Harvard Data Science Review • Issue 6.1, Winter 2024

Data Science at the Singularity

David Donoho[1]

[1]**Department of Statistics, Stanford University, Stanford, California, United States of America**

Published on: Jan 29, 2024

DOI: https://doi.org/10.1162/99608f92.b91339ef

ABSTRACT

Something fundamental to computation-based research has *really* changed in the last ten years. In certain fields, progress is simply dramatically more rapid than previously. Researchers in affected fields are living through a period of profound transformation, as the fields undergo a transition to *frictionless reproducibility* (FR). This transition markedly changes the rate of spread of ideas and practices, affects scientific mindsets and the goals of science, and erases memories of much that came before.

The emergence of FR flows from 3 data science principles that matured together after decades of work by many technologists and numerous research communities. The mature principles involve data sharing, code sharing, and competitive challenges, *however* implemented in the particularly strong form of frictionless open services.

Empirical Machine Learning is today's leading adherent field; its hidden superpower is adherence to frictionless reproducibility practices; these practices are responsible for the striking and surprising progress in AI that we see everywhere; they can be learned and adhered to by researchers in whatever research field, automatically increasing the rate of progress in each adherent field.

Keywords: reproducible computational research, challenge problems paradigm, frictionless reproducibility, frictionless research exchange, emergent superpower, AI Singularity

---
