

Man-Machine Symbiosis to Improve and Accelerate Scientific Discovery

Charles Fracchia – CEO BioBright, Boston, USA

The rise of AI in scientific research is a double-edged sword for science's reproducibility, productivity and progress. To advance scientific knowledge, AI needs to be framed by the correct incentive structures; ones that dramatically increase diverse participation and focus on augmenting human understanding. The complexity and opacity of machine learning techniques bear a very real risk of accelerating the current scientific reproducibility crisis. It is likely to make poor science indistinguishable from good, reflecting the biases of their current inceptors in the process. If we let AI be designed as a black-box, it will become a weapon, in a way akin to how disinformation on the internet can drown out validated facts.

Scientific research is currently grappling with a reproducibility crisis^{1,2}. This crisis is particularly acute in biomedical research, where as few as 11% of experiments were reproducible in a landmark pharmaceutical study.¹ This has devastating effects on the productivity of the field, leading to an average of 12 years and \$1.4 billion to get a drug to market. In the US alone, it is estimated that \$28 billion annually are spent on irreproducible pre-clinical research.³

Of the factors that contribute to the reproducibility crisis, it is clear that academic, publishing, career and financial factors create strong incentives for flashy, unconfirmed results. Digital data availability and fast computational methods are leading to a blurring of the line between *exploratory* and *confirmatory* studies. By using the wrong statistical methods and inadequately segmenting data, scientists are routinely confirming tautological hypotheses that reflect their a-priori ideas. Exploratory studies are meant to lead to the generation of hypotheses, while confirmatory studies are designed to test the validity of selected hypotheses. The data and techniques employed in each are different and reproducible science practice requires careful and deliberate separation. It is far too common that scientists use the same data set to generate their hypotheses and to confirm them. In some fields, such as neuroscience, this tendency is exacerbated by the time and complexity required to generate multiple, distinct observations and sufficient raw data. Finally, the lack of strong, ubiquitous training for statistical analysis is further aggravating the situation and fostering an element of fear, competition and lack of discussion among academics.

The democratization of machine learning approaches is only going to accelerate this trend. The complexity and non-linearity of deep networks adversely affect the human scientist's ability to understand, comprehend and communicate the effect on scientific theories.

We have been advocating for a different approach to AI and informatics in our field: one that puts the Man-Machine relationship at the center of attention. AI tools have to be *designed* to augment human intelligence and improve our understanding of the scientific problem, not to provide black-box answers. Using this approach over the last two and a half years, we – at BioBright – have created tools that have improved scientific processes by more than 20x in neuroscience⁴, 30x in an early drug discovery workflow, doubled the precision of key instruments, and automated the analysis of instrument errors with one of the largest manufacturers in the field.

- [1] Begley CG, Ellis LM (2012) Raise Standards for preclinical cancer research. *Nature* 483, pages 531–533. <https://doi.org/10.1038/483531a>
- [2] 1,500 scientists lift the lid on reproducibility. *Nature News*. <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>
- [3] Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. *PLOS Biology* 13(6): e1002165. <https://doi.org/10.1371/journal.pbio.1002165>
- [4] Richard Harris (NPR Science Correspondent), *Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions*. ISBN-10: 0465097901