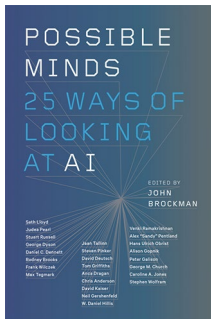# AI minds need to think about energy constraints

Possible Minds: 25 Ways of Looking at AI

Edited by John Brockman

PENGUIN: 2019.
320 PP. £23.99

Possible Minds: 25 Ways Looking at AI is a collection of essays communicated by leading experts who speculate about the future of artificial intelligence and human society. The editor John Brockman, a highly accomplished literary agent, has mingled with the elites of the arts and sciences since the 1960s, gathering news and views, sharing gossip and highlighting unanswered questions.

The essays are sequenced and interleaved by intros and various anecdotes from the editor, such that the collection almost reads like a monograph. The themes stream from machines, via human minds, to forms of dissidence and control in society, and then back to AI and the future of civilization. The book stays clear from discussing anthropomorphic cyborgs or ever-advancing hardware solutions; it is essentially about how the mind works, artificial or real, and makes for an enjoyable read.

Many of the essays revisit the provocative text of Norbert Wiener from nearly 70 years ago, *The Human Use of Human Beings*, where he discusses the benefits and risks of automation for society[1]. Several authors mention recent advances — the rapid acceleration of computing capacity and growth of data — but at the same time emphasize that the societal context of AI mostly remains as Norbert Wiener painted it.

The problem of goal formulation for AI is a motif throughout the book, especially so in the context of value alignment between humans and AI. A suggestive resolution emerges from multiple essays — instead of providing explicit learning objectives for AI, could the way forward be to learn these objectives from examples? This is implied to be similar to reinforcement learning in reverse, or, perhaps, meta-reinforcement learning: an AI could infer the objectives of the rewarder by observing their rewards. Some initial research in this direction is already available[2].

While the experts of *Possible Minds* do not realistically expect AI to control the world anytime soon — at least not in any malicious sense — the perception of continued progress is in the air. So much (and at the same time, so little) progress towards intelligence has already been achieved by AI. DeepBlue, AlphaGo and classifying cat images are often emphasized as tasks that are sufficiently mastered, but narrowly specialized. Despite robot scientists making the news[3], tasks that require creativity, conceptual novelty and curiosity are still out of reach.

Possible Minds delivers a contemporary summary of where machine intelligence stands. There is, however, one perspective that I found absent throughout the book: that of energy. All organisms in the natural world are constantly striving to acquire and control expansive energy for growth and reproduction. Why would machine intelligence be different? Any autonomous, superintelligent AI — either friendly or malicious — would need energy, regardless of the technology used.

Most of the future scenarios in the book imply that intelligent thinking carries no energy cost, or that these costs are negligible. An exception is George Church's essay (chapter 23), where he notes that Watson — winner of *Jeopardy!* — used 85,000 W of energy in real time, while the human brains it was competing against were using 20 W each — three orders of magnitude more energy. A recent study found that the carbon footprint of training a single deep learning model may be larger than the entire lifetime of several passenger cars[4].

Humanity today is heavily reliant on energy reserves, consuming annually about one third of all the net primary energy produced on land and shared among all species. In addition, we consume around double that amount from burning fossil fuels[5], a non-renewable source. When considering 'possible minds' of the future one of the main questions should be: how on Earth will a superintelligent AI find ways to autonomously access and control enormous amounts of extra energy? This remains unanswered in *Possible Minds*. ❏

Reviewed by Indrė Žliobaitė
University of Helsinki, Helsinki, Finland
e-mail: indre.zliobaite@helsinki.fi

References
1. Wiener, N. *The Human Use of Human Beings* (Riverside, 1950).
2. Leike, J. et al. Preprint at https://arxiv.org/abs/1811.07871 (2018).
3. King, R. D., Schuler Costa, V., Mellingwood, C. & Soldatova, L. N. *IEEE Technol. Soc. Mag.* **37**, 40–46 (March, 2018).
4. Strubell, E., Ganesh, A. & McCallum, A. Preprint at https://arxiv.org/abs/1906.02243 (2019).
5. Barnosky, A. *MRS Energy Sustain.* **2**, E10 (2015).