Mechanized Judgment:
Artificial Intelligence and the Scientific Self

Figure 1. An image generated by the Midjourney AI that creates images from text (Prompt: "large scientific laboratory")

Garrepally Tejas

Tuesday/Thursday 2 pm - 4 pm

Artificial Intelligence (AI) can be broadly defined as intelligent systems with the ability to think and learn (Russell S. J., 2010). With many advances in the AI field, an increasing number of AI systems are replacing human judgment in various facets of research. Due to the introduction of, what I call, *mechanized judgment*, the role of human judgment in scientific inquiry cannot be taken for granted. Scientists can now use complex AI models as core tools in various stages of scientific research, reassigning decision-making activities within human-AI research teams (Yongjun et al., 2021). This paper highlights the complementarity of humans with AI and examines how each can bring its strength to scientific decision-making. While AI can supersede human judgment when addressing complexity, human researchers still offer a more comprehensive and intuitive approach to dealing with uncertainty and ambiguity. As AI systems inevitably get more intelligent, social scientists and technologists need to figure out how to allocate tasks optimally within human-AI research teams so that decision-making is more effective. With AI systems as partners for discovery, scientists are better equipped to tackle previously impossible or impractical research.

In the first section, I build on the work of Daston and Galison (2007) and introduce *mechanized judgment* as a new form of epistemic virtue. Subsequently, I highlight how AI can exercise analytical judgment in complex problem domains more efficiently and accurately than humans. Conversely, I also consider the areas in which AI is unable to supersede human judgment in scientific inquiry due to its inherent shortcomings. To conclude, I propose a framework for using AI in research contexts to guide scientific inquiry in the right direction and highlight that the purpose of AI is to augment, not replace, human judgment in expanding the frontiers of scientific knowledge.

The Birth of a New Epistemic Virtue: Mechanized Judgment

In *Objectivity,* Daston and Galison (2007) chart the emergence of scientific objectivity from the eighteenth to the twenty-first century by focusing on scientific atlases - the standard compendia of images used to train the scientific practitioners of each generation. These atlases define the 'working objects' of science, and through attentive historical analysis, one can see the epistemic virtues guiding scientific thought in them. In brief, epistemic virtues are human values that guide how we know what we know. They argue that the ways in which scientists visually conceived and presented objects of their research reflect their implicit commitments to ways of 'doing science' (p. 42).

For the authors, atlases do not just capture a representation of the object in the study; they represent a specific form of "scientific sight" (p. 18). They are visual repositories that serve to calibrate the eyes of scientists by teaching them "what to see and how to see it" (p. 44). The necessity of compiling such atlases lies in the fact that no science can work without standardized objects that guide the investigation of natural objects, which are unrefined by definition (p. 19). However, there are many ways to create such images, namely *truth-to-nature, mechanical objectivity,* and *trained judgment.* These virtues are then used to provide an account of the *scientific self,* a representation of the 'ideal' scientific inquirer as guided by certain epistemic virtues (p. 38).

Prior to the invention of the camera, the self guided by *truth-to-nature* sought to "exclude the accidental [and] eliminate the impure" (p. 59). That is to say, this self's aim was to identify the fixed forms of underlying phenomenal variation (of working objects), a process in which aesthetic judgments were necessary and the artistic self was flaunted (p. 37). In contrast, later *mechanical objectivity*, spawned by the invention of the camera, rejected such judgments as "unacceptable intrusions of subjectivity," hence its ideal were images "untouched by human hands" (p. 43). Subsequently, *trained judgment* emerged, which is the scientific attitude of the human expert who is able to interpret the raw data generated through the earlier mentioned *mechanical objectivity* in a way that identifies meaningful patterns and family resemblances in objects of study (p. 46). Across these three cases, the shifting of epistemic virtues resulted in the transformation of the self (p. 41).

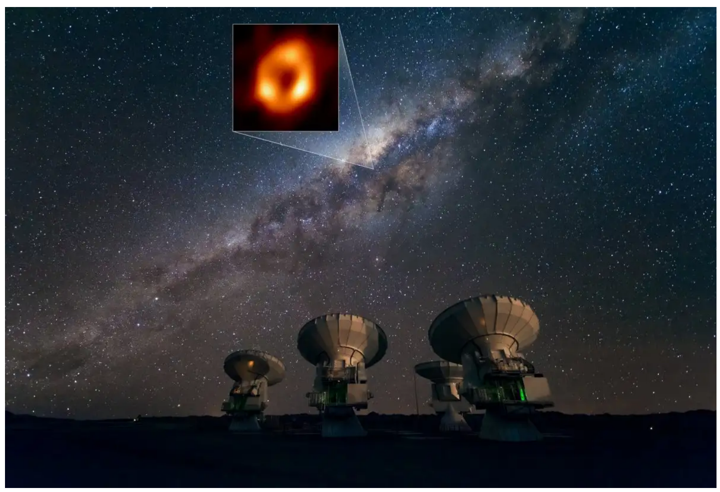

Figure 2. An image from the Event Horizon Telescope (EHT) Collaboration. Each observation campaign can record up to 15 million gigabytes of raw data across all telescopes.

As scientific problems of the 21st century are increasingly grappling with big data, I contend that AI is an inextricable part of a researcher's toolkit in aiding them to keep up with the explosive growth of experimental and observational data. Similar to how aesthetic judgments are necessary for *truth-to-nature*, mechanized judgments (AI) are necessary to tackle this new era of scientific problems given their data-laden nature (Yongjun et al., 2021). This is because Big Data brings about the same issue (Kersting & Meyer, 2018) faced by scientists in studying the mechanically objective image: data is now more "cluttered with incidental detail, compromised by artifacts [and] useless for pedagogy" (Daston and Galison, 2007: 46). Although this time, the vastness of Big Data deny scientists' recourse to trained judgment as it would be too inefficient for them to separate signal from noise to produce the "interpreted image" as compared to AI (Hey T.J. & Hooper V., 2020). Scientists can now rely on AI (the expert) for enhancing images or instrument readings to highlight a pattern or remove noise (Gil, 2017). Out of the fusion of *mechanical objectivity* and *trained judgment* is emerging a new epistemic virtue, *mechanized judgment*, one that is fit for pursuing scientific research in the 21st century.

Figure 3. The activation atlas enables scientists to reveal what patterns and features AI learns from millions of images. Image Credits: https://distill.pub/2019/activation-atlas/

The emergence of *mechanized judgment* has compelled the updating and transformation of atlases in the same way as previous epistemic virtues have. The researchers at OpenAI and Google have created an "activation atlas," which visualizes what a classification network (a type of AI) has learned from vast amounts of data (Carter, et al., 2019). Activation atlases build on feature visualization, a technique for studying what the inner layers of AI can represent. They enable us to see through the eyes of the AI, representing a new form of *scientific sight*. To truly interpret and analyze vast amounts of data, researchers must depend on “sight” provided by AI to reveal the hidden associations and patterns, a task that is even beyond the cognitive abilities of entire human research teams (Hey T.J. & Hooper V., 2020). These images no longer represent a particular object; they are refined products of calculations and intelligent algorithms. This way, the scientific images generated and classified by AI embed within them the collective ways in which scientific research was conducted and its context. To conclude, AI enables researchers to analyze massive repositories of data on a large scale, not just as a representation of the objects’ features but as a fundamental transformation of ways of knowing itself (Gil, 2017).

Figure 4. An image labeled as "fireboat" by the AI model

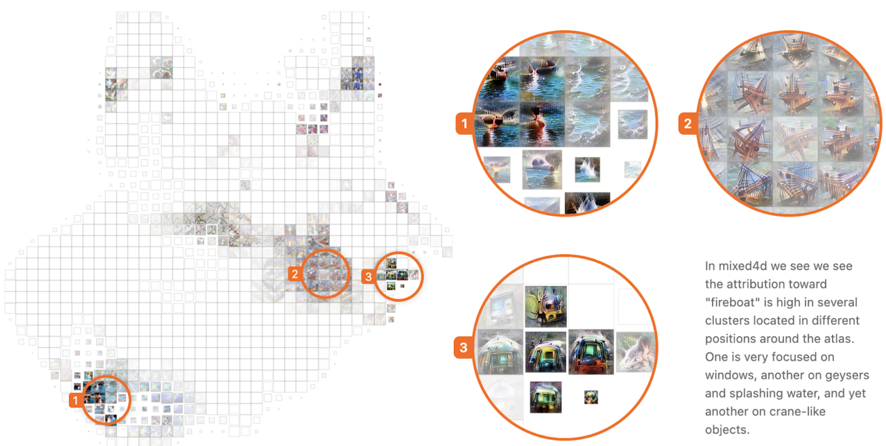

Figure 5. Features with a higher opacity represent what the AI thinks are defining features of a fireboat[1]

[1] For brevity, I have cherry-picked this example. You can explore all the layers and activations in detail in an explorable playground at https://distill.pub/2019/activation-atlas/

Complexity: The AI Advantage

The advent of Big Data has heralded an increase in the complexity of scientific problems, for which the lack of effectiveness of traditional analysis precipitates the need for AI (Yongjun et al., 2021). Areas of research that deal with vast amounts of data require information processing at speeds beyond the cognitive capabilities of even the fastest and smartest human decision-makers. As AI has the advantage of brute force, it can easily handle and analyze huge amounts of data, augmenting researchers' capabilities when dealing with complex problem domains. In this section, I hope to examine a case study and present AI as more effective in overcoming complexity in decision-making than humans, embodying an analytical and brute-force approach.

Biochemist Christian Anfinsen famously postulated that, in theory, a protein's amino acid sequence should fully determine its structure. This hypothesis sparked a five-decade quest to be able to computationally predict a protein's 3D structure as an alternative to expensive and time-consuming experimental methods. Traditional ways of predicting protein structures rely on the collaboration of biochemists and laboratory assistants to conduct experiments and analyze large amounts of data (Jumper J. et al., 2022). A major challenge, however, is that the number of ways a protein could fold before settling into its final 3D structure is astronomically high (Zwanzig R., 1992). As a result, relying on human judgment and the collaborative efforts of scientists to predict protein structures was inefficient, slow, and expensive.

In 2022, DeepMind introduced AlphaFold as a solution to this complex research problem. DeepMind trained AlphaFold using Deep Learning (a type of AI) on publicly available data consisting of ~170,000 protein structures, and as a result, the modeling accuracy of AlphaFold far exceeds that of any current methods (Figure 6). The greater computational information processing capacity and analytical approach of AlphaFold enable it to consistently deliver cheaper and high decision quality as opposed to the slower conventional methods. Evidently, AI can handle decision-making in complex research scenarios far better than humans.

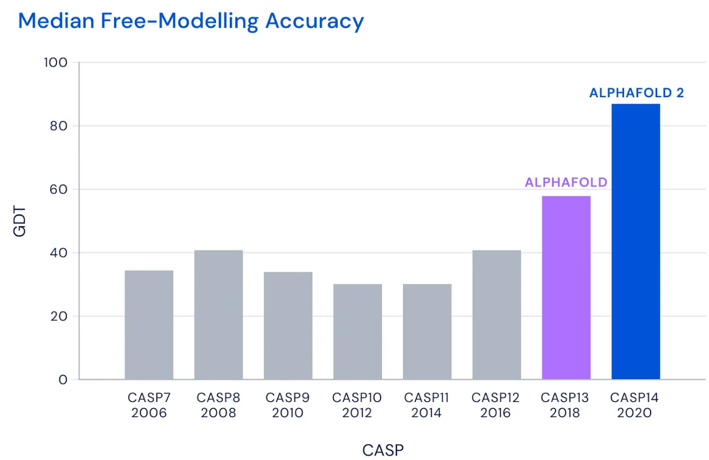

Figure 6. The AlphaFold system achieves a median score of 92.4 GDT overall across all targets in the Critical Assessment of Structure Prediction (CASP), significantly higher than other contenders.

These promising results open up the potential for scientists to use AI as a core tool in scientific research. In fact, in less than a year of its release, AlphaFold is already being used in laboratories around the world, aiding scientists in areas such as drug discovery, vaccine development, and many more[2]. Marcelo Sousa and Megan Mitchell of the University of Colorado Boulder were able to accelerate their research on antibiotic-resistant bacteria by a great deal due to AlphaFold. Understanding the 3D structure of proteins had proven to be extremely challenging and time-consuming with purely experimental methods. However, with AlphaFold, they were able to solve for structures in a matter of minutes. Likewise, Andrei Lupas of the Max Planck Institute for Biology was able to determine the structure of a protein by using AlphaFold

[2] https://unfolded.deepmind.com/

in half an hour, something that he failed to solve for 10 years (Jumper J. et al., 2022). Indeed, AI has catalyzed a huge amount of research and development in new areas that were previously impossible or limited in their scope by the cognitive abilities of humans. Undoubtedly, emerging AI systems possess an exceptional ability to make decisions when addressing complex research issues, accelerating their use in research and enabling humans to work on new problems while AI deals with complexity.

Uncertainty & Ambiguity: The Human Advantage

While AI can make probability-based decisions far more efficiently than humans, they are ill-equipped to tackle novel problems and situations (Guszcza et al., 2017). As is the case in much of cutting-edge research, ambiguity and uncertainty plague the decision-making processes of researchers since there is no precedent or preexisting data. In this section, I posit why a collaborative and intuitive style of decision-making may prove to be more useful.

AI is less capable than human researchers in uncertain or unpredictable environments, particularly outside of a predefined domain of knowledge (Brynjolfsson & McAfee, 2012). To clarify, deep learning (a subset of AI) models do indeed have the ability to generalize beyond the specific data sets they have been trained to analyze, like classifying an image that differs from the one it has "seen" before (Marcus, 2018). However, the problem is that for AI to generalize well in the first place, there needs to be a large amount of data, and the test data must be similar to the training data (p. 6). This sort of brute force approach might work very well in research contexts that have static and finite domains of science, such as speech recognition where data is categorized into a limited set of speech categories (p. 6). However, in contexts where training data is limited but the domain of study is not, many data points have yet to be encountered.

Applying AI judgment is akin to "a square peg jammed into a round hole, a crude approximation when there must be a solution elsewhere" (p. 15). On the converse, humans can learn abstract relationships and make decisions in just a few trials, through implicit means (Marcus, 2001). Indeed even 7-month-old infants can do so, acquiring learned abstract language-like rules from a small number of unlabeled examples, in just two minutes (Marcus, Vijayan, Bandi Rao, & Vishton, 1999). AI judgment, on the other hand, works best when there are thousands, if not millions of training examples, as in the previously mentioned DeepMind's AlphaFold.

Furthermore, scientists often rely on an intuitive approach in the context of cutting-edge research, leveraging insight and qualitative assessment that is rooted in years of collective experience and tacit knowledge. Researchers find it very difficult to articulate the reasons behind certain decisions beyond the fact that they just "feel right" (Galison, 2020). This inherent, unquantifiable perception that comes from within is almost impossible to replicate with AI (Parikh et al., 1994). That is to say, AI is mostly incapable of capturing the inner logic and subconscious patterns of human intuition. Consequently, AI is less likely to mimic human discernment and intuition in more abstract fields of science or unexplored sciences that involve higher levels of ambiguity and uncertainty, compelling the continued necessity of the human hand in these research contexts.

Even in this new era of *mechanized judgment*, experience, insight, and a holistic vision are and will remain capital unique to humans; internalized as subconscious and intuitive thinking processes that cannot be quantified. By this very fact, humans' unique perspectives in approaching uncertain and ambiguous research problems are irreplaceable in scientific inquiry.

It is important to note, however, that complexity, uncertainty, and ambiguity should not be seen as mutually exclusive characteristics of judgment in scientific research. As research problems are often a blend of these characteristics, it only makes sense for these issues to be handled by a blend of analytical and intuitive approaches, with both humans and AI in the loop. Furthermore, the most complex problems may still encapsulate an element of uncertainty, rendering the scientific community indispensable.

Towards Human-AI Symbiosis in Scientific Inquiry

To conclude, the complementarity of humans and AI cannot be understated, and each can bring its own strengths to scientific inquiry. Although AI systems can augment researchers' capabilities in overcoming complexity through a superior analytical approach, the role of human decision-makers and their intuition in dealing with uncertainty and ambiguity of decision-making remains unquestionable. As AI permeates various epistemes, I propose that researchers investigate the fair use of AI to ensure that the type of knowledge produced aligns with human values.

In light of the new *mechanized judgment* paradigm, I call for the collaboration of social scientists and technical experts to create guidelines and frameworks concerning the use of AI in scientific research. As the impacts of research transcend the scientific communities, it is crucial that members of all scientific disciplines agree on set practices for the use of AI. While suggestions for possible frameworks and guidelines are out of this paper's scope, I envision that the frameworks and guidelines take into account the strengths and shortcomings discussed earlier. Ideally, the frameworks should capture the nuances, standards, and practices from all involved epistemes, including the social sciences, without discounting or cherry-picking convenient frameworks. Doing so will ensure the alignment of knowledge produced across various research communities.

In the same way, Anfinsen laid out a challenge far beyond science's reach fifty years ago; there are many aspects of science that remain undiscovered. The rapid progress in the AI field and the fact that "computers plus humans do better than either one alone" (Campbell, 2016) gives mankind greater confidence that AI will become one of humanity's most useful tools in expanding the frontiers of scientific knowledge.

References

Brynjolfsson, E., & McAfee, A. (2012). Winning the race with ever-smarter machines. MIT Sloan Management Review, 53 (2), 53—60.

Carter, et al. (2019). "Activation Atlas", Distill, 2019. https://doi.org/10.23915/distill.00015

Campbell, M. (2016). “20 years later, humans still no match for computers on the Chessboard.” NPR. https://www.npr.org/sections/alltechconsidered/2016/10/24/499162905/20-years-later-humans-still-no-match-forcomputers-on-the-chessboard

Daston, L., & Galison, P. (2007). Objectivity. New York: Zone Books.

Guszcza J., et al. (2017). “Cognitive collaboration: Why humans and computers think better together.” New York, NY: Deloitte University Press.

Galison, P. (Director). (2020). The Edge of All We Know [Documentary]. Sandbox Films.

Gil, Yolanda. (2017). “Thoughtful Artificial Intelligence: Forging a New Partnership for Data Science and Scientific Discovery” ed. Michel Dumontier. *Data Science* 1(1–2): 119–29.

Hey, T.J., & Hooper, V. (2020). AI3SD Video: AI for Science: Transforming Scientific Research.

Jumper, J., Evans, R., Pritzel, A. et al. (2021). “Highly accurate protein structure prediction with AlphaFold”. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kersting, K., & Ulrich M. (2018). “From Big Data to Big Artificial Intelligence?: Algorithmic Challenges and Opportunities of Big Data.” KI - Künstliche Intelligenz 32(1): 3–8.

Marcus, G. (2018). Deep Learning: A Critical Appraisal. http://arxiv.org/abs/1801.00631

Marcus, G. F. (2001). *The Algebraic Mind: Integrating Connectionism and cognitive science*. Cambridge, Mass.: MIT Press.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven month-old infants. Science, 283(5398)(5398), 77-80.

Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc.

Xu, Yongjun et al. (2021). “Artificial Intelligence: A Powerful Paradigm for Scientific Research.” The Innovation 2(4): 100179.

Zwanzig, R., et al. (1992). Levinthal's paradox. Proceedings of the National Academy of Sciences of the United States of America, 89(1), 20–22. https://doi.org/10.1073/pnas.89.1.20