

Possible minds of AI: how about energy?

Indrė Žliobaitė
University of Helsinki, Finland

July 8, 2019

Possible minds 25 ways looking at AI (2019) edited by John Brockman is a collection of essays communicated by leading experts speculating about the future of artificial intelligence and the human society. The editor, a highly accomplished literary agent, has been hanging around the elite of arts and science since the 60s, hearing news and views, sharing gossips and unanswered questions.

Many of the essays revisit the provocative text of Norbert Wiener -- *The Human Use of Human Beings* from nearly 70 years ago, where he argues about the benefits and risks of automation to society. Several essays juggle rapid acceleration of computing capacity and rapid growth of data, yet emphasizing that the societal context of AI mostly remains like Norbert Wiener painted it. Far removed from discussing anthropomorphic cyborgs or ever advancing hardware solutions, the book is essentially about how the mind works, artificial or real. I enjoyed the reading.

The essays are sequenced and interleaved by intros and memory bits from the editor such that the collection almost reads like a monograph. The themes stream from machines, via human minds, to dissidents and forms of control in the society, and back to AI and the future of civilization. The closing essay by Stephen Wolfram is a bit disappointing though, jumping too much and fading away, perhaps since the text was transcribed from an interview.

The motif of goal formulation for AI comes about often in the book. Even more so comes about value alignment between AI and humans. A suggestive resolution emerges from multiple essays – could the way forward be not to provide explicit learning objectives for AI, but rather have the objectives to be learned from examples? This is implied to be similar to reinforcement learning in reverse, or, perhaps, meta-reinforcement learning – from observations of rewards AI would infer the objectives of the rewarder. An initial research in this direction is already available (Leike et al 2018).

While the experts do not realistically expect AI to take control over the world any time soon, at least not in any evil way, the perception of continued progress is in the air. So much and at the same time so little in terms of intelligence can be accomplished by AI already. DeepBlue, AlphaGo or sorting out cat images come about often to emphasize narrowly specialized tasks that AI has already mastered. Tasks that require creativity, conceptual novelty and curiosity, despite robot scientists every now and again making to the news (King et al 2018), are still far out of reach.

Possible minds delivers a contemporary summary of where we stand today in machine intelligence. I strongly missed one perspective, however, that is – how about energy? The natural world is all about acquiring and controlling expansive energy (energy for growth and reproduction), this holds for any organism. How would machine intelligence be different? For any autonomous thinking, friendly or evil, superintelligent AI will need energy no matter the technology. And the more control AI would assume, the more thinking AI would need to do for us or not for us, the more energy will be needed.

Possible minds is about possibilities, not about feasibilities. Most of the future scenarios in the book imply that intelligent thinking carries no energy costs or they are negligible. An exception is George Church's essay (2019) where he notes that Watson, winner of Jeopardy!, used 85 000 wats of energy in real time, while the human brains were using 20 watts each -- three orders of magnitude more energy. A recent study points out that the carbon footprint of training a single deep learning model may be larger than that of several passenger cars combined during their lifetime (Strubell et al 2019). The humanity today heavily relies on energy reserves. Humans annually consume about one third of all the net primary produced energy on land, shared among all the species, plus around double that amount from burning fossil fuels (Barnosky 2015). Two-thirds of human energy consumption comes from fossil fuels that were deposited millions of years ago. Those are deposits, not renewables. If anything, superintelligent AI for starters would need to find ways to autonomously access and control a lot of extra energy! How else?

References

John Brockman (ed.) (2019). Possible Minds: Twenty-Five Ways of Looking at AI. Penguin Pres.

Norbert Wiener (1950). The Human Use of Human Beings. The Riverside Press.

Jan Leike, David Krueger, Tom Everitt, Miljan Martic, Vishal Maini, Shane Legg (2018). Scalable agent alignment via reward modeling: a research direction. arXiv:1811.07871v1 [cs.LG].

Ross D. King, Vlad Schuler Costa, Chris Mellingwood, and Larisa N. Soldatova (2018). Automating sciences: Philosophical and social dimensions. IEEE Technology and Society Magazine, p. 40-46.

George Church (2019). "The Rights of Machines" (Chapter 23). Possible Minds: Twenty-Five Ways of Looking at AI. Penguin Pres.

Emma Strubell, Ananya Ganesh, Andrew McCallum (2019). Energy and Policy Considerations for Deep Learning in NLP. arXiv:1906.02243v1 [cs.CL].

Anthony Barnosky (2015). Transforming the global energy system is required to avoid the sixth mass extinction. MRS Energy & Sustainability: A Review Journal, p. 1-13.