Indian Public Policy Review 2025, 5(6): 125-128
https://doi.org/10.55763/ippr.2025.05.06.005

IPPR

The Fine Print on AI: Debunking AI Myths

Book review of "AI Snake Oil" by Arvind Narayanan & Sayash Kapoor

Adya Madhavan*

The foray of artificial intelligence into practically every domain, has led to a massive volume of claims about its abilities. On the one hand, there are claims about how self-driving cars are likely to be a possibility for all – something that sci-fi has featured for years, which is a growing reality today. On the other hand, there are reports that AI is sentient, which is currently beyond the realm of possibility, and seems like an outlandish claim. Those without a technical background especially, often struggle to weed out facts from fantasies.

Arvind Narayanan and Sayash Kapoor's 'AI Snake Oil' attempts to cut through the noise and answer some of our most burning questions about AI and its abilities. Authored by a PhD student and a Professor of Computer Science at Princeton University, the book has one fundamental objective: to equip the general public with the information they need to identify AI 'snake oil', which is how the authors describe artificial intelligence that doesn't work as claimed. With the advent of models like ChatGPT and Gemini bringing AI into our daily lives, there is much talk of its transformative power. Narayanan and Kapoor aim to demystify the capabilities of AI, and simultaneously clarify many common misconceptions.

The biggest strength of the book lies in the fact that it assumes its readers don't have a tech background. Before delving into the myths and seemingly magical abilities of AI, Narayanan and Kapoor help readers understand what constitutes AI: an umbrella term that is used to describe some dated technology as well. For instance, they provide examples of technologies such as spell check software or robot vacuum cleaners. These are technologies that technically qualify as AI, but are not regarded as such, since they are simple and ubiquitous – in comparison to (more advanced) AI, that is seen as cutting-edge.

The book makes three important arguments: firstly, that predictive AI is inherently flawed, and 'does not and cannot work as advertised'; secondly, that generative AI comes with both its strengths and weaknesses, and should be utilised with caution; and last of all, that the utilisation of AI to mitigate some of the issues of social media is misguided and often oversimplified. Each of their three central arguments is bolstered by a fair amount of context setting through a brief history of the technology and comparisons to other developments that help readers go beyond the surface. For

* Adya Madhavan is a Research Analyst at the Takshashila Institution working with the high-tech geopolitics program.

instance, the chapter on generative AI highlights the development of ImageNet, a database for visual object recognition.

It was interesting to read a book about AI that acknowledged some of the obscurity that exists when classifying technology as artificial intelligence, with cognisance of the fact that there has always been a lack of a constant definition. The fact that once advanced technologies have now faded into the mundane provides much food for thought: perhaps today's advanced technologies, too, will one day be seen the same way.

The authors then distinguish between generative and predictive AI before going into the meat of the book. The authors' conversational tone is jargon-free for the most part, making it a relatively light read despite the dense and somewhat technical subject matter. Their use of examples and analogies that are simple to understand from a layman's perspective makes an otherwise complex topic much more accessible. For example, the authors compare the lack of understanding about AI to 'vehicles' in an alternate universe, wherein that word is essentially an umbrella term that refers to all forms of transportation, leading to chaos and confusion.

While the tone throughout the book emphasises caution, Narayanan and Kapoor do not take away from the many areas where AI excels, such as the automation of mundane tasks and tasks such as image classification and generation. They also acknowledge that they utilise AI for a range of purposes, such as simply doing *better* than they would without its aid, or in areas where it simplifies tasks that would otherwise be tedious and time-consuming, such as generating citations.

Narayanan and Kapoor also tackle a question that is repeatedly brought up: will the advancement of AI be an existential threat to humanity? Their take is that fundamentally, humans are responsible for when and how they deploy AI, and therefore, a sentient, all-powerful superintelligence will not be spontaneously born and take over the world as we know it.

They argue that AI has developed through a 'ladder of generality', where developments are built upon previous progress. The analogy essentially means that each higher rung stands for more flexible systems capable of performing new tasks – more 'general' systems. AI, in its current form, is on the middle rungs of the ladder (pre-trained models and instruction-tuned models). Because we do not know what developments could take place in AI, the ladder of generality has an unknown number of rungs, and one can only speculate what comes next or what direction things will take.

Given that technologies are built upon previous progress, the future development of artificial intelligence will likely follow a similar trajectory and allow people to adapt and implement safeguards. However, there is a caveat—there's no way to know if the current trajectory can be lead to more general AI or if it is a path that leads nowhere. In every wave of AI, researchers have believed that the paradigm they believe in can lead to advanced developments in AI.

The existential threat that AI could prove to be is one of the major concerns about AI in mainstream discourse. Still, Narayanan and Kapoor take it with a pinch of salt and don't seem to give it too much weight. Not to say they aren't cognizant of *any* of the threats posed by AI; they warn

against areas where it can have detrimental effects, but reiterate the need to focus on combating specific issues, instead of fretting over a speculative doomsday scenario with little scientific basis.

They use analogies to explain why it is merely an extreme hypothetical situation and not an actual threat, spending a fair amount of time on going into the explanation. One thought experiment frequently used to justify fears of an existential threat posed by AI is the 'Paperclip Maximiser' experiment. In this experiment, an AI system eradicates humans accidentally in an effort to maximise paperclip production. In this scenario, the AI sees a need for more resources to produce paperclips, and realises the human race is hindering acquiring those resources. However, Kapoor and Narayanan argue that this reasoning assumes that AI is mighty but lacks fundamental concern for human well-being– a flawed premise. They believe that such mindless literalness is not a feature of modern AI systems that have certain in-built safeguards and a more nuanced process of interpretation. An intuitive AI (AGI) – a system even more advanced than modern AI that is in use currently – should be able to discern that this is bad for humanity and exit performing the function if need be.

On the subject of what safeguards need to be implemented if AGI comes into being, the authors argue that this is only a consideration if AGI displays behaviours that are inherently 'power-seeking'. Safeguards can also only be built when there is more clarity as to what the problems of the systems will be. The other alternative could then be to not attempt to build such a system, but this will require surveillance like never before, as well as international cooperation.

Some of the book's recommendations seem to fall short, such as the need to fix 'broken institutions'-- such as underfunded public schools that fuel the need for AI snake oil. These schools lack the infrastructure and number of staff they require, and end up resorting to using quick AI fixes, and sometimes utilise tools that are ineffective or inaccurate. While Narayanan and Kapoor spend a few pages outlining what these broken institutions look like, they spend little over a paragraph explaining what the solutions to this issue are. For some of these systemic problems, it would be useful to have more clarity on how to find solutions.

Other recommendations are arguably great ideas, albeit ones where we are yet to see what implementation will look like on a large scale. For instance, the authors suggest the use of partial lotteries for the allocation of limited resources such as insurance or research grants, instead of using predictive AI to 'optimise' the selection process based on pre-set criteria. The example they use is a partial lottery for university applications, where applicants are selected randomly, out of a larger pool of applicants who meet the wider pre-set criteria. The same logic can be used for research grants, where according to the authors, instead of focusing on grant applications, if the selection process if more random, researchers can focus on actual *research* more.

While the idea of partial lotteries solves for a lot of the issues of both (faulty and biased) AI or human committees, there are decades of acceptance of these systems that will have to be reversed. For instance, currently students focus on volunteering and developing unique skills to stand out in their applications, something which is encouraged by both families and educational institutions alike. If a partial lottery system were to be used, these internalised practices would have to be changed.

Nonetheless, if implemented, it would be interesting to see how such systems would play out in the long run.

Even though the implementation of some of the book's recommendations require long-term change, radical ideas and new thinking could be the need of the hour. In terms of what seems to be the book's focus – providing readers with a non-technical background the information they need to understand this technology and navigate the many gimmicky AI products– the book does an excellent job. Given the emphasis on emerging technologies in today's world, AI Snake Oil equips readers with a compact guide to understanding some of the nuances of artificial intelligence, which are becoming increasingly important.

"AI Snake Oil" by Arvind Narayanan and Sayash Kapoor, 2024, Princeton University Press, Pages 352.