On Finding Good Experiments (in astronomy)

Cheng Soon Ong

30 Sep 2025

Abstract

One of the key choices we have as scientists is to design informative experiments. Using examples from astronomy, we discuss a machine learning perspective of the scientific method. With AI promising accurate predictions, we revisit the question of adaptively designing new measurements that take previous data into account. Following this scientific method framing, we illustrate some recent ideas on using machine learning to recommend experiments in genomics. The goal of this talk is to start a conversation about what it means to recommend experiments in astronomy.
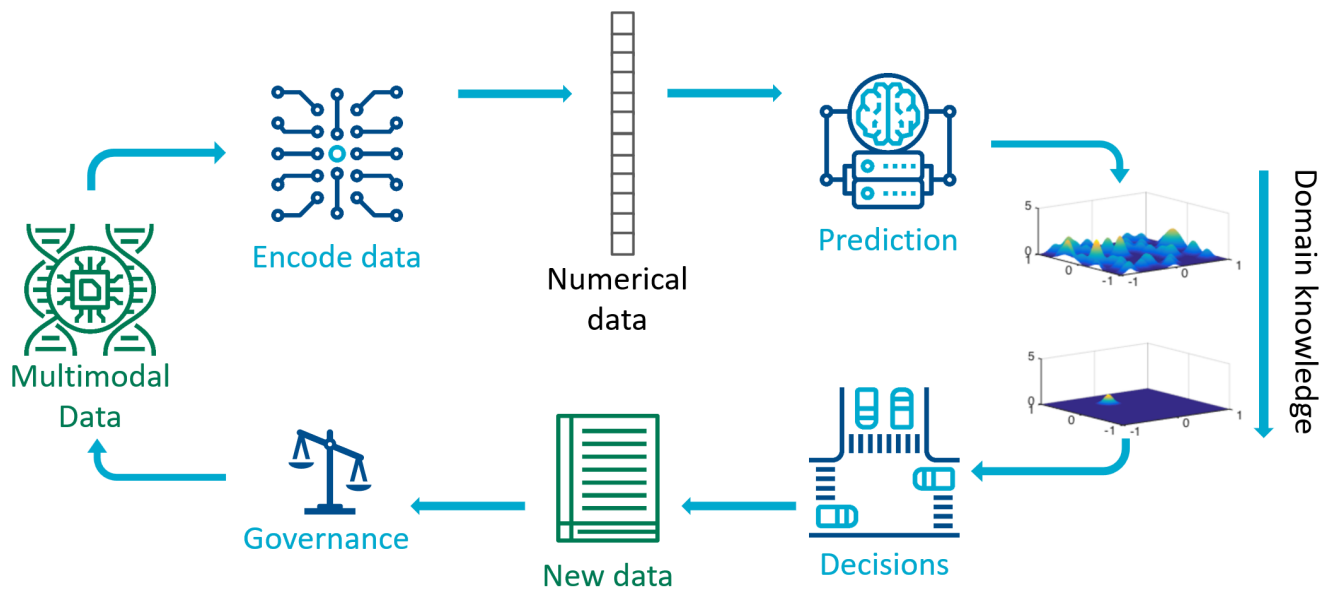
Figure 1: Machine Learning for Scientific Discovery

1

What is an experiment?

- **Encoding data.** Alger, Livingston, McClure-Griffiths, Nabaglo, Wong, Ong; Interpretable Faraday complexity classification; PASA, 2021.
https://www.cambridge.org/core/journals/publications-of-the-astronomical-society-of-australia/article/interpre
faraday-complexity-classification/EB430A90DDFB87B6B973DDF77F4A9666
- **Cross identification** Alger, M. J., Banfield, J. K., Ong, C. S., Rudnick, L., Wong, O. I., Wolf, C., Andernach, H., Norris, R. P., & Shabala, S. S.; Radio Galaxy Zoo: machine learning for radio source host galaxy cross-identification. MNRAS 2018
https://academic.oup.com/mnras/article/478/4/5547/4999919
- **Estimate uncertainty** Yong, Ong, Uncertainty quantification of the virial black hole mass with conformal prediction, MNRAS 2023.
https://academic.oup.com/mnras/article/524/2/3116/7224601
- **Train with physics simulations** Nguyen, Tang, Alger, Marchal, Muller, Ong, McClure-Griffiths; TPCNet: representation learning for HI mapping; MNRAS 2025
https://academic.oup.com/mnras/article/536/1/962/7908519
- **Photometric classification** Tran, Ong, Wolf, PeerJ Comp. Sci. 2018; Combining active learning suggestions
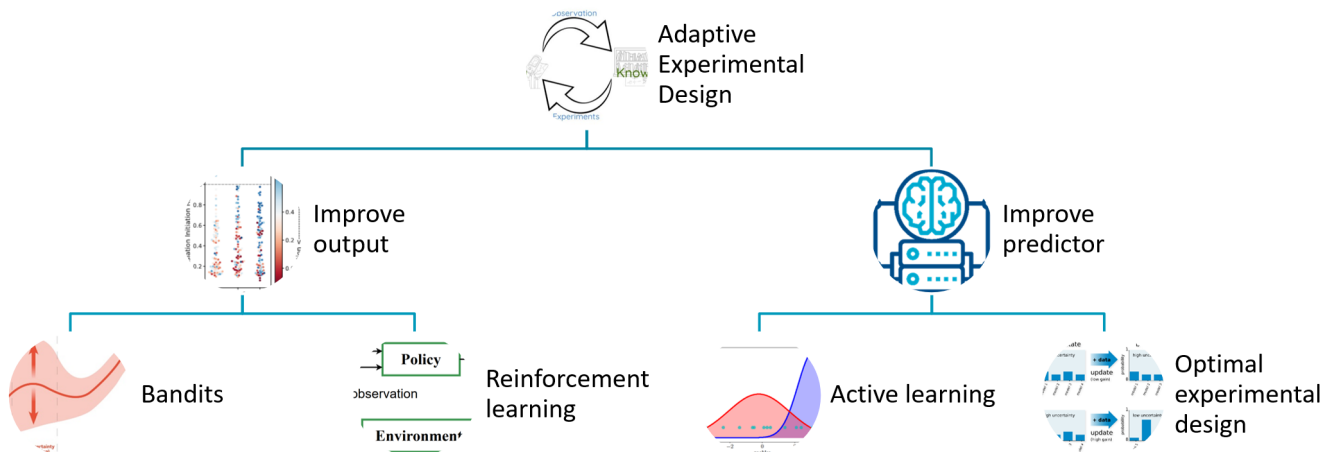https://peerj.com/articles/cs-157/

How do we find them?

Figure 2: Adaptive experimental design

- **Layperson view** Adaptive experiments: Machine learning can help scientific discovery
https://researchoutreach.org/articles/adaptive-experiments-machine-learning-help-scientific-discovery/
- **Overview of learning cycle** Blau, Chades, Ong, Machine Learning for Biological Design, Synthetic Biology, 2024
https://link.springer.com/protocol/10.1007/978-1-0716-3658-9_19

- **Ribosome binding site** Zhang, Holowko, Zumpe, Ong; Machine Learning Guided Batched Design of a Bacterial Ribosome Binding Site; ACS Syn Bio, 2022
  https://pubs.acs.org/doi/10.1021/acssynbio.2c00015
- **Islands of Fitness, generative model** Steinberg, Oliveira, Ong, Bonilla; Variational Search Distributions, ICLR 2025
  https://proceedings.iclr.cc/paper_files/paper/2025/hash/055fc19a3ce780b96cff15ffe738c1f1-Abstract-Conference.html
- **Multiobjective generative model** Steinberg, Wijesinghe, Oliveira, Koniusz, Ong, Bonilla; Amortized Active Generation of Pareto Sets, NeurIPS 2025
  https://neurips.cc/virtual/2025/poster/116473

What does good mean?

- When accurate prediction models yield harmful self-fulfilling prophecies, Patterns, 2025
  https://doi.org/10.1016/j.patter.2025.101229
- **ML and biology** Zhang, Ong; Opportunities and Challenges in Designing Genomic Sequences, 2021
  https://mengyanz.github.io/files/wcb_main.pdf
- **ML and social science** Chades, McGrath, Bohensky, Carter, Coates, Harwood, Islam, Kandanaarachchi, Ong, Reeson, Stone-Jovicich, Paris, Scowell, Wissing, Douglas; Four Compelling Reasons to Urgently Integrate AI Development with Humanities, Social and Economic Sciences; IEEE Trans on Tech and Society, 2025
  https://ieeexplore.ieee.org/document/10960694/
- **Need domain expertise** Hartman, Ong, Powles, Kuhnert; Position: We need responsible, application-driven (RAD) AI research, ICML 2025
  https://icml.cc/virtual/2025/poster/40103