Calculemus!
Robot scientists and the mechanization of scientific reasoning?

Samuli Reijula & Jaakko Kuorikoski
University of Helsinki / TINT

Abstract

We provide a conceptual mapping of recent arguments for the demise of the human scientist due to advances in machine learning, new forms of data, and the automatization of physical experimentation. We do this by considering scientific research as heuristic search in a multidimensional problem space and focus on the role of AI in justification, in contrast to heuristic discovery. We distinguish between four arguments for the mechanization of scientific reasoning based on different technologies (algorithmic analysis of big-data, deep learning, high-throughput experimentation, and adaptive experimentation guided by open-ended search), and highlight their differences in a contrastive manner. We explore which of the interpretations of the discovery-justification distinction still stand and what capabilities are still missing from the full Leibnizian dream of mechanization of scientific reasoning.

Our approach

- AI and experiments conducted by robots seem to blur the distinction between the contexts of discovery and justification
- We view scientific research as search in a high-dimensional space
- AI not merely as a heuristic tool for discovery, but as a source of justification
- Arguments in favor of mechanized reasoning motivated both by claiming that human reasoning restricts search
  - ... too much (prior "theory" as a fundamentally irrational bias)
  - ... and too little (human experimentation as an inefficient means for conducting trial-and-error search)

Four arguments for the mechanization of scientific reasoning

1. Algorithms & big data →
- Detecting all patterns in data, no need for theory, no need for ampliative inference
- "End of the scientific method as we know it" (Anderson 2008):

2. Deep learning
- The procedure (e.g., processing data by CNNs) transforms the search space in an autonomous manner
- Refutes Hempel's (1985) argument against computerized generation of scientific hypotheses (a computer cannot generate new concepts)

3. Robolabs & high-throughput experiments
- Algorithm in charge of the creation of new data points by running a series of automated experiments
  - E.g., drug discovery, materials science
- No need for human hypothesis generation

4. Open-ended search & robolab
- Open-ended search algorithms driven by "novelty" or "curiosity"
- Capable of modifying the search procedure itself (Lavin et al. 2021)
- Closed-loop cycle of experiment design, execution, and learning

Counter-argument

- Reality ≠ Data
- The whole "ground truth" would be impractical and of little intellectual value

- Observational - not generated - data
- Patterns only, no modal knowledge

- Limited creativity:
  - Modifying parameters in an existing setup
  - Objective-based optimization

Literature

Anderson, C. 2008. "The end of theory: The data deluge makes the scientific method obsolete", *Wired magazine*, 16(7), 16-07.

Coutant, A., et al. 2019. "Closed-loop cycles of experiment design, execution, and learning accelerate systems biology model development in yeast", *Proceedings of the National Academy of Sciences*, 116(36), 18142-18147.

Giza, P. 2022. "Automated Discovery Systems, part 2: New developments, current issues, and philosophical lessons in machine learning and data science", *Philosophy Compass*, 17(1), e12802.

Glymour 2004. "Automation of discovery", *Daedalus*, 69-77

Hempel C.G., 1985, "Thoughts on the Limitations of Discovery by Computer", in K. F. Schaffner (ed.), *Logic of Discovery and Diagnosis in Medicine*, 115–122.

Krenn, M., Pollice, R., Guo, S.Y. et al. 2022. "On scientific understanding with artificial intelligence." *Nat Rev Phys* 4, 761–769.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. 1987. *Scientific Discovery: Computational explorations of the creative processes*. MIT press.

Lavin, A. et al.. 2021. "Simulation intelligence: Towards a new generation of scientific methods." arXiv preprint arXiv:2112.03235.

Olivetti, E. A. et al. 2020. "Data-driven materials research enabled by natural language processing and information extraction." *Appl. Phys. Rev.* 7, 041317.

Pearl, J., & Mackenzie, D. 2018. *The Book of Why: The new science of cause and effect*. Basic books.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Hassabis, D. 2020. "Improved protein structure prediction using potentials from deep learning." *Nature,* 577(7792), 706-710.
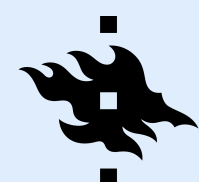
Suchow, J. W., & Griffiths, T. E. 2016. "Rethinking experiment design as algorithm design." In 30th *Conference on Neural Information Processing Systems* (NIPS 2016).

Contact information

samuli.reijula@helsinki.fi, @samulipo

jaakko.kuorikoski@helsinki.fi

TiNT

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI