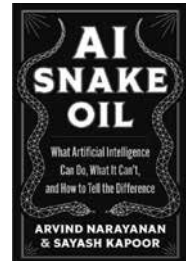


BOOK REVIEW

AI Snake Oil: What Artificial Intelligence Can Do, What It Can't and How to Tell the Difference

Published by: Princeton University Press

ISBN: 9780691249131



GOWRI R

Artificial Intelligence (AI) is proliferating rapidly in various fields such as healthcare diagnostics, human resource management, criminal justice, social media governance as well as in the field of defence, and has generated an unprecedented convergence of expectation, investment and anxiety. About 61 per cent of global venture capital was captured by AI firms from private investors in 2025, while US\$ 33.9 billion was invested in generative AI alone in 2024.^{1,2} Despite the increasing investment and usage in AI, there

Ms **Gowri R** is a Research Associate at the Centre for Aerospace Power and Strategic Studies (CAPSS), New Delhi.

1. "AI Firms Capture 61% of Global Venture Capital in 2025," OECD, February 17, 2026, <https://www.oecd.org/en/about/news/announcements/2026/02/ai-firms-capture-61-percent-of-global-venture-capital-in-2025.html>. Accessed on May 1, 2026.
2. Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Toby Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, Sukrut Oak. "The AI Index 2025 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025, <https://doi.org/10.48550/arXiv.2504.07139>. Accessed on May 11, 2026.

is no clear understanding on how it works. The book *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*, published by Princeton University Press in 2024, helps readers to develop a clear understanding of the capabilities and limitations of contemporary AI. In 2024, *Nature*, one of the world's leading peer-reviewed science journals, recognised this book as one of its ten best books of the year. The authors compare the hype of AI to that of oldtime snake oil peddlers who exploited people's trust, which lacks scientific evidence, and, thus, named the book "*AI Snake Oil*"

The authors of the book are:

1. Arvind Narayanan is a professor of Computer Science at Princeton University and director of the Centre for Information Technology Policy (CITP). He led the Princeton Web Transparency and Accountability Project whose early research was to demonstrate that machine learning algorithms reflect and perpetuate cultural stereotypes.
2. Sayash Kapoor is a senior fellow at Mozilla and a Computer Science doctoral candidate at Princeton. He is known for peer-reviewed research on the fundamental limits of predictive models.

Together, they combine deep technical expertise with social and policy dimensions of AI deployment, a combination that distinguishes this book from other popular AI discourse.

SUMMARY OF THE BOOK

The first chapter establishes the conceptual framework by introducing three types of AI snake oil: AI that fundamentally does not and cannot work; AI that works but is overhyped and leads to over-reliance, replacing human expertise, rather than augmentation; and, finally, AI that works well but is harmful in its application. This typology is the book's most durable contribution and it enables readers to not only question "Does this AI work?" but also "What kind of harm does it create and what solution is appropriate?"

Chapter Two, “How Predictive AI Goes Wrong?”, analyses why predictive AI fails during deployment. These failures arise due to various reasons such as training data that embeds historical human biases and generates inequality, differences in real world environmental conditions, the phenomenon of “over-automation” and different cultural specificities across different regions of the world.

Chapter Three, “Why Can’t AI Predict the Future?” deepens this analysis by confronting the epistemological foundations of predictive AI. The natural interest of people to learn about the future always being the intuitive developments of predictive AI. However, the authors argue that human social outcomes are structurally resistant to prediction because human lives are driven by chance events, feedback loops and the interventions of other actors. Case studies such as the Fragile Families Challenge of 2017 to predict the life of children at the age of fifteen resulted in no correct prediction, despite being provided with rich data.

Chapter Four “How AI Works?”, provides rich technical genealogy of Large Language Models (LLM) and image generators from the early perceptron of 1958 through the “AI winters” of the 1970s and 1980s, to the pivotal AlexNet breakthrough of 2012, when an eight-layer deep neural network achieved superior image classification performance. This chapter traces the scaling of training datasets from hundreds of thousands of labelled images to billions of scraped web pages and images, including the over five billion annotated images used to train stable diffusion. It walks the reader through the arithmetic token generation concept which explains how a single ChatGPT response of a few hundred words requires approximately one quadrillion arithmetic operations. This chapter also examines the biases in labelling the images that lead to sexualise and racialise outputs, and it confronts the ethical failure of AI companies that used artists’ work to train commercial systems without consent, credit or compensation. The authors acknowledge that GenAI works impressively well but with ethical shortcuts that are difficult to correct.

Chapter Five addresses the most popular question of whether advanced AI constitutes an existential risk to humanity. The authors explain that Advanced Generative Intelligence (AGI) has a long way to go as each step of the ladder takes time to implement and fulfil. It is disadvantageous to stop the growth of any technology and the main concern as per the authors is:

“What will humans do with AI?”, which invokes the several questions with plausible answers such as “Will the autonomous decision that AI makes be acceptable humans?” and “Do humans really understand and analyse the results provided by autonomous decision-making AI systems or will they blindly believe it?”.

Chapter 6 seeks an answer to the question, “Why Can’t AI Fix Social Media?”. The authors identify seven structural limitations of AI for content moderation: (1) its inability to grasp context and cultural nuances; (2) under-investment in non-Western language capabilities leading to documented failures in Myanmar, Ethiopia, and South Asia; (3) the temporal mismatch between static trained models and a constantly evolving linguistic and social landscape; (4) adversarial manipulation by inimical actors who quickly learn to evade classifiers; (5) the difficulty of predicting the real-world consequences of platform interventions; (6) the distorting effects of legal regulation produces “collateral censorship” as platforms over-moderate to minimise legal exposure; and (7) the fundamentally political character of content policy, which cannot be delegated to a technical system without also delegating the political decisions embedded in it. The chapter implies that institutions have to adopt ethical AI algorithms to avoid the dangers of content moderation.

Chapter Seven, “Why Do Myths About AI Persist?” is about the anatomy of the AI hype ecosystem. The authors trace the self-reinforcing cycle where companies, researchers, university communications offices, journalists, and public figures collectively produce and maintain inflated expectations

about AI capabilities. The case study of deployment of AI in hospitals explains that the myth-making is not merely a matter of honest misunderstanding but involves active commercial incentives, inadequate journalism, and structural failures of transparency and reproducibility in AI research itself.

The final chapter, “Where Do We Go From Here?”, turns from diagnosis to prescription. The authors argue for a regulatory framework that neither treats AI as self-governing nor demands the creation of entirely new legal machinery. The authors argue that there is a need for the vigorous enforcement of existing consumer protection, anti-discrimination, and data privacy regulations. They analyse the emerging global landscape such as the European Union’s risk-based AI Act, which requires transparency on high-risk AI applications in hiring, educational access, and worker management; the General Data Protection Regulation (GDPR); the Digital Services Act; and the United States of America’s approach of delegating AI oversight to sector-specific agencies. The book urges careful, evidence-driven choices about AI use to minimise harm and protect future generations.

CONCLUSION

AI Snake Oil focusses on providing public awareness by differentiating types of AI, especially generative and predictive AI. It achieves this purpose by providing suitable evidence-based case studies. It not only provides policy related aspects but also dives deeply into the technical background of models, implementation, and application of AI without using technical jargon. This helps the audience from the fields who wish to learn about AI, such as policy-makers and the general public, understand which AI serves what purpose by providing deep technical knowledge. Thus, it produces a balance between technical depth and outcomes that impact the general public. This book also implies that the already prevailing laws can be applied for AI regulation instead of waiting to make new AI laws and Acts. Further, the book suggests that users of AI need to understand the training data used for development, the existing cultural differentiation in countries, and the real-world factors that may affect the results and predictions. However, the book

underestimates the growth of AI technology and its current increase in accuracy. Though it explains how AI solutions can affect real-world problems, it does not speak about the use of AI in the military systems which is being seen increasingly in the contemporary world.