the 2000s, when scholars like Gavetti (2005), Levinthal (2011), and Powell et al. (2011) mobilized cognitive and psychological insights to explain strategic behavior more realistically.

Behavioral strategy focuses on several core dimensions. The first of these are cognition and mental models. Strategic decision-makers rely on cognitive schemas, analogies, frames, and heuristics to interpret environments. These mental models shape what firms perceive, how they attend to information, and how they make sense of competitive contexts. Research by Gary and Wood (2011), Eggers and Kaplan (2013), and Kaplan (2011) demonstrates how managerial cognition affects strategic choices under ambiguity, complexity, and competitive pressure.

In addition, behavioral strategy focuses on biases and heuristics. Strategic choices are systematically shaped by psychological biases, overconfidence, loss aversion, anchoring, escalation of commitment, and others (Powell et al. 2011). Although biases can create performance traps, heuristics sometimes facilitate effective strategizing by simplifying complexity.

Shifting from a purely economic approach, behavioral strategy also takes emotion and social influence into consideration. Recent extensions acknowledge affect, identity, status, and social dynamics as important determinants of strategic behavior (McDonald and Westphal 2013). Emotions influence risk-taking, conflict resolution, alliance decisions, and entrepreneurial judgment.

The importance of behavioral realism in strategy theory is also foregrounded. The core argument of behavioral strategy is that realistic assumptions about human cognition and emotion improve explanatory power (Gavetti 2012). Instead of assuming perfect rationality, behavioral strategy assumes bounded, structured, and socially embedded rationality.

This perspective has become essential for understanding how decision-makers formulate strategy, interpret competitive signals, and navigate uncertainty, making it a crucial foundation for analyzing how AI alters managerial cognition and strategic judgment.

2.2. Microfoundations of Strategy

While behavioral strategy emphasizes the cognitive and psychological processes of individuals, the microfoundations movement addresses a broader question: How do individual-level actions and interactions generate firm-level outcomes?

The microfoundational agenda emerged around 2010 as scholars grew increasingly dissatisfied with “black box” explanations of capabilities, routines, and performance heterogeneity. Research by Felin and Foss (2005), Felin et al. (2015), and Barney and Felin (2013) emphasized the need for theories that explicitly articulate how individuals, through their attributes, behaviors, and interactions, create and modify organizational-level constructs. This interpretation is extended by incorporating contemporary AI-driven dynamics, offering a novel synthesis beyond classical frameworks.

Three themes define microfoundational work. The first is *individual heterogeneity*. Microfoundation theorists believe that capabilities and routines originate from the skills, motivations, knowledge, and cognitive differences among individuals (Felin and Foss 2005). Even highly abstract constructs such as adaptive capabilities or absorptive capacity must ultimately be grounded in individuals’ abilities to sense, interpret, and act.

The second theme involves *interactions and coordination mechanisms*. Individuals rarely act in isolation; organizational outcomes arise through structured interactions supported by coordination mechanisms, relational contracts, communication patterns, and shared understandings. Microfoundations emphasize how patterns of interaction, such as communication networks or cross-functional teaming, produce emergent outcomes (Ployhart and Moliterno 2011).

Finally, microfoundations address *aggregation and emergence*. A key contribution of microfoundations is to clarify how individual actions accumulate into organizational phenomena. This involves understanding emergence: firm-level capabilities result not from simplistic summation but from complex, interdependent interactions. Organizational routines, for example, emerge from repeated interaction patterns but acquire stability through collective enactment.

The microfoundational lens is valuable for studying AI because AI alters who (or what) performs actions, how individuals interact with systems, and how firm-level outcomes emerge from human-AI hybrids. The shift from purely human microfoundations to sociotechnical microfoundations challenges the boundary conditions of the field.

2.3. Ecosystems, Platforms, and Meta-Organizations

Over the past decade, ecosystems and platforms have become defining concepts in strategy research. Traditional strategy theories conceptualized competition as occurring between firms within industries or markets. Interconnected network thinking expands this by focusing on loosely coupled, multiactor systems in which value creation depends on complementarities among firms, users, regulators, and institutional actors. Ecosystem theory has evolved from early work on innovation ecosystems (Adner 2006; Autio and Thomas 2014) to more nuanced understandings of platform-based value creation (Gawer 2014; Gawer and Cusumano 2014). The modular architecture of digital platforms (Baldwin and Clark 2000) enables new forms of ecosystem orchestration. The emergence of AI platforms introduces additional complexity in how value is created and captured across eco- system participants (Jacobides et al. 2019).

Three distinctive ideas underpin ecosystems research, The first relates to the *primacy of complementarities*. Firms create value not in isolation but through complementarities, among modules, technologies, and actors. The success of a digital hub depends on the quality and diversity of complementary innovations (Hannah and Eisenhardt 2018; Kapoor and Agarwal 2017).

The second theme involves *distributed and interdependent innovation*. In ecosystems, innovation is distributed across multiple firms and actors. No single firm controls all critical resources (Adner 2017). Meta-organizations, organizational forms where membership is composed of other organizations, shape coordination, governance, and conflict resolution.

The third idea is that of *winner-take-most dynamics and governance* (Schrepel and Pentland 2025). Ecosystems and platforms often exhibit network effects, scale economies, and data-driven learning loops that produce dominant positions for a few firms. This creates new governance challenges, including power asymmetries, access control, and rule-setting authority. Interconnected network theory provides a compelling framework for understanding AI because AI development, deployment, and improvement occur in highly interdependent environments (Shur-Ofry 2024). FMs rely on data ecosystems, model hubs, cloud infrastructures, and communities of developers. Competitive dynamics increasingly unfold not only between firms but between ecosystems of algorithms, data flows, and complementarities (Schrepel and Pentland 2025).

2.4. Stakeholder RBV

Classic RBV focuses on firm-controlled resources as sources of advantage. Stakeholder RBV represents a major shift, arguing that participating entities, not just shareholders, play central roles in resource creation, development, and appropriation (Barney et al. 2021). In a sense, it expands the boundaries where firm-level research can be enacted.

Some of the key premises of stakeholder RBV include:

Stakeholders as Resource Contributors: Stakeholders (employees, suppliers, customers, communities, regulators) contribute tangible and intangible resources: knowledge, legitimacy, relationships, trust, data, and social support. These resources shape a firm's ability to generate and sustain strategic edge (McGahan 2021).

Co-Creation and Co-Specialization: Value emerges through collaborative interactions among firms and stakeholders. Co-specialized assets, resources whose value depends on relational fit, create mutual dependence. The more stakeholders are integrated into capability development, the more they influence strategic outcomes (Harrison et al. 2010).

Stakeholder Power and Value Appropriation: Stakeholders vary in their power, claims, salience, and ability to capture value. Stakeholder RBV highlights distributional questions: who benefits from resource creation, and under what conditions? This perspective suggests that value creation is a fundamentally relational process, whereby the firm's competitive success depends on its ability to effectively manage and integrate the diverse interests of its stakeholder network (Freeman et al. 2021).

Ethical and Sociopolitical Considerations: Many strategic resources are socially constructed, making ethical and political issues such as fairness, equity, and transparency, central to understanding competitive advantage. This is particularly relevant for AI, where issues of data rights, privacy, surveillance, and algorithmic fairness introduce new stakeholder tensions. The potentially severe consequences of AI mismanagement, whether through intentional misuse or unintended harm, necessitate moving beyond discourse on "human-centered AI" toward frameworks that actively enhance human capabilities and wellbeing. Strategic advantage increasingly depends on organizations' ability to deploy AI in ways that demonstrably elevate human agency, augment rather than diminish human capacities, and create value that stakeholders experience as genuinely beneficial rather than extractive (Kashyap et al. 2024; Samuel et al. 2024). Firms that treat ethical considerations as constraints to be minimized risk stakeholder backlash and regulatory intervention, while those that embed human enhancement principles into AI strategy may gain legitimacy-based advantages and more durable stakeholder relationships.

Stakeholder RBV is essential to understanding how AI impacts strategy because AI depends on data derived from stakeholders, shapes stakeholder experiences, and redistributes value and risk across stakeholder groups (Ozdemir et al. 2023; Kashyap et al. 2024). The successful reinforcement of ecological sustainability at the ecosystem level often depends on the diversity of the stakeholder network, as collaborations with secondary stakeholders are specifically associated with the development of eco-innovations (Ozdemir et al. 2023).

2.5. SAP

SAP shifts attention from strategy as a property of firms to strategy as something people do, emphasizing everyday activities, tools, and routines through which strategizing is enacted. Users of SAP (Whittington 2006) argue that traditional perspectives obscure the micro-level practices that produce strategy. Current approaches to SAP focus on emphasizing dynamism in strategic theory and practice (Prashantham and Healey 2022).

SAP focuses on three interrelated dimensions:

Practitioners: It is important to ask who we refer to when we talk about strategists. SAP researchers believe that beyond top managers, strategists include middle managers, consultants, analysts, and increasingly, technological artifacts and systems.

Practices: What tools, methods, and techniques do strategists use? Practices include forecasting tools, decision frameworks, performance dashboards, analytical techniques, and planning routines.

Praxis: What do strategists actually do in day-to-day settings? Praxis includes meetings, presentations, planning cycles, negotiations, sensemaking, and interactions.

A distinguishing feature of SAP is its materiality: strategizing is mediated by artifacts, spreadsheets, slide decks, dashboards, analytics tools. This makes SAP particularly relevant for understanding how AI transforms strategizing because AI creates new tools, embeds itself in existing practices, and acts as a quasi-strategist by participating in decision processes.

2.6. Why These Five Trends Matter for AI

These five trends represent distinct but complementary theoretical developments in strategy. They jointly provide:

- A behavioral lens on cognition and decision-making
- A micro-level lens on individuals and interactions
- A system-level lens on ecosystems and platforms
- A stakeholder lens on value creation and claims
- A practice-based lens on strategizing-in-action

Together, they enable a multilayered understanding of strategic management, one that is well-suited to capturing the profound changes introduced by AI. Each trend reveals a different way in which AI destabilizes the assumptions of traditional strategy theories. In subsequent sections, we examine how AI intersects with these areas and how foundational constructs, such as strategic edge, firm boundaries, and the nature of resources, must be reconceptualized in light of AI.

3. How AI Intersects with and Challenges Strategy Theory

AI is not simply a new technology. It is a general-purpose, adaptive, and increasingly agentic system that fundamentally reshapes how firms sense opportunities, make strategic decisions, configure resources, coordinate activities across organizational boundaries, structure network relationships, and ultimately innovate. Unlike prior technologies that automated predefined tasks, AI systems learn, adapt, understand queries and users better, and generate novel solutions, transforming not only the efficiency of strategic processes but the fundamental nature of strategy (Samuel et al. 2022, 2026).

While earlier digital technologies influenced communication, transaction costs, or knowledge flows, AI transforms core strategic mechanisms: sensing, learning, prediction, recombination, attention allocation, and resource deployment. In this section, we show how AI reshapes, extends, and in some cases destabilizes the assumptions embedded in the five theoretical areas outlined earlier. To ground the discussion, we draw upon relevant strategy research, where scholars have begun exploring AI's strategic implications.

3.1. AI and Behavioral Strategy

Behavioral strategy rests on psychological realism, cognition, heuristics, biases, and limited rationality. AI interacts with every part of this foundation, reshaping the informational environment, altering cognitive processes, and introducing a new quasi-agent with its own distinct characteristics.

Below, we highlight four ways in which our theoretical understanding of behavioral strategy is affected by the emergence of strategy in firm practices.

AI alters attention structures and cognitive frames: AI-based systems change how managers perceive and prioritize issues by filtering information (recommender systems, predictive dashboards), by highlighting patterns not readily identifiable by humans, and by structuring interpretive frames through model outputs, confidence scores, or anomaly alerts. Shrestha et al. (2019) show how AI reshapes managerial cognition by augmenting sensemaking processes. AI generates synthetic representations of environments, probability distributions, risk scores, clusters, that serve as cognitive scaffolds, influencing what managers attend to.

AI may attenuate certain biases but amplify others: AI can reduce some biases (anchoring or availability heuristics) by offering more stable baselines or by countering human overconfidence with statistical calibration. Conversely, AI may amplify confirmation bias as humans selectively attend to algorithmic outputs that align with their beliefs. Strategy research on algorithmic decision support (Ransbotham et al. 2021) warns that humans may overweight algorithmic authority while misunderstanding model error. This raises an important theoretical question: Does AI reduce limited rationality or merely shift its locus?

AI introduces algorithmic biases and model persuasion: AI systems exhibit their own forms of bias, stemming from data sampling, model architectures, or training objectives. These algorithmic biases can shape strategic decisions in ways that differ from human psychological biases but are nonetheless consequential (Mehrabi et al. 2021). AI may also exhibit persuasive power: the capacity to steer attention, frame options, or nudge decision-makers. As high-stakes strategic decisions become increasingly mediated by AI, behavioral strategy must incorporate human—algorithm joint cognition.

AI changes the emotional landscape of strategizing: Managers may experience anxiety, distrust, or overconfidence in response to AI systems. Emotions influence how they incorporate model outputs and whether they accept or reject AI recommendations. This emotional interface between humans and technology becomes a new determinant of strategic behavior.

3.2. AI and Microfoundations

Microfoundations explain how individual actions produce firm-level outcomes. AI challenges these assumptions by altering the identity of actors, the nature of interactions, and the mechanisms of aggregation and emergence. Below, we highlight three key ways in which the emergence of AI affects our understanding of the microfoundations of strategy.

AI as a strategic actor: Traditional microfoundations assume that individuals are human, cognitively bounded, embedded in social structures, and motivated by incentives, beliefs, and identities. AI violates multiple assumptions simultaneously. It is non human, computationally powerful, trained on large-scale data rather than experiential learning, and motivated (in a sense) only by optimization objectives (Garvey et al. 2021). These represent fundamental shifts in our understanding of human-non-human interactions within firms. AI becomes a new class of microfoundational actor, neither human nor tool but a hybrid entity performing tasks normally attributed to individuals (forecasting, design, analysis, problem-solving). This challenges the anthropocentric foundations of strategy theory.

AI alters interaction patterns and coordination mechanisms: Human-AI hybrids create new interaction structures. These include delegation chains in which AI performs tasks that humans previously coordinated, predictive coordination, where agents synchronize based on algorithmic forecasts, and hidden interdependencies, because AI models may embed correlations invisible to humans. Strategy research on digital coordination (von Krogh et al. 2023) provides early insights into how algorithmic systems reshape organizational interactions.

Aggregation dynamics change in human-AI systems: Capabilities arise not simply from human actions but from human-AI co-evolution. Learning loops evolve differently. On one hand, model updates may alter individual behavior, while on the other, individuals fine-tuning models may alter model outputs. This collective behavior feeds new data back into the system. In other words, AI introduces reflexive microfoundations, where emergent outcomes depend on recursive interactions between people and algorithms.

3.3. AI, Ecosystems, and Platforms

AI challenges interconnected network theory at its core. It changes the architecture of complementarities, the nature of multiactor coordination, and the basis of strategic edge. Below, we highlight four ways in which the world of ecosystems and platforms is fundamentally altered by the emergence of AI.

AI increases interconnected network interdependence: Training data, FMs, cloud infrastructures, and application layers form deeply interdependent ecosystems. No single firm controls all necessary components. This aligns with strategy research on interconnected network complexity (Hannah and Eisenhardt 2018), but AI increases interdependence by embedding models into cloud compute layers, data pipelines, development frameworks, and API-based service layers. Such interdependence amplifies both cooperation (shared learning) and competition (model differentiation).

AI contributes to the emergence of FMs: FMs (GPT-like architectures) serve as digital hub cores around which complements, including domain-specific models, datasets, applications, and agentic workflows, develop. Research on digital hub architectures (Kapoor and Agarwal 2017) suggests that core modules determine system evolution. In AI ecosystems, the core becomes updatable, fine-tunable, and partly opaque. This raises new strategic questions: Who owns the model? Who controls training data? Who captures value from downstream complements?

AI reinforces winner-take-most dynamics: AI-driven ecosystems benefit from data network effects, algorithmic feedback loops, scale economies in training, and deployment breadth effects (model quality improves with diverse use cases). Cockburn et al. (2018) show that firms leading in AI enjoy durable performance advantages. These dynamics intensify with FMs, which become increasingly difficult for late entrants to match.

Network governance becomes algorithmic: Governance is increasingly mediated by factors such as model access controls, API pricing, safety constraints, fine-tuning permissions, and bias mitigation protocols (Ustahaliloğlu 2025). Interconnected network governance thus becomes model governance, requiring strategy theorists to integrate insights from computer science, ethics, and institutional theory (Rahwan et al. 2019).

3.4. AI and Stakeholder RBV

Stakeholder RBV emphasizes co-created resources and interdependent value. In that respect, it forced us to consider the environment in which organizations operate as the domain of their praxis. AI further reshapes this understanding of the role of stakeholders in a firm. Below, we highlight four ways in which it does so:

Stakeholders generate the most critical resource: Data are the foundation of AI. Yet data originate from stakeholders. Customers generate usage data. Communities generate content data, employees generate operational data, suppliers generate transactional data, regulators generate compliance data, and so forth (Jones and Tonetti 2020). Thus, stakeholder participation becomes integral to AI-related strategic edge. The RBV's traditional assumption that firms own their resources is contradicted by AI, where data sources are shared, contested, or co-created.

Stakeholder power increases with data rights: Stakeholders gain power when they control consent, data access, data deletion rights, and the ability to restrict or condition data sharing. Regulatory bodies also become dominant stakeholders (Constantinides et al. 2018). GDPR, CCPA, and emerging AI-specific regulations redefine property rights over data, shifting power dynamics.

AI generates new stakeholder risks: Stakeholder RBV emphasizes fairness and value distribution. AI creates risks for each stakeholder group (Floridi et al. 2018). Customers face privacy loss and algorithmic discrimination. Workers face automation and deskilling. Communities face surveillance, mis/disinformation, finally, partners face dependency on digital hub-controlled models. The strategic management of stakeholder risks becomes a capability in itself.

Stakeholders participate in AI co-creation: Open-source communities, developers, and crowd workers contribute to training, validating, and fine-tuning models. Value creation becomes multisided, and value appropriation becomes deeply contested (Nambisan et al. 2017).

3.5. AI and SAP

SAP focuses on strategizing as enacted practice, what people do in real situations. AI reshapes the materiality, temporality, and agency of strategizing. Below, we highlight four ways in which it does so.

AI becomes a strategizing tool and companion: AI systems create new strategy tools, including predictive dashboards, text and image generation tools, simulation engines, automated scenario generators, and workflow-oriented agentic systems. Strategizing becomes AI-augmented and, increasingly, AI-shaped.

AI alters the temporal structure of strategizing: Strategy once unfolded in cycles, annual planning, quarterly reviews. AI introduces continuous strategy, through real-time pattern recognition, always-on monitoring, continuous experimentation, and instantaneous simulation. This temporal shift breaks with long-established strategy routines.

AI as a "Quasi-Practitioner": SAP calls for attention to the practitioners of strategy. AI raises challenging questions: Is AI a strategist? Does it participate in sensemaking? Can it influence agenda-setting? Early theoretical commentary suggests that AI may become a quasi-practitioner, performing tasks traditionally reserved for analysts, consultants, or managers. This expands the ontological boundaries of who participates in strategizing.

AI reshapes meetings, routines, and interactions: Agentic AI tools transcribe meetings, summarize discussions, generate strategic options, and highlight inconsistencies. This changes interaction patterns, power dynamics, what counts as "expertise," the role of middle managers, and how strategists justify and legitimate decisions (Faraj et al. 2018). These changes affect not just outcomes but the practice of strategy itself.

3.6. Summary

AI intersects with strategy theory in profound ways:

- It alters cognition, attention, bias, and sensemaking.
- It changes microfoundations by introducing nonhuman actors.
- It transforms ecosystems into model-centered, data-driven structures.
- It redefines stakeholder resources and power.
- It reshapes strategizing as an enacted, material practice.

Each theoretical area highlights a different dimension of AI's transformative power.

In the next segment, we synthesize these insights into a comparative table contrasting pre-AI and post-AI determinants of strategic edge across the five domains.

4. Comparative Framework and Post-AI Transformations

To synthesize the theoretical developments described in earlier sections, this segment presents a comparative table contrasting pre-AI perspectives on the determinants of strategic edge with emerging post-AI perspectives across the five theoretical domains: behavioral strategy, microfoundations, ecosystems and platforms, stakeholder RBV, and SAP. In the following the table, we provide an in-depth analytical narrative explaining each transformation and its implications for strategy theory.

Table 1 provides a visual synthesis of the theoretical changes introduced by AI. Figure 1 depicts the impacts of AI on each theoretical domain, with the central role of AI as disruptor, integrator, and change agent driving the transformation from traditional to AI-infused strategic imperatives. Below, we expand on these transformations, explaining how each theoretical area must evolve to capture the mechanisms of advantage in AI-infused organizations and ecosystems.

4.1. Behavioral Strategy: From Human Limited Rationality to Hybrid Cognitive Architectures

Before AI became hegemonic across firms, competitive advantage from a behavioral standpoint hinged on the ability of managers to display heightened cognition, marshal their heuristic judgment, not fall prey to biases, and generally display better judgment at a human level. While these skills are not necessarily neutralized by AI, an additional dimension emerges, that of *human-AI joint cognition*. Attention and sensemaking take on an algorithmic character as well. Below, we highlight three key ways in which a firm's behavioral strategy is impacted by AI.

From cognitive advantage to hybrid cognitive advantage: Before AI, strategic edge was strongly tied to managerial cognition, how effectively leaders perceived opportunities, interpreted competitive signals, and avoided biases (Kaplan 2008). The behavioral part of strategic edge came from superior attention, superior heuristics, or superior sensemaking.

In the AI era, cognitive advantage is increasingly determined not only by human capabilities but by how firms integrate AI systems into cognitive processes (Jarrahi 2018). AI detects patterns too complex for humans, reallocates attention through alerts, flags, and prioritization, and reframes decision problems via simulations or model-generated alternatives. Strategizing becomes a human–algorithm collaboration, with algorithmic cognition supporting, shaping, or sometimes overriding human judgment (Faraj et al. 2018).

Algorithmic bias as a strategic variable: Where behavioral strategy once emphasized human biases, firms must now manage human biases, algorithmic biases, and biases emerging from human–algorithm interaction loops.

Table 1: Pre-AI and post-AI perspectives on determinants of competitive advantage.

Theory	Pre-AI Perspective	Post-AI Perspective
Behavioral Strategy	Advantage hinges on superior managerial cognition, effective heuristics, reduced biases, and high-quality executive judgment. Strategic outcomes depend on how managers interpret environments.	Advantage depends on *human–AI joint cognition*, algorithmic augmentation of attention and sensemaking, the governance of model biases, and firms' ability to integrate algorithmic and human judgment. Strategic outcomes reflect hybrid cognitive architectures.
Microfoundations	Firm-level capabilities and performance arise from individual-level skills, interactions, motivations, and emergent routines. All microfoundations are human.	Capabilities arise from *sociotechnical ensembles* of humans and AI systems. AI functions as a quasi-actor embedded in routines, altering interaction patterns, delegation structures, and emergent dynamics. Microfoundations include algorithmic agents.
Ecosystems and Platforms	Advantage emerges through complementarities, control of platform architectures, network effects, and superior governance of multiactor systems.	Advantage emerges from control of *AI foundation models*, data ecosystems, compute resources, and model governance. Ecosystems become model-centric; competition shifts to learning loops, fine-tuning networks, and AI capability diffusion.
Stakeholder RBV	Stakeholders contribute unique resources (knowledge, legitimacy, relationships), but firms primarily control and appropriate resources. Value creation occurs in firm-stakeholder relationships.	Stakeholders generate and control critical AI resources (data, feedback, content). Advantage depends on fair governance of data rights, ethical AI practices, stakeholder trust, and *co-creation* in model training and validation. Stakeholder power increases.
Strategy-as-Practice (SAP)	Strategy emerges through human practices: meetings, analyses, interpretations, routines, and tools shaping the everyday doing of strategy.	Strategy emerges through *AI-augmented practices*: automated analyses, AI-generated insights, simulations, and agentic workflows. AI becomes a quasi-practitioner influencing routines, interpretations, and power dynamics.

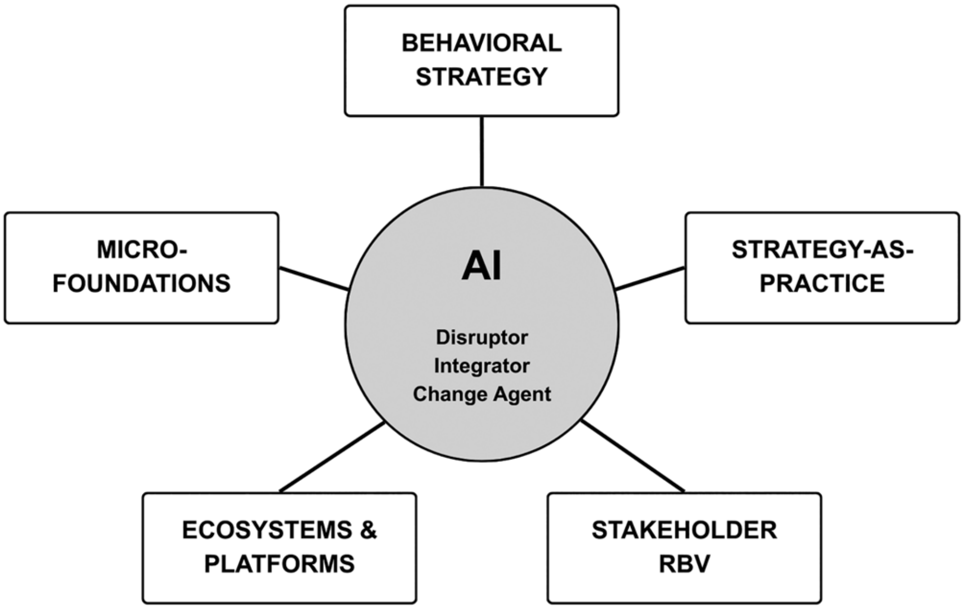


Figure 1: Strategy imperatives impacted by AI.

Strategic edge may depend on how well firms mitigate and harness these biases (Barocas and Selbst 2016; Jarrahi 2018).

Strategic emotion in human-AI interaction: Trust, anxiety, and overreliance on algorithms become emotional factors shaping strategy (Dietvorst et al. 2015). Organizations that manage emotional adaptation to AI (through training, transparency, or governance) outperform those that struggle with AI-induced discomfort or resistance (Glikson and Woolley 2020).

Table 2 synthesizes how AI transforms behavioral strategy across three dimensions. Human cognition evolves into hybrid cognition where algorithmic systems augment managerial sensemaking through pattern detection and cognitive scaffolds. Heuristics and biases expand to include managing both human and algorithmic biases, raising the question of whether AI reduces bounded rationality or merely shifts its locus. Limited rationality gives way to an emotional landscape where trust, anxiety, and adaptation to AI systems become strategic determinants.

These transformations reveal that cognitive advantage now depends on how effectively firms integrate algorithmic and human judgment while managing the emotional interface between humans and technology.

4.2. Microfoundations: From Human Capabilities to Sociotechnical Capabilities

As Table 1 suggests, microfoundations in the pre-AI era were predicated upon individual-level skills, interactions, motivations, and emergent routines. It was a *sine qua non* that all microfoundations were human. The terrain has shifted considerably in the post-AI era. Microfoundations now necessarily include algorithmic agents. Agentic AI gives rise to quasi-actors embedded in routines. They function to transform interaction patterns, disrupt delegation structures, and reconfigure emergent dynamics for better or for worse. Below, we highlight three ways in which the terrain of microfoundations has been transformed by AI.

AI as a microfoundational actor: Microfoundations theory must expand its ontology. Pre-AI microfoundations were explicitly human: individuals, their skills, their interactions. Post-AI, many traditionally human tasks, forecasting, classification, text generation, diagnosis, resource allocation, are performed by AI systems (Raisch and Krakowski 2021). This generates hybrid microfoundations, comprising humans, algorithms, interaction rules, data flows, and feedback loops. New "individual-level" heterogeneity emerges from differences in model architectures, training regimes, and fine-tuning (Felin et al. 2015).

Changing interaction patterns: Human-AI collaboration introduces new forms of coordination. These include algorithm-mediated coordination (AI recommends who should do what), predictive synchronization (AI predicts others' actions and informs teammates), and multiagent interactions (multiple AI agents interacting with humans). Capabilities emerge from interdependencies across human and algorithmic nodes.

Emergence in hybrid systems: Classic emergence involved repeated human interaction. Post-AI emergence depends on model updates, data drift, reinforcement learning feedback, and human adaptation to algorithmic cues. The locus of emergence shifts toward dynamic human–AI co-evolution (Murray et al. 2021; Shrestha et al. 2021).

AI necessitates attention to the transformation of microfoundations from purely human actors to sociotechnical capabilities. Individual capabilities must expand to include AI systems performing traditionally human tasks (forecasting, classification, text generation), generating hybrid microfoundations comprising humans, algorithms, and data flows (see Table 3).

Human interactions can be transformed into algorithm-mediated coordination where AIs can be deployed to recommend task allocation, enables predictive synchronization, and create multiagent interactions. Skill-based heterogeneity would yield to adaptivity through continuous reconfiguration (Felin et al. 2015), where competitive advantage stems from real-time adjustment rather than accumulated knowledge stocks (Teece 2018). These changes will fundamentally alter how firm-level outcomes emerge from individual-level actions and shift the locus of emergence toward dynamic human-AI co-evolution.

4.3. Ecosystems and Platforms: From Modular to Model-Centric

In the 21st century, platforms began to replace products as a source of competitive advantage in diversified firms, especially in the space of technology-enabled commerce. It was believed that competitive advantage emerged in these platforms through complementarities, control of platform architectures, network effects, and superior governance of multiactor systems. All those determinants of performance differentials remain salient, but a new layer emerges in the post-AI landscape. The new source of competitive advantage is AI FMs, data ecosystems (Iansiti and Lakhani 2020), compute resources, and model governance (Schrepel and Pentland 2025). The model becomes

Table 2: Behavioral strategy: Pre-AI to post-AI transformation.

Pre-AI Theoretical Dimension	Post-AI Theoretical Dimension(s)	Transformation Description	Key Mechanisms	Research Implications
Human Cognition Strategic edge tied to managerial cognition, how effectively leaders perceived opportunities, interpreted competitive signals, and avoided biases (Gavetti 2005; Gary and Wood 2011; Eggers and Kaplan 2013; Kaplan 2011).	**Hybrid Cognition** Cognitive advantage determined by how firms integrate AI systems into cognitive processes. AI detects patterns too complex for humans, reallocates attention, and reframes decision problems (Shrestha et al. 2019).	Strategizing becomes a human–algorithm collaboration, with algorithmic cognition supporting, shaping, or sometimes overriding human judgment.	• Recommender systems, predictive dashboards • Filtering information • Cognitive scaffolds • Alerts, flags, prioritization	Research should examine how AI augmentation affects managerial mental models, decision heuristics, and strategic frames (Riedl et al. 2024). How do firms integrate algorithmic and human judgment into cognitive processes?[*] What governance mechanisms manage emotional adaptation to AI?[*]
Heuristics and Biases Strategic choices shaped by psychological biases (overconfidence, loss aversion, anchoring, escalation of commitment). Heuristics sometimes facilitate strategizing by simplifying complexity (Powell et al. 2011).	**Algorithmic Biases** Firms must manage human biases, algorithmic biases, and biases emerging from human–algorithm interaction loops. AI may reduce some biases but amplify others (Ransbotham et al. 2017).	Strategic edge depends on how well firms mitigate and harness biases. Raises question: Does AI reduce limited rationality or merely shift its locus?	• Statistical calibration • Confirmation bias amplification • Algorithmic authority • Data sampling biases	Under what conditions does AI amplify versus mitigate cognitive biases? (Kahneman and Klein 2009) How do firms manage human biases, algorithmic biases, and interaction loop biases?[*]
Limited Rationality Behavioral strategy assumes bounded, structured, and socially embedded rationality rather than perfect rationality (Simon 1947; Cyert and March 1963).	**Emotional Landscape of Strategizing** Trust, anxiety, and overreliance on algorithms become emotional factors shaping strategy. Organizations that manage emotional adaptation to AI outperform those struggling with AI-induced discomfort (McDonald and Westphal 2013).	Emotional interface between humans and technology becomes a new determinant of strategic behavior through training, transparency, or governance.	• Manager anxiety, distrust • Overconfidence responses • Training, transparency • Governance mechanisms	How do managers calibrate trust in algorithmic versus human judgment? What organizational practices facilitate emotional adaptation to AI?[*]

[*] Theorized in this paper.

Table 3: Microfoundations: Pre-AI to post-AI transformation.

Pre-AI Theoretical Dimension	Post-AI Theoretical Dimension(s)	Transformation Description	Key Mechanisms	Research Implications
Individual Capabilities Microfoundations were explicitly human: individuals, their skills, their interactions. Capabilities and routines originate from skills, motivations, knowledge, and cognitive differences among individuals (Felin and Foss 2005; Felin et al. 2015; Barney and Felin 2013).	**Sociotechnical Capabilities** Many traditionally human tasks (forecasting, classification, text generation, diagnosis, resource allocation) are performed by AI systems. This generates hybrid microfoundations, comprising humans, algorithms, interaction rules, data flows, and feedback loops.	Microfoundations theory must expand its ontology. New individual-level heterogeneity emerges from differences in model architectures, training regimes, and fine-tuning quality.	• AI performing human tasks • Hybrid microfoundations • Model architectures • Training regimes, fine-tuning	Future research must theorize how agency is distributed across human-AI ensembles, drawing on actor-network theory (Latour 2005; Callon 1984), distributed cognition frameworks, and computational organizational theory. Under what conditions do algorithmic agents complement versus substitute for human judgment?[*] How do power dynamics shift when AI systems mediate strategic decisions?[*]
Human Interactions Organizational outcomes arise through structured human interactions supported by coordination mechanisms, relational contracts, communication patterns, and shared understandings (Felin et al. 2015).	**Human-AI Coordination** Human-AI collaboration introduces new forms of coordination: algorithm-mediated coordination, predictive synchronization, and multiagent interactions. Capabilities emerge from interdependencies across human and algorithmic nodes (von Krogh et al. 2023).	Agentic AI gives rise to quasi-actors embedded in routines. They function to transform interaction patterns, disrupt delegation structures, and reconfigure emergent dynamics.	• Algorithm-mediated coordination • Predictive synchronization • Multiagent interactions • Quasi-actors in routines	Future work should integrate insights from human-computer interaction (Sidji et al. 2024), cognitive science, and behavioral decision theory. What coordination mechanisms enable effective human-AI collaboration at scale?[*] What new governance forms emerge when responsibility is distributed between humans and machines?[*]
Skill-Based Heterogeneity Individual-level heterogeneity in skills creates firm-level capability differences. Competitive advantage stems from learning capability, the ability to accumulate, codify, and leverage knowledge over time (Teece et al. 1997; Eisenhardt and Martin 2000).	**Adaptivity through Continuous Reconfiguration** Advantage shifts toward adaptivity, the capacity to continuously reconfigure in response to real-time feedback. AI systems enable this through continuous sensing, pattern recognition, and adjustment rather than episodic learning cycles (Winter 2003; Zollo and Winter 2002).	Post-AI emergence depends on model updates, data drift, reinforcement learning feedback, and human adaptation to algorithmic cues. The locus of emergence shifts toward dynamic human-AI coevolution.	• Model updates • Data drift • Reinforcement learning feedback • Human-AI coevolution	What organizational structures enable continuous adaptivity rather than episodic learning cycles?[*] How do firms build adaptive capacity without sacrificing stability required for operational effectiveness?[*] What governance mechanisms ensure adaptive systems remain aligned with strategic intent?[*]

[*] Theorized in this paper.

salient to the ecosystem and AI capability diffusion becomes key. Below, we highlight four key elements of that transformation.

FMs become the new digital hub cores: Historically, platforms coordinated complementarities (apps, suppliers, modules). With AI, FMs become the core modules of ecosystems, generating capabilities across domains (language, vision, code), lowering the cost of complement creation (agents, apps, copilots), and serving as focal points for innovation communities (Bommasani et al. 2021; Bresnahan and Greenstein 1999). Innovation in FM-centric ecosystems emerges through two mechanisms: *emergent capability discovery*, where developers uncover unanticipated problem-solving abilities (such as cross-domain reasoning or novel task performance) that enable applications not envisioned during model development, and *architectural homogenization*, where standardized FM interfaces allow rapid deployment across diverse sectors with minimal adaptation (Bresnahan and Trajtenberg 1995; Baldwin and Clark 2000). For example, a breakthrough in legal document analysis can immediately inform medical diagnosis, customer service, or software development applications. Control over FMs thus becomes a primary source of interconnected network power, as these dual innovation mechanisms create network effects that reinforce the centrality of dominant models.

AI amplifies winner-take-most competition: FM ecosystems fundamentally reshape competitive dynamics by introducing self-reinforcing learning advantages. Model quality compounds over time through more data, larger user bases, more diverse fine-tuning tasks, and network effects from shared model improvements. This creates *learning-based strategic edges* that differ fundamentally from traditional network effects. While conventional network effects increase value through adoption, learning-based advantages improve the core product itself. Each interaction generates training data that enhance model performance, attracting more users and generating more data in a compounding cycle. Laggards face increasingly insurmountable barriers because catching up requires not just matching current capabilities but replicating the cumulative learning embedded in leaders' models (Agarwal et al. 2022). Competition thus shifts from contestable markets where late entrants can leapfrog incumbents, to cumulative-advantage markets where early leaders' learning curves create persistent performance gaps that challenge traditional strategy assumptions about competitive mobility.

Ecosystem governance becomes technical: Governance shifts from managing human complementarities to managing model access and API-based rules (Gillespie 2014) These include rate limits, safety constraints, fine-tuning permissions, model versioning, and compliance requirements. Firms with superior model governance architectures (transparent, safe, flexible) outperform on interconnected network health. Critically, governance decisions become *encoded in technical infrastructure* rather than negotiated through contracts or social norms. Access controls, usage policies, and quality standards are embedded directly into APIs, creating quasi-regulatory systems that shape complementor behavior automatically (Plantin et al. 2018). This technical embedding of governance rules accelerates enforcement but also raises new strategic questions about who designs these architectures, how power is distributed across governance layers, and whether complementors can effectively contest or exit from technically-mediated systems.

Multi-interconnected network intensifies interdependence: AI models depend on global data sources, interoperable tools, and cross-ecosystem infrastructures (Tilson et al. 2010; Yoo et al. 2010). Firms compete not just as stand-alone platforms but as interdependent ecosystems, where one ecosystem's model may rely on another's compute layer, while complementors cross-deploy applications across model ecosystems and industry boundaries blur around cross-model interoperability. The ecosystem becomes the key unit of analysis, yet traditional ecosystem theory assumes relatively stable boundaries and clear hierarchies. AI ecosystems exhibit *fluid boundaries* where participants simultaneously cooperate (sharing infrastructure, data standards, pre-trained models) and compete (differentiating on fine-tuning, application layers, user relationships), creating coopetition dynamics that are more complex and dynamic than platform-era competition (Klinger et al. 2020). Strategic advantage increasingly depends on orchestrating relationships across multiple, overlapping ecosystems rather than dominating a single, bounded platform.

In summary, AI will drive the transformation of ecosystems and platforms from modular architectures to model-centric structures (see Table 4).

FMs will replace traditional platform cores, generating capabilities across domains and lowering creation costs through emergent capabilities and architectural homogenization. Platform coordination will evolve into technically-mediated governance where API-based rules, access controls, and quality standards are embedded directly into the infrastructure, creating quasi-regulatory systems (Jacobides et al. 2019). We anticipate that traditional network effects will give way to learning-based strategic edges where model quality compounds through self-reinforcing data cycles, creating cumulative advantages and persistent performance gaps for early leaders. As a result, competition will shift from contestable markets to overlapping ecosystems where participants simultaneously cooperate and compete through shared infrastructure and cross-model interoperability.

Table 4: Ecosystems and platforms: Pre-AI to post-AI transformation.

Pre-AI Theoretical Dimension	Post-AI Theoretical Dimension(s)	Transformation Description	Key Mechanisms	Research Implications
Modular Complementarities Platforms coordinated complementarities (apps, suppliers, modules). Competitive advantage emerged through complementarities, control of platform architectures, network effects, and superior governance of multiactor systems (Hannah and Eisenhardt 2018; Kapoor and Agarwal 2017).	**Foundation Models as Digital Hub Cores** Foundation models become core modules of ecosystems, generating capabilities across domains (language, vision, code), lowering the cost of complement creation (agents, apps, copilots), and serving as focal points for innovation communities.	Control over foundation models becomes a primary source of interconnected network power. Innovation emerges through emergent capability discovery and architectural homogenization.	• Foundation models (FMs) • Emergent capability discovery • Architectural homogenization • Standardized FM interfaces	How do firms orchestrate multi-actor learning systems? (Hannah and Eisenhardt 2018) Future work should integrate ecosystem theory (Adner and Kapoor 2010) with research on open innovation (Chesbrough 2003), standards competition, and collective intelligence. What governance mechanisms align incentives when value creation depends on shared data and models?[*]
Platform Coordination Governance managed human complementarities through contracts and social norms. Platforms coordinated value creation across multiple firms with relatively stable boundaries and clear hierarchies.	**Model-Centric Ecosystems with Technical Governance** Governance shifts to managing model access and API-based rules. Governance decisions become encoded in technical infrastructure rather than negotiated. Access controls, usage policies, and quality standards are embedded directly into APIs (Xu et al. 2025).	AI ecosystems exhibit fluid boundaries where participants simultaneously cooperate (sharing infrastructure, data standards, pre-trained models) and compete (differentiating on fine-tuning, application layers, user relationships).	• API-based rules • Rate limits, safety constraints • Fine-tuning permissions • Quasi-regulatory systems	In AI ecosystems, each participant's interactions generate data that improves a shared foundation model, creating knowledge spillovers that benefit all members including competitors (Khan et al. 2024). How is power distributed across governance layers in technically-mediated systems?[*] Can complementors effectively contest or exit from technically-mediated systems?[*]

Table 4: **(continued)—Ecosystems and platforms: Pre-AI to post-AI transformation.**

Pre-AI Theoretical Dimension	Post-AI Theoretical Dimension(s)	Transformation Description	Key Mechanisms	Research Implications
Network Effects Traditional network effects increase value through adoption. Competition occurred in contestable markets where late entrants could leapfrog incumbents.	**Learning-Based Strategic Edges** Model quality compounds over time through more data, larger user bases, more diverse fine-tuning tasks, and network effects from shared model improvements. Learning-based advantages improve the core product itself. Each interaction generates training data that enhances model performance (Cockburn et al. 2018).	Competition shifts from contestable markets to cumulative-advantage markets where early leaders' learning curves create persistent performance gaps. Laggards face insurmountable barriers because catching up requires replicating cumulative learning embedded in leaders' models.	• Self-reinforcing learning • Training data enhancement • Compounding cycle • Cumulative learning barriers	How do firms compete in cumulative-advantage markets where early leaders' learning curves create persistent performance gaps?* What strategies enable laggards to overcome cumulative learning barriers?* How does strategic advantage depend on orchestrating relationships across multiple, overlapping ecosystems?*

---

*Theorized in this paper.

4.4. Stakeholder RBV: From Firm-Controlled Resources to Stakeholder-Governed Data and Trust

Stakeholders were always considered key to a firm's competitive advantage, but it was felt that they did so by contributing unique resources (knowledge, legitimacy, relationships). Despite their salience, it was believed that the primary firms controlled and appropriated resources (Hillman and Keim 2001). Things have transformed rather radically in this realm. In the post-AI landscape, stakeholders generate and control critical AI resources (data, feedback, content). The gains therefrom will have to be shared as their power increases. Moreover, competitive advantage is predicated upon the on fair sharing of data rights. Below, we highlight four themes relating to the transformation of firm-stakeholder relationships in the post-AI era.

Stakeholders own or control critical inputs: The most important AI resource, data, originates with stakeholders. This reverses the logic of resource control. Before AI became hegemonic, firms owned their resources (factories, IP, proprietary routines). In the post-AI world of firms, stakeholders generate the key resource (data), and firms must govern access and use responsibly (Zuboff 2019; Khatri and Brown 2010). This creates a new basis for advantage: trusted access to stakeholder-generated resources.

Data rights become strategic assets: Data protection laws, consent frameworks, and community norms shape what data firms can use, how they can use it, and whether stakeholders will cooperate with them over the short and long term. Firms with stronger reputations for fairness, privacy, transparency, and reciprocity gain a long-term resource advantage.

Stakeholder risks become strategic constraints: Stakeholders may resist AI adoption for a variety of reasons, including privacy fears, AI phobia, job displacement, and algorithmic unfairness (Samuel et al. 2025). Managing stakeholder risks is no longer just CSR, it becomes a capability that determines whether firms can sustain AI-driven advantage.

Stakeholders become co-developers of capabilities: In the AI era, it becomes imperative for firms to partner with their stakeholders through feedback, ratings, content creation, fine-tuning interactions, and crowdsourced testing. stakeholders help develop and refine AI capabilities. Firms cannot monopolize value; they must orchestrate cocreation (Prahalad and Ramaswamy 2004; Laursen and Salter 2006).

We predict that stakeholder RBV will evolve from firm-controlled resources to stakeholder-governed data ecosystems. The most critical AI resource, data, originates with stakeholders, reversing traditional resource control logic and creating new bases for advantage through trusted access (see Table 5).

Data rights will serve as strategic assets, and protection laws, consent frameworks, and community norms will shape what firms can use and whether stakeholders cooperate. As a result, resource creation will shift from firm-directed processes to co-creation opportunities where firms partner with stakeholders through feedback, ratings, content creation, and crowdsourced testing. Managing stakeholder risks around privacy, displacement, and algorithmic fairness will become a capability that determines whether firms can sustain AI-driven advantages.

4.5. SAP: From Human Strategizing to AI-Augmented Strategizing

As Table 1 suggests, our pre-AI concepts of strategizing primarily involved human practices. Transforming strategy work into a verb involved meetings, analyses, artifacts like documents and spreadsheets, and routines that highlighted the performative elements of strategy. Things are a bit different in the post-AI landscape. Today, the "verbing" of strategy includes AI-augmented practices. These include automated analyses, AI-generated insights, simulations, and agent powered workflows. AI becomes a quasi-practitioner influencing routines, interpretations, and power dynamics (Jarzabkowski and Kaplan 2015). Below, we highlight four ways in which SAP has been transformed by AI.

AI becomes a ubiquitous strategy tool: Strategizing becomes materially transformed, through real-time dashboards replacing periodic reports, generative AI producing draft strategic alternatives, simulation engines evaluating competitive moves, and agentic AI conducting environmental scanning automatically. These practices change not only outputs but how strategists act.

AI changes the rhythm of strategy work: Strategy becomes dynamic and continuous, requiring always-on monitoring, instant experimentation, automated scenario testing, and continuous learning loops (Faraj et al. 2018; Raisch and Krakowski 2021). This contrasts sharply with pre-AI episodic planning.

AI as a quasi-practitioner: AI joins the cast of strategists. It participates in meetings through summaries and insights, produces analytic content, flags risks and opportunities, and sometimes even initiates strategic options (Jarzabkowski and Kaplan 2015). This raises new questions about who "owns" strategic decisions, how credit or accountability should be assigned, and what becomes of expertise when AI can generate equivalent outputs.

Table 5: Stakeholder RBV: Pre-AI to post-AI transformation.

Pre-AI Theoretical Dimension	Post-AI Theoretical Dimension(s)	Transformation Description	Key Mechanisms	Research Implications
Firm-controlled Resources Stakeholders contributed unique resources (knowledge, legitimacy, relationships). Primary firms controlled and appropriated resources (factories, IP, proprietary routines).	**Stakeholder-Generated Data** The most important AI resource, data, originates with stakeholders. This reverses the logic of resource control. Stakeholders generate the key resource (data), and firms must govern access and use responsibly (Freeman 1984; Freeman et al. 2007).	Creates a new basis for advantage: trusted access to stakeholder-generated resources. Competitive advantage is predicated upon fair sharing of data rights.	• Stakeholder data generation • Reversed resource control logic • Responsible governance • Trusted access	How should firms balance value appropriation with fair treatment of stakeholders who provide training data?[*] Research should examine questions of algorithmic justice (Mittelstadt et al. 2016), data rights, and stakeholder voice in AI governance (Floridi et al. 2018). What accountability mechanisms ensure AI systems respect stakeholder interests? (Diakopoulos 2016)
Stakeholder Contributions Stakeholders contributed resources but had limited power over resource use and value appropriation. Firms primarily controlled strategic decisions about resource deployment.	**Data Rights as Strategic Assets** Data protection laws, consent frameworks, and community norms shape what data firms can use and whether stakeholders will cooperate. Firms with stronger reputations for fairness, privacy, transparency, and reciprocity gain long-term resource advantage.	Stakeholders gain power when they control consent, data access, data deletion rights, and ability to restrict or condition data sharing. Regulatory bodies become dominant stakeholders.	• Data protection laws • Consent frameworks • Fairness, privacy, transparency • Regulatory power	How do power asymmetries shift when firms control algorithms mediating stakeholder interactions?[*] How do data rights and consent frameworks reshape competitive dynamics?[*] What mechanisms enable stakeholders to exercise voice in AI governance?[*]
Resource Creation Firms created and controlled resources. Stakeholders provided inputs but firms determined resource use and captured most value. Resource creation was largely firm-directed.	**Co-Creation and Ethical Governance** Firms partner with stakeholders through feedback, ratings, content creation, fine-tuning interactions, and crowdsourced testing. Stakeholders help develop and refine AI capabilities. Firms cannot monopolize value and must orchestrate co-creation.	Managing stakeholder risks (privacy fears, job displacement, algorithmic unfairness) becomes a capability that determines whether firms can sustain AI-driven advantage. Value creation becomes multisided and value appropriation becomes deeply contested.	• Feedback, ratings • Content creation • Fine-tuning interactions • Crowdsourced testing	How do firms orchestrate value co-creation with stakeholders in AI development?[*] What governance structures enable fair value appropriation across stakeholder groups?[*] How do firms manage stakeholder risks (privacy, displacement, unfairness) as strategic capabilities?[*]

[*] Theorized in this paper.

Power dynamics shift within strategizing groups: AI redistributes power by empowering junior analysts with sophisticated analysis tools, reducing reliance on experience-based intuition, shifting influence toward those who understand AI systems, and altering the expertise hierarchy. The practice of strategy becomes a negotiation between human expertise, algorithmic suggestions, and institutional expectations (Christin 2020).

In short, SAP will transform from episodic planning to continuous, AI-augmented strategizing (see Table 6).

The temporal rhythm will shift from periodic reviews and annual planning cycles to constant monitoring, instant experimentation, and continuous learning loops enabled by real-time dashboards. Consequently, AI-augmented practices where automated analyses, generative AI, simulation engines, and agentic workflows will materially transform how strategy is performed (Faraj et al. 2018; Raisch and Krakowski 2021). We also predict a shift in power dynamics when AI redistributes influence by empowering junior analysts with sophisticated tools (Raisch and Krakowski 2021), reducing reliance on experience-based intuition (Faraj et al. 2018), and altering expertise hierarchies by democratizing access to *technologies of rationality* (Jarzabkowski and Kaplan 2015). AI will join the cast of strategists as a quasi-practitioner, participating through summaries, producing analytic content, and flagging opportunities (Orlikowski and Scott 2008; Raisch and Krakowski 2021). This raises fundamental questions about who owns strategic decisions and what constitutes expertise when AI can generate equivalent outputs (Faraj et al. 2018)

4.6. A New Architecture of Strategic Edge: Individual, Firm, and Ecosystem Transformations

Across all five theoretical areas, AI has introduced a paradigm shift. Cognition has become hybrid, capabilities have become sociotechnical, ecosystems have become model-centric, resources have become stakeholder-generated, and strategizing has become AI-augmented. Traditional sources of advantage, economies of scale, cost leadership, differentiation, resource uniqueness—remain relevant but are increasingly overshadowed by learning-based advantages, data network effects, model governance capabilities, hybrid cognitive systems, and ecosystem orchestration skills.

AI functions simultaneously as disruptor, integrator, and change agent. As a disruptor, AI upends established business models and competitive positions through rapid capability advancement and unpredictable emergent behaviors (Christensen et al. 2015). As an integrator, AI's architectural homogeneity in FMs enables the combination of value creation activities across previously disconnected domains, while its probabilistic reasoning and generative capabilities provide normative guidance that reshapes decision processes and substitutes for human labor. As a change agent, AI's accelerating development since 2022 has compressed timescales for strategic adaptation, forcing organizations to respond to technological shifts that once unfolded over decades.

Building on these roles and the theoretical transformations outlined above, we define AI in strategic management terms as: *an adaptive general-purpose transformative system with powerful but incomplete capabilities in cognition, learning, and creativity that fundamentally reshapes how firms sense opportunities, make decisions, configure resources, coordinate activities, and innovate.* This new perspective necessitates expansions to established strategy frameworks and the development of new theoretical constructs. Our definition emphasizes not only what AI is ontologically, but what AI does strategically. It transforms the mechanisms through which competitive advantage is created, sustained, and appropriated. AI does not simply modify strategy theory; it rewrites it.

While the five theoretical perspectives examined above illuminate distinct facets of AI's impact on strategy, they collectively reveal a deeper multilevel architecture of transformation. AI does not simply alter individual theories in isolation; rather, it operates simultaneously across three fundamental levels of strategic analysis: individual cognition, firm-level capabilities, and ecosystem-level coordination. These levels are interconnected through cascading effects and feedback loops that amplify AI's transformative power.

At the individual level, AI fundamentally reshapes managerial cognition. Traditional strategy scholarship has long emphasized bounded rationality, cognitive biases, and limited information processing capacity as constraints on strategic decision-making (Simon 1947; Kahneman and Klein 2009; Gavetti and Levinthal 2000). AI introduces cognitive augmentation mechanisms including algorithmic decision support (Shrestha et al. 2021), distributed cognition systems (Riedl et al. 2024), and hybrid problem-solving capabilities (Raisch and Fomina 2025; Passerini et al. 2025) that transform individual cognition into hybrid human-AI architectures (Jarrahi 2018; Murray et al. 2021). Managers need no longer be bounded by the confines of their biological cognitive limitations. Instead, they can rely on AI-augmented judgment, enhanced information processing (Samuel et al. 2022), and complementary human-AI capabilities (Raisch and Krakowski 2021) that create new forms of causal reasoning and expertise (Glikson and Woolley 2020). This shift from bounded rationality to hybrid cognition represents a fundamental change in the microfoundations of strategic behavior (Eggers and Kaplan 2013; Gavetti 2012).

At the firm level, AI can transform the nature of competitive advantage. In the past competitive advantages were the result of differences in valuable, rare, inimitable, and non-substitutable resources. These were complemented by

Table 6: Strategy-as-practice: Pre-AI to post-AI transformation.

Pre-AI Theoretical Dimension	Post-AI Theoretical Dimension(s)	Transformation Description	Key Mechanisms	Research Implications
Episodic Planning Strategy unfolded in cycles through periodic reports, annual planning, and quarterly reviews. Human strategists conducted analyses using artifacts like documents and spreadsheets (Whittington 2006; Jarzabkowski and Spee 2009).	**Continuous Strategy** Strategy becomes dynamic and continuous, requiring always-on monitoring, instant experimentation, automated scenario testing, and continuous learning loops. Real-time dashboards replace periodic reports.	AI changes the rhythm of strategy work. This temporal shift breaks with long-established strategy routines, contrasting sharply with pre-AI episodic planning.	• Real-time dashboards • Always-on monitoring • Instant experimentation • Continuous learning loops	How does AI change micro-level practices of strategizing (meetings, planning cycles, decision routines)? (Jarzabkowski and Spee 2009; Whittington 2006) What new forms of strategizing emerge when AI enables real-time scenario analysis?[*] Future work should employ ethnographic methods, practice theory, and sociomateriality frameworks (Orlikowski 2007; Leonardi 2011).
Human Strategists Strategizing involved human practices through meetings, analyses, and routines that highlighted performative elements of strategy. Strategists included top managers, middle managers, consultants, and analysts.	**AI-Augmented Practices** Strategizing becomes materially transformed through automated analyses, AI-generated insights, simulations, and agent-powered workflows. Generative AI produces draft strategic alternatives, simulation engines evaluate competitive moves, and agentic AI conducts environmental scanning automatically.	AI becomes a ubiquitous strategy tool. These practices change not only outputs but how strategists act.	• Automated analyses • AI-generated insights • Simulation engines • Agentic AI workflows	How can practitioners navigate tensions between algorithmic recommendations and experiential judgment?[*] How do AI-augmented practices change the material and temporal structure of strategizing?[*] What role does AI play as a quasi-practitioner in strategic decision-making?[*]
Periodic Reviews Strategy work followed established hierarchies where expertise came from experience. Power derived from knowledge accumulation and senior position in organizational hierarchy.	**Real-Time Adaptation** AI redistributes power by empowering junior analysts with sophisticated analysis tools, reducing reliance on experience-based intuition, shifting influence toward those who understand AI systems, and altering the expertise hierarchy. AI participates in meetings through summaries and insights, produces analytic content, and flags risks and opportunities.	Power dynamics shift within strategizing groups. The practice of strategy becomes a negotiation between human expertise, algorithmic suggestions, and institutional expectations. Raises questions about who owns strategic decisions and how credit or accountability should be assigned.	• Empowering junior analysts • Reducing experience-based intuition • AI system understanding • Altered expertise hierarchy	How do power dynamics shift when AI systems participate in strategic decision-making?[*] Who owns strategic decisions when AI initiates strategic options?[*] What becomes of expertise when AI can generate equivalent outputs?[*]

[*] Theorized in this paper.

organizational capabilities, routines, and hierarchical coordination within clearly defined firm boundaries (Barney 1991; Teece et al. 1997). AI introduces opportunities for data accumulation, algorithmic capability building, sociotechnical integration, and dynamic reconfiguration that can shift the basis of advantage from resource stocks to learning velocities (Agarwal et al. 2022; Iansiti and Lakhani 2020; Raisch and Krakowski 2021). Post-AI competitive advantage will depend on learning-based advantages, sociotechnical capabilities that blend human and algorithmic actors, human-AI collaborative routines, porous organizational boundaries, and fluid, AI-enabled coordination structures (Faraj et al. 2018; Orlikowski and Scott 2008; Raisch and Fomina 2025). Firms will no longer compete primarily on what they own, but on how rapidly they learn.

At the ecosystem level, AI can reconfigure the locus and logic of competition. Traditional ecosystem theories emphasized network effects, modular architectures, platform governance, and complementary relationships within relatively stable value creation structures (Jacobides et al. 2018; Adner 2017). AI introduces FMs as organizing hubs, data ecosystems as competitive arenas, cross-platform learning as a source of spillovers, and AI-driven orchestration mechanisms as enablers of continuous co-evolution (Bommasani et al. 2021; Schrepel and Pentland 2025; Xu et al. 2025). The result is a shift toward model-centric ecosystems where advantage will accrue to those who control FMs, leverage data network effects, manage AI-driven complementarities, orchestrate distributed intelligence, and participate in dynamic value co-creation (Gawer and Cusumano 2014; Jones and Tonetti 2020). Competition will move from controlling modular interfaces to shaping learning systems that span organizational boundaries.

Note that these three levels do not operate independently. As Figure 2 illustrates, transformation at each level both emerges from and reinforces changes at adjacent levels (Felin and Foss 2005; Ployhart and Moliterno 2011). Individual-level cognitive transformations shape the development of firm capabilities, which in turn influence ecosystem-level dynamics. Simultaneously, feedback loops operate in reverse. Ecosystem-level structures constrain and enable firm-level strategies, which shape the cognitive tools and frameworks available to individual managers. AI amplifies these cross-level dynamics through three reinforcement mechanisms visible in the figure: refocusing at the individual level (as hybrid cognition continuously updates attention and interpretation) (Raisch and Fomina 2025), reconfiguration at the firm level (as learning-based advantages require ongoing capability renewal) (Teece 2018), and reinforcement at the ecosystem level (as model-centric structures entrench and evolve). Cross-level feedback loops further illustrate how ecosystem dynamics feedback to reshape both firm capabilities and individual cognition, creating a self-reinforcing cycle of AI-driven transformation.

This multilevel synthesis reveals why AI represents more than an incremental technological shift. AI forces a reconceptualization of strategy's foundational assumptions by simultaneously transforming cognition, capabilities, and

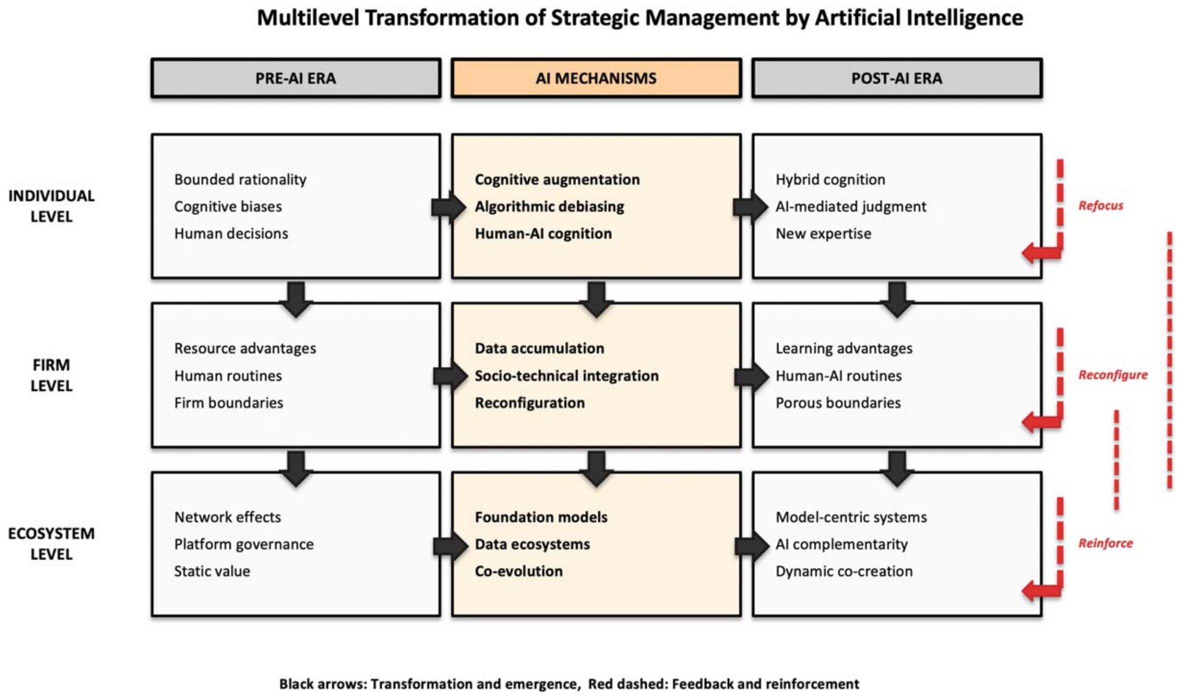


Figure 2: Multilevel transformation of strategic management by AI.

ecosystems, and by creating powerful feedback loops that amplify change across levels (Gavetti 2012; Powell et al. 2011). The traditional separation between internal (firm-level) and external (industry-level) sources of advantage will become increasingly untenable when learning processes, data flows, and algorithmic systems span and connect all three levels (Nambisan et al. 2017). A comprehensive theory of AI-infused strategy must therefore account for this multilevel, interconnected, and dynamically reinforcing architecture of transformation.

5. Conclusion and Empirical Research Agenda

AI is reshaping the foundations of strategic management. It alters what managers perceive, how organizations coordinate, how ecosystems evolve, how value is created and distributed, and how strategic work is performed. Based on this premise, we offer an AI-infused synthesis of strategy theory by (1) reframing classic strategy debates through an AI lens, and (2) extending five major theoretical domains—behavioral strategy, microfoundations, ecosystems, stakeholder RBV, and SAP, to incorporate the distinctive features of the AI landscape. We conclude by synthesizing the paper's contributions and implications for theory, empirical research, and practice.

In this paper, we have tried to accomplish two interrelated tasks. We have tried to reframe the classic debate in strategic management by viewing it through an AI lens and also offered contemporary extensions to five contemporary theories in strategic research in light of the new AI landscape.

5.1. Reframing Classic Strategy Debates through AI

The longstanding divide between external industry configuration (Porter 1980, 1985) and resource-based explanations (Barney 1991; Wernerfelt 1984) for strategic edge can be reframed through the lens of AI. We note that external industry configuration is increasingly shaped by model-centric ecosystems, data access regimes, and AI governance architectures (Bommasani et al. 2021; Schrepel and Pentland 2025).

Resource-based advantages depend on data richness, fine-tuning pipelines, sociotechnical capabilities, and hybrid learning loops (Agarwal et al. 2022; Raisch and Krakowski 2021). Therefore, AI serves as a synthesizing mechanism, enabling a unified perspective where strategic edge arises from orchestrating learning systems embedded in both the external environment and internal firm capabilities (Iansiti and Lakhani 2020).

5.2. Extending Five Contemporary Strategy Theories

Each contemporary theoretical trend gains depth when integrated with AI. Behavioral Strategy becomes the study of hybrid cognition and human-AI joint decision-making (Powell et al. 2011; Raisch and Fomina 2025). Microfoundations expand to include algorithmic actors, human-AI interaction structures, and sociotechnical emergence (Felin et al. 2015; Murray et al. 2021). Ecosystems theory becomes model-centric, emphasizing data flows, FM cores, and AI-driven complementarities (Jacobides et al. 2018; Xu et al. 2025). Stakeholder RBV foregrounds stakeholder-generated resources (data, feedback), fairness constraints, and ethical governance (Freeman et al. 2021; Floridi et al. 2018). Finally, SAP evolves into AI-augmented strategizing, where agentic tools shape routines, interactions, and legitimacy (Jarzabkowski and Spee 2009; Whittington 2006a). We conclude through our analysis that rather than replacing existing theories, AI extends their explanatory reach.

5.3. Implications for Strategy Research

AI challenges foundational assumptions of strategy theory. Several implications follow:

Strategy must incorporate nonhuman actors as part of microfoundations theory: Microfoundations theory (Felin et al. 2015; Felin and Foss 2005) assumes all strategic actors are human, with firm-level outcomes emerging from individual attributes, interactions, and routines. AI disrupts this by introducing algorithmic agents that exhibit goal-directed behavior, learn from experience, and influence outcomes (Gutiérrez 2024; Kellogg et al. 2020). Future research must theorize how agency is distributed across human-AI ensembles, drawing on actor-network theory (Latour 2005; Callon 1984), distributed cognition frameworks (Riedl et al. 2024; Sidji et al. 2024), and computational organizational theory. The following research questions deserve attention. Under what conditions do algorithmic agents complement versus substitute for human judgment? How do power dynamics shift when AI systems mediate strategic decisions (Shrestha et al. 2019)? What new governance forms emerge when responsibility is distributed between humans and machines (Floridi et al. 2018)?

Cognition becomes distributed between humans and AI, requiring new models of decision-making: Research should examine how AI augmentation affects managerial mental models, decision heuristics, and strategic frames

(Riedl et al. 2024; Gary and Wood 2011). Key questions: How does reliance on AI-generated insights alter cognitive processes underlying strategic choices (Kaplan 2011)? Under what conditions does AI amplify versus mitigate cognitive biases (Kahneman and Klein 2009; Dietvorst et al. 2015)? How do managers calibrate trust in algorithmic versus human judgment (Glikson and Woolley 2020)? Future work should integrate insights from human-computer interaction (Sidji et al. 2024), cognitive science, and behavioral decision theory (Camerer 2003).

Strategic advantage shifts from learning capability to adaptivity: AI fundamentally reorients competitive advantage away from learning capabilities, which were traditionally viewed as the ability to accumulate, codify, and leverage knowledge over time (Teece et al. 1997; Eisenhardt and Martin 2000) toward adaptivity: the capacity to continuously reconfigure in response to real-time feedback. While learning capability emphasizes building knowledge stocks that become durable resources (Grant 1996; Nonaka 1994), adaptivity emphasizes fluid responsiveness where insights immediately trigger reconfigurations rather than being stored for future use (Winter 2003; Zollo and Winter 2002). AI systems enable this shift because they operate through continuous sensing, pattern recognition, and adjustment rather than episodic learning cycles (Faraj et al. 2018). Organizations that excel at accumulating expertise may struggle in AI-infused environments if they cannot translate insights into immediate action. Future research should explore: What organizational structures enable continuous adaptivity rather than periodic learning cycles (Eisenhardt 1989)? How do firms build adaptive capacity without sacrificing the stability required for operational effectiveness (Benner and Tushman 2003)? What governance mechanisms ensure adaptive systems remain aligned with strategic intent? How do culture, leadership, and incentives sustain adaptivity (Samuel et al. 2023) without creating organizational exhaustion or strategic drift (March 1991)?

Ecosystems become model-driven structures rather than purely modular architectures: AI introduces a new organizing logic centered on FMs, data flows, and algorithmic interdependencies (Xu et al. 2025; Bommasani et al. 2021). In AI ecosystems, each participant's interactions generate data that improves a shared FM, creating knowledge spillovers that benefit all members including competitors (Khan et al. 2024; Jones and Tonetti 2020). This shifts competitive advantage from resource ownership to learning speed and ecosystem position (Iansiti and Lakhani 2020). Key questions: How do firms orchestrate multiactor learning systems (Hannah and Eisenhardt 2018)? What governance mechanisms align incentives when value creation depends on shared data and models (Gawer and Cusumano 2014; Jacobides et al. 2019)? Future work should integrate ecosystem theory (Adner and Kapoor 2010; Adner 2017) with research on open innovation (Chesbrough 2003; von Hippel 2005), standards competition (Katz and Shapiro 1985), and collective intelligence (von Krogh et al. 2012).

Stakeholder theory becomes intertwined with data governance and algorithmic fairness: AI makes stakeholders the primary source of critical inputs (data, feedback, content) while exposing them to algorithmic decisions affecting their welfare (Freeman 1984; Freeman et al. 2007). This creates new challenges: How should firms balance value appropriation with fair treatment of stakeholders who provide training data (Zuboff 2019)? What accountability mechanisms ensure AI systems respect stakeholder interests (Diakopoulos 2016)? How do power asymmetries shift when firms control algorithms mediating stakeholder interactions (Plantin et al. 2018)? Research should examine questions of algorithmic justice (Mittelstadt et al. 2016; Barocas and Selbst 2016), data rights (Khatri and Brown 2010), and stakeholder voice in AI governance (Floridi et al. 2018).

SAP must recognize AI as a participant in strategizing: AI transforms strategizing by generating scenarios, recommending actions, automating analyses, and reconfiguring workflows (Jarzabkowski and Spee 2009; Whittington 2006a). Key questions: How does AI change micro-level practices of strategizing (meetings, planning cycles, decision routines) (Jarzabkowski and Kaplan 2015)? What new forms of strategizing emerge when AI enables real-time scenario analysis (Doshi et al. 2025)? How can practitioners navigate tensions between algorithmic recommendations and experiential judgment (Kahneman and Klein 2009)? Future work should employ ethnographic methods, practice theory, and sociomateriality frameworks (Orlikowski 2007; Leonardi 2011; Orlikowski and Scott 2008).

These shifts redefine core constructs such as capability, resource, boundary, rivalry, and coordination (Teece 2018).

5.4. Implications for Methodologies

The integration of AI into strategic management research presents methodological complexities that extend beyond the scope of this conceptual paper. Unlike traditional strategic phenomena where methodological approaches can be relatively standardized, AI-infused strategy research requires context-specific designs that account for multiple interacting factors. The appropriate methodology for studying AI's strategic implications varies systematically based on the theoretical domain under investigation, the nature and role of the AI system itself, the level of risk associated with its deployment, and the organizational and industry context in which it operates. These factors interact in ways that make it impractical to prescribe uniform methodological templates. Instead, researchers must thoughtfully match methods to the specific characteristics of their empirical setting (Bhimani 2020; Christin 2020).

We identify four primary dimensions that should guide methodological choices in AI-infused strategy research, each requiring careful consideration when designing empirical studies. These dimensions can form the foundation for future inquiry into appropriate research designs in this emerging domain:

Strategy domain: The theoretical area under study determines the level of analysis, temporal scale, and appropriate causal inference strategies. For example, behavioral strategy studies examining cognitive effects would benefit from experimental designs that isolate decision-making processes (Powell et al. 2011), while ecosystem research requires network-level data and longitudinal observation of multiactor dynamics (Adner and Kapoor 2010).

Nature of AI engagement: The type of AI system (predictive analytics, generative models, computer vision, robotics, agentic systems) and its organizational role shape data requirements and feasible research designs (Dwivedi et al. 2021). For instance, predictive AI systems with clear outputs enable large quantitative studies and natural experiments, whereas agentic multiagent systems with emergent behaviors require simulation modeling, fine-grained process level data, and intensive case analyses (Klinger et al. 2020).

Level of risk: The potential for harm shapes ethical constraints on research design, ranging from permissible experimentation to observation-only approaches (Floridi et al. 2018). Consider that low-risk consumer applications allow field experiments and A/B testing, while human-critical systems (e.g., healthcare diagnosis, autonomous vehicles) require quasi-experimental designs and participatory methods that prioritize stakeholder safety, and infrastructure-critical systems (energy grids, financial networks) necessitate simulation-based approaches where live experimentation is not feasible.

Organizational and industry context: Organizational characteristics (size, structure, culture) and industry dynamics (competitive intensity, regulatory regime, stakeholder power) affect data accessibility, identification strategies, and generalizability (Bloom et al. 2019). Large established firms provide archival data suitable for multilevel quantitative analysis, while startups require intensive case approaches. On the other hand, heavily regulated industries (healthcare, finance) demand close collaboration with oversight bodies and emphasize observational designs, whereas fast-moving technology sectors create natural experiments through continuous AI deployment (Cockburn et al. 2018).

These four dimensions interact to create a complex methodological landscape. A comprehensive treatment of how researchers should navigate these trade-offs, including guidance on matching specific research designs to specific configurations of domain, AI type, risk level, and context, represents a substantial undertaking that merits dedicated development in future work. Our aim in this paper has been to establish the theoretical foundations for understanding how AI transforms strategy. Translating these theoretical insights into rigorous empirical research requires equally careful methodological development.

5.5. Implications for Managerial Practice

Building hybrid cognitive capabilities: Managers must develop competencies for leading organizations where cognition is distributed between humans and AI (Raisch and Fomina 2025; Jarrahi 2018), which include cultivating algorithmic literacy, fostering cultures balancing data-driven insights with experiential judgment (Samuel et al. 2023), designing governance mechanisms to clarify when humans should defer to or override algorithms (Glikson and Woolley 2020), and investing in training for effective human-AI collaboration. Leaders must model constructive collaboration and create norms to legitimize both algorithmic inputs and human intuition (Shrestha et al. 2019).

Orchestrating AI-enabled learning: Competitive advantage increasingly depends on learning velocity - the ability to rapidly improve systems through data and experimentation (Iansiti and Lakhani 2020; Agarwal et al. 2022). This requires mechanisms for continuous data collection and model retraining, organizational structures enabling rapid experimentation (Eisenhardt 1989), metrics tracking learning progress and algorithmic performance, and cultures rewarding learning from successes and failures (March 1991; Samuel et al. 2023). Managers must also navigate tensions between protecting proprietary data and participating in open ecosystems accelerating collective learning (Chesbrough 2003; Laursen and Salter 2006).

Designing governance mechanisms: AI ecosystems require new governance approaches addressing data sharing, intellectual property, quality standards, value allocation, and conflict resolution (Gawer and Cusumano 2014; Constantinides et al. 2018). Managers should proactively design governance mechanisms rather than allowing de facto arrangements to emerge organically (Boudreau 2010).

Navigating ethical dilemmas: AI adoption confronts managers with pressing questions. How can we address algorithmic bias and ensure fairness (Barocas and Selbst 2016; Mehrabi et al. 2021)? What transparency should we provide to different stakeholders (Diakopoulos 2016; Samuel 2024)? How do we manage job displacement and workforce transitions (Bessen et al. 2025; Brynjolfsson and McAfee 2014)? Where should we set boundaries around surveillance and privacy (Zuboff 2019)? What is our responsibility for our impacts on society (Floridi et al.

2018)? Managers must decide whether to establish ethics committees, conduct impact assessments, invest in bias detection, develop explainability practices, and support workforce reskilling. They must consider a fundamental question: Is our role simply to comply with regulations, or should we proactively shape AI development in socially responsible directions?

Anticipating competitive shifts: AI may fundamentally alter competitive dynamics (Christensen et al. 2015; Iansiti and Lakhani 2020). Managers must determine the answers to the following questions. How will AI reshape customer preferences and switching behaviors? Will AI lower entry barriers for new competitors (Bresnahan and Greenstein 1999)? Will AI commoditize our capabilities (Brynjolfsson and McAfee 2014)? How will AI affect economies of scale, scope, and firm boundaries (Williamson 1985)? Under different scenarios, how might industry structure evolve? What early warning signals should we monitor to anticipate rather than react to disruption (Christensen 1997)?

Building organizational resilience: AI introduces new uncertainties demanding resilience. Managers must decide. How much slack should we maintain for experimentation and failure recovery (March 1991)? How can we balance risky AI investments with stable core businesses (Benner and Tushman 2003)? What rapid response capabilities do we need for algorithmic failures? How can we foster cultures treating failure as learning (Levinthal and March 1993)?

Ultimately, AI strategy should be integrated into core strategic planning, not siloed within IT departments (Ransbotham et al. 2017). Managers who successfully navigate these challenges will position their organizations to thrive in an AI-infused competitive landscape. Those who treat AI as merely another technology risk strategic obsolescence.

5.6. Final Synthesis

AI challenges strategic management to rethink its foundational assumptions. Just as industrial organization economics, resource-centric perspectives, and behavioral theories transformed strategy over prior decades, AI represents the next major theoretical frontier. It introduces new actors (algorithmic agents), new resources (data and models), new forms of competition (learning races), and new organizing logics (interconnected network orchestration).

By analyzing five influential strategy theories, articulating a comprehensive theory of agentic AI, and proposing a wide-ranging empirical agenda, this paper seeks to provide a foundation for the next generation of strategy research. The field now has an opportunity to integrate insights from behavioral science, computer science, organizational theory, and ethics to build richer explanations of strategic edge in an AI-infused world.

The central message of this paper is clear: AI does not merely necessitate changes to the practice of strategy; it compels changes to the theory of strategy. The challenge ahead is to empirically test, refine, and extend the ideas outlined here so that the discipline can continue to offer rigorous, relevant insights in an era where intelligence, human and artificial, is at the core of strategic competition.

References

Adner, R. 2006. "Match Your Innovation Strategy to Your Innovation Ecosystem." *Harvard Business Review* **84**, no. 4: 98–107.

Adner, R. 2017. "Ecosystem As Structure: An Actionable Construct for Strategy." *Journal of Management* **43**, no. 1: 39–58.

Adner, R., and R. Kapoor. 2010. "Value Creation in Innovation Ecosystems: How the Structure of Technological Interdependence Affects Firm Performance in New Technology Generations." *Strategic Management Journal* **31**, no. 3: 306–333.

Agarwal, A., J. S. Gans, and A. Goldfarb. 2022. "The Simple Economics of Artificial Intelligence." *Journal of Economic Perspectives* **36**, no. 4: 3–24.

Aghion, P., B. F. Jones, and C. I. Jones. 2019. "Artificial Intelligence and Economic Growth." NBER (Working paper). Accessed March 14, 2026. https://www.nber.org/system/files/working_papers/w23928/w23928.pdf

Autio, E., and L. D. W. Thomas. 2014. "Innovation Ecosystems." In *The Oxford Handbook of Innovation Management*, edited by M. Dodgson, D. M. Gann, and N. Phillips, 204–288. Oxford University Press.

Baldwin, C. Y., and K. B. Clark. 2000. *Design Rules: The Power of Modularity*, volume 1. MIT Press.

Barney, J. 1991. "Firm Resources and Sustained Competitive Advantage." *Strategic Management Journal* **12**, no. 5: 99–120. https://doi.org/10.1002/smj.4250120108

Barney, J., and T. Felin. 2013. "What are Microfoundations?" *Academy of Management Perspectives* **27**, no. 2: 138–155. https://doi.org/10.5465/amp.2012.0107

Barney, J. B., D. J. Ketchen, Jr., and M. Wright. 2021. "Resource-Based Theory and the Value Creation Framework." *Journal of Management* **47**, no. 7: 1936–1955.

Barocas, S., and A. D. Selbst. 2016. "Big Data's Disparate Impact." *California Law Review* **104**, no. 3: 671–732. https://doi.org/10.2139/ssrn.2477899

Benner, M. J., and M. L. Tushman. 2003. "Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited." *Academy of Management Review* **28**, no. 2: 238–256.

Bessen, J., M. Goos, A. Salomons, and W. van den Berge. 2025. "Automatic Reaction: What Happens to Workers at Firms That Automate?" *The Review of Economics and Statistics* **107**, no. 1: 125–139.

Bhimani, A. 2020. "Digital Data and Management Accounting: Why We Need to Rethink Research Methods." *Journal of Management Control* **31**: 9–23. https://doi.org/10.1007/s00187-020-00295-z

Bhuyan, S. S., V. Sateesh, N. Mukul, A. Galvankar, A. Mahmood, M. Nauman, et al. 2025. "Generative Artificial Intelligence Use in Healthcare: Opportunities for Clinical Excellence and Administrative Efficiency." *Journal of Medical Systems* **49**, no. 1: 10.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, et al. 2019. "What Drives Differences in Management Practices?" *American Economic Review* **109**, no. 5: 1648–1683.

Bommasani, R., D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, et al. 2021. "On the Opportunities and Risks of Foundation Models." arXiv (Working paper). Accessed March 14, 2026. https://arxiv.org/abs/2108.07258

Boudreau, K. J. 2010. "Open Platform Strategies and Innovation: Granting Access vs. Devolving Control." *Management Science* **56**, no. 10: 1849–1872.

Bresnahan, T. F., and S. Greenstein. 1999. "Technological Competition and the Structure of the Computer Industry." *The Journal of Industrial Economics* **47**: 1–40.

Bresnahan, T. F., and M. Trajtenberg. 1995. "General Purpose Technologies: 'Engines of Growth'?" *Journal of Econometrics* **65**, no. 1: 83–108. https://doi.org/10.1016/0304-4076(94)01598-T

Brynjolfsson, E., and A. McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.

Callon, M. 1984. "Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay." *The Sociological Review* **47**: 196–223.

Camerer, C. F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Chesbrough, H. 2003. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business School Press.

Chidipothu, N., R. Anderson, J. Samuel, A. Pelaez, J. Esguerra, and M. N. Hoque. 2025. "Improving Large Language Model (LLM) Performance with Retrieval Augmented Generation (RAG): Development of a Transparent Generative AI University Support System for Educational Purposes." *Journal of Big Data and Artificial Intelligence* **3**, no. 1: 102–122.

Christensen, C. M. 1997. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press.

Christensen, C. M., M. E. Raynor, and R. McDonald. 2015. "What Is Disruptive Innovation?" *Harvard Business Review* **93**, no. 12: 44–53.

Christin, A. 2020. "The Ethnographer and the Algorithm: Beyond the Black Box." *Theory and Society* **49**: 897–918.

Cockburn, I. M., R. Henderson, and S. Stern. 2018. "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis." In *The Economics of Artificial Intelligence: An Agenda*, edited by A. Agrawal, J. Gans, and A. Goldfarb, 115–146. University of Chicago Press.

Constantinides, P., O. Henfridsson, and G. G. Parker. 2018. "Platforms and Infrastructures in the Digital Age." *Information Systems Research* **29**, no. 2: 381–400.

Cyert, R. M., and J. G. March. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall.

Diakopoulos, N. 2016. "Accountability in Algorithmic Decision Making." *Communications of the ACM* **59**, no. 2: 56–62.

Dietvorst, B. J., J. P. Simmons, and C. Massey. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err." *Journal of Experimental Psychology: General* **144**, no. 1: 114–126.

DiMaggio, P. J., and W. W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* **48**, no. 2: 147–160.

Doshi, A. R., J. J. Bell, E. Mirzayev, and B. S. Vanneste. 2025. "Generative Artificial Intelligence and Evaluating Strategic Decisions." *Strategic Management Journal* **46**, no. 3: 583–610.

Dwivedi, Y. K., L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, et al. 2021. "Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy." *International Journal of Information Management* **57**.

Eggers, J. P., and S. Kaplan. 2013. "Cognition and Capabilities." *Academy of Management Annals* **7**, no. 1: 293–338. https://doi.org/10.1080/19416520.2013.769318

Eisenhardt, K. M. 1989. “Making Fast Strategic Decisions in High-Velocity Environments.” *Academy of Management Journal* **32**, no. 3: 543–576.

Eisenhardt, K. M., and J. A. Martin. 2000. “Dynamic Capabilities: What Are They?” *Strategic Management Journal* **21**, no. 10–11: 1105–1121. https://doi.org/10.1002/1097-0266(200010/11)21:10/11≤1105::AID-SMJ133≥3.0.CO;2-E

Faraj, S., S. Pachidi, and K. Sayegh. 2018. “Working and Organizing in the Age of the Learning Algorithm.” *Information and Organization* **28**, no. 1: 62–70.

Felin, T., and N. J. Foss. 2005. “Strategic Organization: A Field in Search of Micro-Foundations.” *Strategic Organization* **3**, no. 4: 441–455.

Felin, T., N. J. Foss, and R. E. Ployhart. 2015. “The Microfoundations Movement in Strategy and Organization Theory.” *Academy of Management Annals* **9**, no. 1: 575–632.

Floridi, L., J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, et al. 2018. “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.” *Minds and Machines* **28**, no. 4: 689–707.

Freeman, R. E. 1984. *Strategic Management: A Stakeholder Approach*. Pitman Publishing.

Freeman, R. E., S. D. Dmytriyev, and R. A. Phillips. 2021. “Stakeholder Theory and the Resource-Based View of the Firm.” *Journal of Management* **47**, no. 7: 1757–1770.

Freeman, R. E., K. Martin, and B. L. Parmar. 2007. “Stakeholder Capitalism.” *Journal of Business Ethics* **74**, no. 4: 303–314. https://doi.org/10.1007/s10551-007-9517-y

Galaz, V., M. A. Centeno, P. W. Callahan, A. Causevic, T. Patterson, I. Brass, et al. 2021. “Artificial Intelligence, Systemic Risks, and Sustainability.” *Technology in Society* **67**: 101741.

Garvey, M. D., J. Samuel, and A. Pelaez. 2021. “Would you Please Like My Tweet?! An Artificially Intelligent, Generative Probabilistic, and Econometric-Based System Design for Popularity-Driven Tweet Content Generation.” *Decision Support Systems* **144**: 113497.

Gary, M. S., and R. E. Wood. 2011. “Mental Models, Decision Rules, and Performance Heterogeneity.” *Strategic Management Journal* **32**, no. 6: 569–594. https://doi.org/10.1002/smj.899

Gavetti, G. 2012. “Toward a Behavioral Theory of Strategy.” *Organization Science* **23**, no. 1: 267–285.

Gavetti, G., and D. Levinthal. 2000. “Looking Forward and Looking Backward: Cognitive and Experiential Search.” *Administrative Science Quarterly* **45**, no. 1: 113–137.

Gavetti, G., D. A. Levinthal, and J. W. Rivkin. 2005. “Strategy-Making in Novel and Complex Worlds: The Power of Analogy.” *Strategic Management Journal* **26**, no. 8: 691–712. https://doi.org/10.1002/smj.475

Gawer, A. 2014. “Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework.” *Research Policy* **43**, no. 7: 1239–1249.

Gawer, A., and M. A. Cusumano. 2014. “Industry Platforms and Ecosystem Innovation.” *Journal of Product Innovation Management* **31**, no. 3: 417–433.

Gillespie, T. 2014. “The Relevance of Algorithms.” In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by T. Gillespie, P. J. Boczkowski, and K. A. Foot. MIT Press. https://doi.org/10.7551/mitpress/9780262525374.003.0009

Glikson, E., and A. W. Woolley. 2020. “Human Trust in Artificial Intelligence: Review of Empirical Research.” *Academy of Management Annals* **14**, no. 2: 627–660.

Grant, R. M. 1996. “Toward a Knowledge-Based Theory of the Firm.” *Strategic Management Journal* **17**, no. S2: 109–122.

Guerras-Martín, L., A. Madhok, and Á. Montoro-Sánchez. 2014. “The Evolution of Strategic Management Research: Recent Trends and Current Directions.” *BRQ Business Research Quarterly* **17**, no. 2: 69–76.

Gutiérrez, J. L. M. 2024. “On Actor-Network Theory and Algorithms: ChatGPT and the New Power Relationships in the Age of AI.” *AI and Ethics* **4**, no. 4: 1071–1084.

Haim, A., and S. C. Meyer. 2024. “Competition between AI Foundation Models: Dynamics and Policy Recommendations.” *Industrial and Corporate Change* **33**, no. 4: 819–845. https://doi.org/10.1093/icc/dtae023

Hannah, D. P., and K. M. Eisenhardt. 2018. “How Firms Navigate Cooperation and Competition in Nascent Ecosystems.” *Strategic Management Journal* **39**, no. 12: 3163–3192.

Harrison, J. S., D. A. Bosse, and R. A. Phillips. 2010. “Managing for Stakeholders, Stakeholder Utility Functions, and Competitive Advantage.” *Strategic Management Journal* **31**, no. 1: 58–74

Hillman, A. J., and G. D. Keim. 2001. “Shareholder Value, Stakeholder Management, and Social Issues: What’s the Bottom Line?” *Strategic Management Journal* **22**, no. 2: 125–139.

Hossain, S., M. Fernando, and S. Akter. 2025. “Digital Leadership: Toward a Dynamic Managerial Capability Perspective of Artificial Intelligence-Driven Leader Capabilities.” *Journal of Business Research* **189**: 115103. https://doi.org/10.1016/j.jbusres.2024.115103

Iansiti, M., and K. R. Lakhani. 2020. *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World*. Harvard Business Review Press.

Jacobides, M. G., C. Cennamo, and A. Gawer. 2018. “Toward a Theory of Ecosystems.” *Strategic Management Journal* **39**, no. 8: 2255–2276.

Jacobides, M. G., A. Sundararajan, and M. W. Van Alstyne. 2019. “Platforms and Ecosystems: Enabling the Digital Economy.” https://www3.weforum.org/docs/WEF_Digital_Platforms_and_Ecosystems_2019.pdf

Jarrahi, M. H. 2018. “Artificial Intelligence and the Future of Work: Human–AI Symbiosis in Organizational Decision Making.” *Business Horizons* **61**, no. 4: 577–586.

Jarzabkowski, P., and S. Kaplan. 2015. “Strategy Tools-in-Use: A Framework for Understanding ‘Technologies of Rationality’ in Practice.” *Strategic Management Journal* **36**: 537–558.

Jarzabkowski, P., and A. P. Spee. 2009. “Strategy-as-Practice: A Review and Future Directions for the Field.” *International Journal of Management Reviews* **11**, no. 1: 69–95. https://doi.org/10.1111/j.1468-2370.2008.00250.x

Jones, C. I., and C. Tonetti, C. 2020. “Nonrivalry and the Economics of Data.” *American Economic Review* **110**, no. 9: 2819–2858.

Jordan, M. I., and T. M. Mitchell. 2015. “Machine Learning: Trends, Perspectives, and Prospects.” Accessed March 14, 2026. https://www.cs.cmu.edu/∼tom/pubs/Science-ML-2015.pdf

Kahneman, D., and G. Klein. 2009. “Conditions for Intuitive Expertise: A Failure to Disagree.” *American Psychologist* **64**, no. 6: 515–526.

Kaplan, S. 2008. “Framing Contests: Strategy Making under Uncertainty.” *Organization Science* **19**, no. 5: 729–752.

Kaplan, S. 2011. “Research in Cognition and Strategy: Reflections on Two Decades of Progress and a Look to the Future.” *Journal of Management Studies* **48**, no. 3: 665–695.

Kapoor, R., and S. Agarwal. 2017. “Sustaining Superior Performance in Business Ecosystems: Evidence from Application Software Developers in the iOS and Android Smartphone Ecosystems.” *Organization Science* **28**, no. 3: 531–551.

Kashyap, R., Y. Samuel, L. W. Friedman, and J. Samuel. 2024. “Artificial Intelligence Education & Governance—Human Enhancive, Culturally Sensitive and Personally Adaptive HAI.” *Frontiers in Artificial Intelligence* **7**: 1443386.

Katz, M. L., and C. Shapiro. 1985. “Network Externalities, Competition, and Compatibility.” *American Economic Review* **75**, no. 3: 424–440.

Kellogg, K. C., M. A. Valentine, and A. Christin. 2020. “Algorithms at Work: The New Contested Terrain of Control.” *Academy of Management Annals* **14**, no. 1: 366–410.

Khan, A. 2024. “The Intersection of Artificial Intelligence and International Trade Laws: Challenges and Opportunities.” *IIUM Law Journal*, **32**, no. 1: 103–152. https://doi.org/10.31436/iiumlj.v32i1.912

Khatri, V., and C. V. Brown. 2010. “Designing Data Governance.” *Communications of the ACM* **53**, no. 1: 148–152.

Klinger, J., J. Mateos-Garcia, and K. Stathoulopoulos. 2020. “Deep Learning, Deep Change?” *Research Policy* **49**, no. 7: 104042.

Latour, B. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.

Laursen, K., and A. Salter. 2006. “Open for Innovation: The Role of Openness in Explaining Innovation Performance Among UK Manufacturing Firms.” *Strategic Management Journal* **27**, no. 2: 131–150. https://doi.org/10.1002/smj.507

Leonardi, P. M. 2011. “When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies.” *MIS Quarterly* **35**, no. 1: 147–167.

Levinthal, D. A. 2011. “A Behavioral Approach to Strategy—What’s the Alternative?” *Strategic Management Journal* **32**, no. 13: 1517–1523. https://doi.org/10.1002/smj.963

Levinthal, D. A., and J. G. March. 1993. “The Myopia of Learning.” *Strategic Management Journal* **14**, no. S2: 95–112.

March, J. G. 1991. “Exploration and Exploitation in Organizational Learning.” *Organization Science* **2**, no. 1: 71–87.

McDonald, M. L., and J. D. Westphal. 2013. “Access Denied: Low Mentoring of Women and Minority First-Time Directors and Its Negative Effects on Appointments to Additional Boards.” *Academy of Management Journal* **56**, no. 4: 1169–1198.

McGahan, A. M. 2021. “Integrating Insights from the Resource-Based View of the Firm into the New Stakeholder Theory.” *Journal of Management* **47**, no. 7: 1734–1756. https://doi.org/10.1177/01492063211021656

Mehrabi, N., F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. 2021. "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys* **54**, no. 6: 1–35.

Mittelstadt, B. D., P. Allo, M. Taddeo, S. Wachter, and L. Floridi. 2016. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society* **3**, no. 2. https://doi.org/10.1177/2053951716679679

Murray, A., J. Rhymer, and D. G. Sirmon. 2021. "Humans and Technology: Forms of Conjoined Agency in Organizations." *Academy of Management Review* **46**, no. 3: 552–571.

Nambisan, S., K. Lyytinen, A. Majchrzak, and M. Song. 2017. "Digital Innovation Management: Reinventing Innovation Management Research in a Digital World." *MIS Quarterly* **41**, no. 1: 223–238.

Nelson, R. R., and S. G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Harvard University Press.

Nonaka, I. 1994. "A Dynamic Theory of Organizational Knowledge Creation. *Organization Science* **5**, no. 1: 14–37.

Orlikowski, W. J. 2007. "Sociomaterial Practices: Exploring Technology at Work." *Organization Studies* **28**, no. 9: 1435–1448.

Orlikowski, W. J., and S. W. Scott. 2008. "Sociomateriality: Challenging the Separation of Technology, Work and Organization." *Academy of Management Annals* **2**, no. 1: 433–474. https://doi.org/10.1080/19416520802211644

Ozdemir, S., J. C. Fernandez de Arroyabe, V. Sena, and S. Gupta. 2023. "Stakeholder Diversity and Collaborative Innovation: Integrating the Resource-Based View with Stakeholder Theory." *Journal of Business Research* **164**: 113955.

Passerini, A., G. Aryo, P. Minervini, B. Sayin, and K. Tentori. 2025. "Fostering Effective Hybrid Human–LLM Reasoning and Decision Making." *Frontiers in Artificial Intelligence* **7**.

Plantin, J.-C., C. Lagoze, P. N. Edwards, and C. Sandvig. 2018. "Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook." *New Media & Society* **20**, no. 1: 293–310.

Ployhart, R. E., and P. T. Moliterno. 2011. "Emergence of the Human Capital Resource: A Multilevel Model." *Academy of Management Review* **36**, no. 1: 127–150.

Porter, M. E. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press.

Porter, M. E. 1985. *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press.

Powell, T. C., D. Lovallo, and C. R. Fox. 2011. "Behavioral Strategy." *Strategic Management Journal* **32**, no. 13: 1369–1386.

Prahalad, C. K., and V. Ramaswamy. 2004. "Co-Creation Experiences: The Next Practice in Value Creation." *Journal of Interactive Marketing* **18**, no. 3: 5–14. https://doi.org/10.1002/dir.20015

Prashantham, S., and M. P. Healey. 2022. "Strategy-As-Practice Research: Reflections on Its Rationale, Approach, and Contributions." *Journal of Management Studies* **59**, no. 8: e1–e17.

Rahwan, I., M. Cebrian, N. Obradovich, J. Bongard, J.-F Bonnefo, C. Breazeal, et al. 2019. "Machine Behaviour." *Nature* **568**: 477–486. https://doi.org/10.1038/s41586-019-1138-y

Raisch, S., and K. Fomina. 2025. "Combining Human and Artificial Intelligence: Hybrid Problem-Solving in Organizations." *Academy of Management Review* **49**, no. 4: 808–830.

Raisch, S., and S. Krakowski. 2021. "Artificial Intelligence and Management: The Automation–Augmentation Paradox." *Academy of Management Review* **46**, no. 1: 192–210.

Ransbotham, S., F. Candelon, D. Kiron, B. LaFountain, and S. Khodabandeh. 2021. "AI and the Innovation Ecosystem." *MIT Sloan Management Review* **63**: 35–44.

Ransbotham, S., D. Kiron, P. Gerbert, and M. Reeves. 2017. *Reshaping Business with Artificial Intelligence: Closing the Gap between Ambition and Action*. MIT Sloan Management Review.

Riedl, C., S. Savage, and J. Zvelebilova. 2024. "AI's Social Forcefield: Reshaping Distributed Cognition in Human-AI Teams." arXiv (Working paper). https://arxiv.org/abs/2407.17489.

Samuel, J. 2021. "A Call for Proactive Policies for Informatics and Artificial Intelligence Technologies." *Scholars Strategy Network*. Accessed March 14, 2026. https://scholars.org/contribution/call-proactive-policies-informatics-and-artificial-intelligence-technologies

Samuel, J. 2024. "The Critical Need for Transparency and Regulation Amidst the Rise of Powerful Artificial Intelligence Models." *Scholars Strategy Network*. Accessed March 14, 2026. https://scholars.org/contribution/critical-need-transparency-and-regulation

Samuel, J., R. Kashyap, Y. Samuel, and A. Pelaez. 2022. "Adaptive Cognitive Fit: Artificial Intelligence-Augmented Management of Information Facets and Representations." *International Journal of Information Management* **65**: 102505.

Samuel, J., T. Khanna, J. Esguerra, S. Sundar, A. Pelaez, and S. S. Bhuyan. 2025. "The Rise of Artificial Intelligence Phobia! Unveiling News-Driven Spread of AI Fear Sentiment Using ML, NLP and LLMs." *IEEE Access* **13**: 125944–125969.

Samuel, J., A. Tripathi, and E. Mema. 2024. "A New Era of Artificial Intelligence Begins—Where Will It Lead Us?" *Journal of Big Data and Artificial Intelligence* **2**, no. 1: 1–15.

Samuel, J., M. Ye, T. Khanna, Y. Yao, X. Lin, R. Anderson, and A. Gupta. 2026. "Finder, Evaluator, Explainer, Generator (FEEG): A Bloom's Taxonomy-Based Query Classification Framework for LLMs and Generative AI." *Expert Systems with Applications* **304**: 130755.

Samuel, Y., M. Brennan-Tonetta, J. Samuel, R. Kashyap, V. Kumar, S. Krishna Kaashyap, et al. 2023. "Cultivation of Human Centered Artificial Intelligence: Culturally Adaptive Thinking in Education (CATE) for AI." *Frontiers in Artificial Intelligence* **6**: 1198180.

Schrepel, T., and A. Pentland. 2025. "Competition between AI Foundation Models: Dynamics and Policy Recommendations." *Industrial and Corporate Change* **34**, no. 5: 1085–1103.

Shrestha, Y. R., S. M. Ben-Menahem, and G. von Krogh. 2019. "Organizational Decision-Making Structures in the Age of Artificial Intelligence." *California Management Review* **61**, no. 4: 66–83.

Shrestha, Y. R., V. Krishna, and G. von Krogh. 2021. "Augmenting Organizational Decision-Making with Deep Learning Algorithms." *Journal of Business Research* **129**: 135–148. https://doi.org/10.1016/j.jbusres.2020.09.068

Shur-Ofry, M. 2024. "A Networks-of-Networks Perspective on AI Policy." *Network Law Review*.

Sidji, M., W. Smith, and M. J. Rogerson. 2024. "Adopting the Theory of Distributed Cognition for Human-AI Cooperation." In *Proceedings of the 36th Australian Conference on Human-Computer Interaction*, 486–491. Association for Computing Machinery.

Simon, H. A. 1947. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. Free Press.

Teece, D. J. 2018. "Business Models and Dynamic Capabilities." *Long Range Planning* **51**, no. 1: 40–49.

Teece, D. J., G. Pisano, and A. Shuen. 1997. "Dynamic Capabilities and Strategic Management." *Strategic Management Journal* **18**, no. 7: 509–533.

Tilson, D., K. Lyytinen, and C. Sørensen. 2010. "Digital Infrastructures: The Missing IS Research Agenda." *Information Systems Research* **21**, no. 4: 748–759. https://doi.org/10.1287/isre.1100.0318

Tripathi, A., J. Samuel, M. Brennan-Tonetta, H. Nguyen, and E. Mema. 2025. "When Machines Create! Envisioning Our Future with the Transformative Power of Generative AI." *Journal of Big Data and Artificial Intelligence* **3**, no. 1: 1–4.

Ustahaliloğlu, M. K. 2025. "Artificial Intelligence in Corporate Governance." *Corporate Law and Governance Review* **7**, no. 1. https://doi.org/10.22495/clgrv7i1p11

von Hippel, E. 2005. *Democratizing Innovation*. MIT Press. https://doi.org/10.7551/mitpress/2333.001.0001.

von Krogh, G., S. Haefliger, S. Spaeth, and M. W. Wallin. 2012. "Carrots and Rainbows: Motivation and Social Practice in Open-Source Software Development." *MIS Quarterly* **36**, no. 2: 649–676.

von Krogh, G., Q. Roberson, and M. Gruber. 2023. "Recognizing and Utilizing Novel Research Opportunities with Artificial Intelligence." *Academy of Management Journal* **66**, no. 2: 367–373.

Wan, X., and P. Zhao. 2024. "Research on the Impact of Artificial Intelligence on Corporate Competitive Advantage." *Frontiers in Business, Economics and Management* **17**: 438–443.

Wernerfelt, B. 1984. "A Resource-Based View of the Firm." *Strategic Management Journal* **5**, no. 2: 171–180.

Whittington, R. 2006. "Completing the Practice Turn in Strategy Research." *Organization Studies* **27**, no. 5: 613–634. https://doi.org/10.1177/0170840606064101

Williamson, O. E. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. Free Press.

Winter, S. G. 2003. "Understanding Dynamic Capabilities." *Strategic Management Journal* **24**, no. 10: 991–995. https://doi.org/10.1002/smj.318

Wu, P., and Y. Wang. 2021. "The Impact of Artificial Intelligence on Student Learning Outcomes: A Meta-Analysis." *The Asia-Pacific Education Researcher*.

Yoo, Y., O. Henfridsson, and K. Lyytinen. 2010. "The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research." *Information Systems Research* **21**, no. 4: 724–735. https://doi.org/10.1287/isre.1100.0322

Xu, F., Y. Zhang, L. Chen, and H. Wang. 2025. "The Economics of AI Foundation Models: Openness, Competition, and Governance." arXiv (Working paper).

Zollo, M., and S. G. Winter. 2002. "Deliberate Learning and the Evolution of Dynamic Capabilities." *Organization Science* **13**, no. 3: 339–351. https://doi.org/10.1287/orsc.13.3.339.2780

Zuboff, S. 2019. *The Age of Surveillance Capitalism*. PublicAffairs.

Journal of Big Data and Artificial Intelligence

The Journal of Big Data and Artificial Intelligence
publishes one volume of high quality scholarly and practitioner articles on AI and data annually, along with special issues on a rolling basis. Accepted articles are made available online for early access—to submit articles, please visit.

http://JBDAI.org

Journal of Big Data and
Artificial Intelligence