Review Essay

Artificial Super Intelligence: Are we there yet?

Artificial Intelligence: A guide to thinking humans by Melanie Mitchell. London: Penguin, 2019.

The World Economic Forum's recent report on the fourth industrial revolution (*The Future of Jobs Report*, 2018) outlines artificial intelligence (AI) as one of the major drivers to innovation and a potential disruptor to job markets around the world. Schools of thought, around the use and development of AI, range from intelligent machines like autonomous cars or language translators assisting us in our daily lives, to notions of malevolent robots, inspired by science fiction movies, capable of bringing upon us the destruction of the world we live in. With this background, Mitchell's book is timely and serves to explain the realities of contemporary AI research and why she thinks an all-encompassing artificial general intelligence (AGI) is many years away from realization. She focuses on the history of AI research while addressing the origins of this 'hype problem of AI'. She also provides technological explanations to some of the achievements made in the domain replete with examples and a geeky sense of humor. The book outlines various chapters segregated into sections focusing on the background and history of AI research, its applications in the visual, language and games domain and lastly focusing on ethics in AI and questions on meaning-making that she believes AI researchers should focus on.

Mitchell begins the book by recounting a personal experience with her PhD advisor Douglas Hofstadter (author of G.E.B) at Google. Hofstadter feels threatened that in our not very distant future, humankind’s creativity, our capability to produce emotions and consciousness could be easily reproduced and mechanized by an AGI. Few other well-known names in the technology industry, the likes of Ray Kurzweil of Google (who believes in the idea of an AI 'Singularity', an age where machines are smarter than us), Stephen Hawking, Elon Musk and Bill Gates have also expressed their belief that an AGI could be an existential threat to human kind. She terms this notion or belief that we are close to building an AGI, which is smarter than human beings and capable of destroying us, collectively as the ‘hype problem of AI’. She deconstructs this notion in the subsequent chapters in the book. While researchers have successfully built intelligent machines for language translation (e.g. Google translate) or machines that can win in the game of Go or Chess, she challenges the notion of building intelligent machines that can do everything - one AI program that translates language, beats you at Go and Chess as well as drives you to work and back home safely!

In writing this book, Mitchell has accomplished outlining a brief history of AI. From John McCarthy's intent to coin the term ‘artificial intelligence’ to differentiate this science from the then prominent Cybernetics theory, to the more recent debates on the ethics of AI. Though she indicates that the book was not intended to serve as general history or survey on the topic, laying this historical foundation helps readers understand some of the reasons behind the so called 'AI Winter' (the period from the 1960's to the late 1980's where no real progress was being shown by then prominent symbolic AI architectures) and its subsequent 'AI Spring' (the period after 1988 when DARPA announced AI being more important an innovation than the atom bomb). This foundation also helps in understanding the roots of AI in symbolic architectures and its more recent focus on sub-symbolic architectures during the 'AI Spring'.

Some of the common uses of AI these days surround image recognition (used in face recognition or autonomous car driving) and natural language processing (used in language translation). Mitchell provides a big picture overview on the inner workings of these algorithms. She also explains supervised and unsupervised learning algorithms and alludes to some of the big problems in contemporary AI approaches including adversarial attacks on AI, the difficulty in building transparent machines or explainable AI algorithms. She uses examples from some Google researchers’ work on adversarial attacks, where scientists changed a few pixels on an image of a yellow school bus to dupe an image recognition algorithm to classify the image as an ostrich! The promise of such witty examples in different parts of the book, peppered with analogies like comparing the concept of an explainable AI to your school mathematics teacher expecting you to 'show your work' than just the answer, helps make this book a great read.

An interesting theory Mitchell builds upon is the notion that AI researchers are driven primarily by competition and the need to win. Most of the accomplishments in AI in the last decade were due to researchers competing against each other at competitions like Pascal Visual Object Challenge (PVOC), where in 2012 the first promising AI algorithm could classify 85% of 3 million images correctly. This started, what she terms as, the AI gold rush, where companies started channeling more investments into building AI capabilities not just from an infrastructure standpoint, but also investing in building their own capability to perform scientific AI research. Researchers were also driven by the idea of building an algorithm which can defeat humans in a game of Chess (like IBM's DeepBlue beat Gary Kasporov in 1997) or Go (like DeepMind’s AlphaGo beat Lee Sedol in 2016) or even Jeopardy! (like IBM's Watson won against few human contestants in 2011). While each of these algorithms were designed and implemented separately using different architectures and datasets, Mitchell warns against anthropomorphizing these wins. For example, saying IBM Watson beat humans at Jeopardy! should question the real intelligence of Watson in 'beating' humans at anything else other than Jeopardy! (IBM subsequently tried to extend Watson to consume millions of books on medical literature in order for it to be able to diagnose cancer. It failed miserably (Strickland, 2019))

The way human beings learn new concepts is inherently different from the way a machine learns the same concepts. We ask questions and demand more information when we do not understand something. Children, for instance, are inherently curious about their environment and actively explore new things. They also learn to categorize or generalize very quickly, just by looking at or learning from a few examples or apply their learnings from one area to another (in machine learning terminology this is called transfer learning). To indicate transfer learning in human beings, Mitchell cites examples like that of an adult who can drive a car in one city can drive a car in another city very easily. Or the way we use and understand metaphors, for instance ‘falling in love’ does not mean physically falling down. Comparatively, machines take longer to generalize like us and need millions of data points in order to effectively categorize objects (The PVOC competition previously discussed, had 3 million images to learn from). Plus, the process of tuning parameters and hyperparameters needed for these machine learning algorithms is something akin to alchemy. Goes to justify the range of salaries some of the AI research magicians draw, not just at companies like Google or Facebook, but also at non-profits like OpenAI (Metz, 2018). With this context, Mitchell encourages AI researchers to focus on building understanding or meaning-making within AI. Some of the references on meaning-making, generalizing and transfer learning that Mitchell alludes to are in the area of dog-walking. When we encounter the term dog-walking, pictures of people out for a walk or a jog in the park in the morning, with a dog on a leash come to mind. But if we encounter a dog-walker on a bicycle (human being riding a bicycle while walking the dog on a leash), we can still intuitively correlate the image to dog-walking. Unfortunately, AI algorithms do not generalize these concepts well and may not recognize this activity as dog-walking, unless it has previously been fed images with a dog-walker on a bicycle. Human beings learn slowly and most of our intuition is based on making connections and analogies. Citing her own work in symbolic AI research during her PhD, where she built a program called 'Copycat' to build analogies given only a few data points, she emphasizes on the need for AI algorithms to create abstract concepts to build something akin to common sense knowledge and build upon analogy oriented symbolic AI approaches in the future. Some of the other interesting initiatives she cited in her book include Doug Lenat's 'Cyc' project. This project aims to create an encyclopedia of all unwritten common-sense knowledge that humans possess. While Mitchell seems doubtful of this long running initiative to be able to encode all human knowledge, who is to say Cyc will not have its own 'Spring' when AI machines start aiming for meaning-making?

While companies focus on the domains where AI can be applied, very little is being done, on the topic of AI regulation and ethics, and that too only in large companies. Giving some contemporary examples on the potential misuse of AI, like Amazon which markets its face-recognition solution to police departments in the United States or Face First which provides a service which applies face-recognition to alert shops in malls when either litigious individuals or high valued customers arrive at their premises, Mitchell highlights some of the ethical concerns researchers should bear in mind while building such solutions. Historically AI algorithms have been biased against minority groups because the datasets fed into training these algorithms lack any sufficient information representative of such groups. Mitchell believe that solutions like face-recognition systems, especially if employed for law enforcement, can have a negative impact on minority groups if unregulated. And that this regulation should not be left just to governments but should be based on the cooperation of governments with multi-national companies, various non-profit think tanks and universities who are contributing in different ways towards research in AI.

While this book proves to be a comprehensive text on AI research, the only aspect that seemed to be missing was that of the promising set of researchers in the area of genetic algorithms who aim to develop AI algorithms based on the concept of evolution. These algorithms are not very popular because of the amount of time and resources needed to train them, but recently they have shown promise because of the idea of open-endedness in their learning (as compared to Cybernetics focused, feedback-based learning methods currently in vogue). It is surprising that Mitchell only mentions genetic algorithms in the book in passing, as she also authored a book on the topic few years back. Also, 'AI Spring' which Mitchell lushly describes in the book ignores the fact that the research on AI would not have achieved its current velocity if the computer hardware industry (especially the graphical processing units from the gaming industry) had not made its own progress on innovation. Applied AI, even for language translation or self-driving cars, would not have been possible if not for the great leaps made in innovating faster processing units for computers. One would think companies like NVIDIA would be referred to in the same breath as Google or Facebook, but surprisingly these innovations are hardly mentioned. Then again, she did mention in the preface that this was not supposed to be an AI history book.

Towards the end of the book, Mitchell goes back to some of the main themes of the book to answer questions on AGI. She believes that an AI ‘superintelligence’ should be the least of humankind’s worries right now and that we should be more worried about giving too much autonomy to AI systems without understanding its limitations. Deep fakes or unreliable face-recognition systems can cause much harm to us in the near term than a fantastical malevolent AI of the future.

I highly recommend this book to anyone who is interested in learning about AI. If one is new to the field (or is an expert) and wants to build a breadth of understanding in AI research, this book will prove to be a great starting point. Mitchell’s lucid prose and examples make for a great reading on a complex topic that is artificial intelligence.

References

Metz, C. (2018, April 19). A.I. Researchers Are Making More Than $1 Million, Even at a Nonprofit. *The New York Times*. https://www.nytimes.com/2018/04/19/technology/artificial-intelligence-salaries-openai.html

Strickland, E. (2019, April 2). *How IBM Watson Overpromised and Underdelivered on AI Health Care—IEEE Spectrum*. IEEE Spectrum: Technology, Engineering, and Science News. https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care

The Future of Jobs Report. (2018). World Economic Forum. http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

Purnima Kamath
National University of Singapore