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

**By Arvind Narayanan and Sayash Kapoor**
**Princeton University Press**
**September 24, 2024**
**357 pages**

## Navigating the AI Hype: Why *AI Snake Oil* is Crucial for Interpreting Advancements and Inherent Limitations

**Komal Vij,**

**Stanford University**

> *Snake oil: a substance with no real medicinal value sold as a remedy for all diseases* (Oxford English Dictionary)

In *AI Snake Oil*, Arvind Narayanan and Sayash Kapoor dismantle the AI hype with a cautious and informed lens, equipping readers with a foundation to do the same. Both authors—renowned computer scientists and TIME's 100 Most Influential People in AI—approach the subject with expertise and clarity, making the book accessible and insightful for general readers, technical professionals, and policymakers alike.

From spellcheck to self-driving cars, the authors convincingly show how the definition of "AI" continues to shift as the field evolves. They address widespread misconceptions by providing expert guidance on how to evaluate AI claims and distinguish fact from fiction. Narayanan and Kapoor effectively analyze the trade-offs of predictive and generative AI, helping readers understand how these systems work and, crucially, where their limitations lie. By grounding their arguments in technical knowledge, they help deflate inflated expectations.

The book opens with a clever analogy, comparing discussions of AI to calling all forms of transport "vehicles"—lumping together cars, bikes, spacecraft, and more. This, the authors argue, is similar to how AI is overgeneralized, leading to confusion. They warn that "there's almost nothing one can say in one breath that applies to all types of AI" (p. 13). The tongue-in-cheek definition—"AI is whatever hasn't been done yet"—highlights the tendency to label only the newest innovations as "true" AI.

This book provides clarity for readers overwhelmed by hype. The virality of Narayanan's MIT lecture on AI snake oil demonstrates public appetite for informed skepticism. The authors note, "most of us suspect that a lot of the AI around us is fake, but we don't have the vocabulary or the authority to question it" (p. 17). This book serves as a remedy—giving readers, even those with limited technical background, the tools to interrogate misleading claims.

Narayanan and Kapoor critique predictive AI systems by emphasizing their unintended consequences. They offer a concrete definition of a model as "a set of numbers that mathematically specify how the system should behave" (p. 39), and show how such models—though seemingly objective—are often used in high-stakes decisions like hiring or loan approval. A standout example

involves an AI model used to predict pneumonia outcomes, which incorrectly concluded that asthma patients were at lower risk and should be discharged early. The model's misjudgment, based on training data correlations, could have led to fatal outcomes—underscoring the authors' key point: "a good prediction is not a good decision" (p. 42).

The book further explains where predictive AI thrives—such as spam detection—and where it struggles, especially in complex, low-data social contexts like predicting eviction risk. Narayanan and Kapoor emphasize that while improved data can enhance predictions in some domains, others may remain inherently unpredictable.

On generative AI, the authors bring both enthusiasm and caution. Declaring themselves "enthusiastic users of generative AI" (p. 99), they describe how tools like ChatGPT can produce text, images, and video. But misconceptions often give rise to panic. For instance, when a New York Times journalist tested Bing's chatbot and received eerily sentient responses, many were alarmed. The authors explain that these results stem from training data—fictional stories of sentient AI—that the model reproduces. Generative models are not reasoning beings but pattern matchers, and anthropomorphizing them only fuels misunderstanding.

Even if training data were limited to only true statements, the nature of generative models—probabilistic pattern completion—means they will still "hallucinate" and generate false outputs (p. 138). This insight helps readers temper their expectations and interpret outputs more critically.

Although the book stresses AI's limitations, the authors are not dismissive of progress. They cite promising uses, like the *Be My Eyes* app, which pairs visually impaired users with a vision-capable AI assistant. By acknowledging both potentials and pitfalls, Narayanan and Kapoor provide a balanced, technically grounded guide for navigating AI's present and future.

The book's central message is both sobering and empowering: AI is not magic, and it cannot solve all our problems. But with informed skepticism and clear-eyed understanding, we can harness it more ethically and effectively. *AI Snake Oil* is an indispensable contribution to the growing literature on AI and society.