

Does AI Prediction Scale to Decision Making?

Teppo Felin and Mari Sako

Artificial intelligence (AI) excels at prediction.¹ Large language models (LLMs), for example, are remarkable at predicting the next word and stringing together fluent text. AI's predictive power extends beyond language to generate moving visuals and audio. AI models are trained with vast amounts of data and they use the statistical associations and patterns in the data to generate outputs.

But does AI's ability to predict scale to decision making? Do relatively mundane (albeit impressive) forms of prediction—like predicting the next word—extend to reasoning in novel situations and to decision making in the real world?²

Many argue that the answer is “yes.” Predictive algorithms are now widely used to make decisions in varied domains like healthcare, finance, law and marketing.³ AI models are said to not only solve problems they have encountered in their training data, but also new problems. Some argue that LLMs have an “emergent” capability to reason.⁴ And with more and more data and computational power, some argue that AI will surpass humans in *any* cognitive or decision-making task. For example, in their book *Noise: A Flaw in Human Judgment*, Kahneman and colleagues argue that “all mechanical prediction techniques, not just the most recent and more sophisticated ones, represent significant improvements on human judgment.” In short, “mechanical prediction is superior to people.”⁵ Humans are biased and computationally limited, while machines are objective and neutral. And AI's capacity for predictive judgment is said to extend well beyond mundane problem solving—specifically, to human decision making under uncertainty.⁶

We concur that the predictive capabilities of AI are remarkable. But in this column, we argue that there are limits to AI prediction. We elaborate on the nature of such limits, which in our view apply to a broad range of highly consequential real-world decisions. These decisions require the human capacity for what we call “counter-to-data” reasoning, extending beyond data-driven prediction.

The Limits of AI Prediction

The key limit of AI prediction derives from the fact that the input or “raw material” AI uses to make predictions is *past* data—the data the AI has been trained with. AI therefore is necessarily always backward-looking, as its outputs are a function of inputs. The statistical learning that forms the basis of any prediction cannot somehow bootstrap the future, particularly if it involves data that is “out of distribution,” that is, data the AI hasn't encountered before. AI prediction is based on probabilistically sampling past associations or existing correlations from its

training data, with an eye toward *likely* and probable outcomes. But past-oriented AI has no mechanism for making predictions or generating unique outputs well beyond its training data.

To offer a simple illustration of this, in one reasoning task humans and LLMs were both presented with the sequence transformation “a b c d → a b c e” and asked to apply the same transformation to “i j k l.” Both humans and LLMs could readily deduce that the analogous transformation would result in “i j k m,” by abstracting the concept of a “successor” (incrementing the last character). However, when the sequence was modified to use a permuted alphabet, LLMs frequently failed in situations that humans can easily solve. For example, if we give the following, permuted alphabet [a u c d e f g h i j k l m n o p q r s t b v w x y z] and ask humans to name the next letter after “a” they will readily say “u.” Or if we give the pattern “a u c d → a u c e” and ask for the equivalent transformation for “q r s t” → q r s ?, then humans can easily follow the prompt and respond “b.” LLMs however struggle with these types of tasks.⁷

Or to offer another example, LLMs have problems with simple word puzzles like: “Alice has 4 brothers and 1 sister. How many sisters does Alice’s brother have?” Many LLMs fail and answer that Alice’s brother has one sister.⁸ These types of problems are also illustrated by the fact that even slight changes in the wording or structure of common reasoning and problem solving tasks lead AI prediction to falter or fail completely.⁹ These failures highlight the reliance of LLMs on *previously encountered, surface-level patterns that are memorized* from training data—and the inability of LLMs to engage in novel, on-the-fly reasoning.

Other clever experiments have been developed to see whether LLMs are simply reciting what they have encountered in their training data, or whether they can actually engage in novel reasoning. For example, AI researcher Francois Chollet’s so-called “abstract reasoning corpus” (ARC) tests the ability of AI to solve *new* problems, tasks the AI has not encountered previously.¹⁰ Chollet “hides” these tasks by *not* posting them on the Internet, to ensure LLMs are not trained on them. The tasks are strikingly simple: many of them can readily be solved by a child. Chollet has even offered a one-million-dollar prize to allow anyone to submit an algorithmic model to solve these simple, “hidden” problems better than humans. But so far humans significantly outperform any algorithm—including highly sophisticated LLMs, including OpenAI’s o3 model unveiled in December 2024.¹¹ The bottom line is that not just LLMs but AI more broadly presently fail to solve novel tasks, while humans (and even children) can do so routinely.¹²

The tasks used to illustrate how AI fails to engage in reasoning are quite simple. They highlight just how out of reach more complex forms of problem solving and novel decision making can be for AI, at least for now. That said, we certainly recognize that the existing capabilities of AI—in retrieving information and writing fluent text—are remarkable. The applications of this are sure to improve and be transformative. But the claim that AI’s ability to predict translates to reasoning and decision making in novel situations is overstated. AI prediction is tightly

coupled with the data or problems it has been trained with, encountered, and essentially, memorized. AI can summarize—or generate derivative outputs from—the information it has encountered, but this does not somehow translate to solving new problems or bootstrapping new data.

AI's Data-Driven Prediction vs. Human “Counter-to-Data” Reasoning

AI is superior to humans when it comes to the processing of data and information. Algorithms are superior to humans in what Kahneman, Sibony and Sunstein call “predictive judgment,” the process of estimating an outcome based on past data, such as predicting a candidate’s future success or the likelihood of an event such as fraud based on past outcomes. Algorithms do not have preferences or values, which can introduce random variability (called “noise”). AI sticks to the data and objective facts. Importantly, this advantage of algorithms over humans is said to apply not only to recurrent decisions, but even to rare, one-of-a-kind decisions.¹³

While data is important, in many instances of decision making, data might in fact be wrong, contested, or not (yet) available. This is inherently the case for highly consequential, “forward-looking” decisions.¹⁴ In these instances, AI algorithms perform particularly poorly. And interestingly, it’s here that the very things that are seen as the “bugs” of human judgment—seeming biases, idiosyncratic preferences, and disagreement—in fact turn out to be extremely useful. To illustrate, in the context of technology startups, a recent MIT study shows that *disagreement* among human judges and experts offers the best predictor of the eventual economic value and success of a startup.¹⁵ That is, startups that are contrarian and idiosyncratic—not predictable—create the most economic value. Disagreement and idiosyncrasy are precisely what AI cannot capture well.

Many forms of AI are “autoregressive,” meaning they generate outputs sequentially by using prior data points to statistically predict future values. While this approach works well in stable environments, there are limits to applying it in evolving and uncertain ones. For instance, there is strong evidence that data-driven investors (in venture capital) that use AI—like machine learning and predictive analytics—tilt their investments towards startups that are “backward-similar” and therefore less innovative and novel, and less likely to achieve major success (like an IPO).¹⁶ *AI is great at mirroring what led to success in the past, but it is lousy at anticipating what might lead to success in the future.*

What gets lost with the emphasis placed on AI’s data-driven prediction is the human ability to engage in what might be termed “counter-to-data” reasoning. This form of reasoning takes several forms. First, humans can disagree and have different interpretations of even seemingly conclusive data and evidence. And second, humans can hypothesize and experiment to generate *new* data to prove things that might appear highly contrarian or implausible. This is the very basis of the scientific method. As scientists well know, any appeal to data is meaningless without some kind of theory; and theories tell us what might constitute relevant data and experiments. It is this logic that also should be at the heart of decision making under uncertainty.

The bottom line is that when knowledge evolves and grows—and when new data is needed—humans have an advantage over AI algorithms. Humans can disregard seemingly conclusive data, reason forward and experiment. The most consequential decisions humans make are often highly idiosyncratic.

Consider Airbnb as a brief illustration of how humans can reason counter-to-data and experiment to generate *new* (or different) data. The founders of the startup—in mid-2000s—were met with significant skepticism when they proposed to start a company to use vacant homes as an alternative to hotel accommodation. There was no data or evidence to suggest that this might be plausible. Sophisticated investors and lodging experts dismissed their idea. But the Airbnb founders believed, “counter to the data,” that if they could solve the problems of trust between strangers and create a way to efficiently match travelers, then they would be able to realize their belief.¹⁷ The founders thus ignored existing evidence about the implausibility of their idea and instead committed to realize their idea through causal reasoning and experimentation. They hypothesized that if they created a platform where people could “match” with those providing accommodation, and if they could generate trust amongst strangers (for example, with a ratings system), then their previously implausible-sounding idea might become a reality. For this, they needed to experiment to generate evidence, new data, for the plausibility of their idea.

Figure 1. Comparing AI and Human Decision Making



Figure 1 offers a visual summary of the contrast between the data-driven approach of AI versus human reasoning. To the left, AI takes a well-known bottom-up approach starting with data, leading to prediction and decision. To the right, humans do not start with all the relevant data but with “counter-to-data” reasoning, with new data generated through the process of reasoning and experimentation for a decision to be made. While some assert that AI’s ability to utilize masses of data and predict is useful for decision making under uncertainty,¹⁸ we argue that for many human decisions, forward-looking counter-to-data reasoning and new data are needed. Humans can disagree about existing data and engage in novel experimentation and problem solving to generate the evidence needed. Humans can ignore data they might disagree with and generate experiments to generate alternative evidence.

Comparative Advantages of AI and Humans

A good way to summarize the AI-human tension is to recognize their respective strengths and limitations (see Table 1), depending on the nature of decision, situation or problem that we are talking about. In some situations, the predictive abilities of AI far surpass human abilities. When problems are well-defined and recurring, and when lots of relevant data is available, using mechanistic prediction can yield useful outputs and decisions. But in other types of situations and environments, humans surpass AI. That is, when problems are open-ended, ill-defined or controversial (because they challenge accepted social norms or received wisdom), and when data is sparse or not available—or when there is disagreement on the data—relying on humans yields better results.

Importantly, many if not most AI-based tools are based on statistical averages, patterns and frequencies that are not amenable to one-off or individualized decision making under uncertainty.¹⁹ In these types of decision making, data is the eventual *outcome* of top-down factors, rather than—as with AI—the starting input (see Figure 1). In human decision making, data is the outcome of thinking in counterfactual ways, as well as intervening in the world through experimentation. These steps cannot be automated.

Table 1. Summary of Decisions and Problems Suited for AI vs Humans

	Artificial Intelligence	Humans
Types of problems	Structured, well-defined problems with clear parameters and solutions	Ill-defined, open-ended, or controversial problems requiring problem formulation
Input	Data	Counterfactual and causal reasoning
Focus	Prediction and pattern recognition	Abstract, causal reasoning
Approach	Bottom-up, data-driven	Top-down, theory-driven
Temporal focus	Backward-looking, uses general patterns from past data	Forward-looking and idiosyncratic, anticipates and plans for uncertain futures
Causal understanding	Identifies statistical relationships and correlations	Engages in causal reasoning and hypothesizing
Level of specificity	General probabilities, frequencies and averages	Individualized focus, extremes and idiosyncrasies
Novelty	Recombines known data and patterns to create variation	Generates novel data and new associations
Useful contexts	Operations, routine decisions in highly stable environments, pattern recognition	Novel decision making, strategy, idiosyncratic decisions in unpredictable environments

Conclusion

AI prediction is everywhere. Prediction certainly has its uses, particularly when abundant data are available and decisions are “predictable.” But when dealing

with uncertain and data-sparse environments—or when data is contested or not yet available—more forward-looking forms of reasoning are needed. We argue that human cognition plays a central role here. Humans can engage in counter-to-data reasoning about the plausibility of outcomes that presently lack data. It's this causal reasoning that enables humans to intervene in their surroundings and to experimentally generate new data. As a result, under certain circumstances, human decision-making involves a very different set of steps from data-driven prediction.

Such human capability is of increasing importance, as we consider how to differentiate ourselves from others in a world of easy access to various AI models. Certainly human-centered AI (when humans interact with AI as an input into human decisions) has received a lot of attention, and will undoubtedly become increasingly important. But clear thinking around the nature of the problems being addressed must precede any discussion about the nature of such interaction.

¹References

Agarwal, A., Gans, J., & Goldfarb, A. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Press, Cambridge, MA, 2022.

² McCarthy, J. Generality in artificial intelligence. *Communications of the ACM*, 30(12), 1030–1035 (1987).

³ Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M., & Moradi, S. Big data and predictive analytics: A systematic review of applications. *Artificial Intelligence Review*, 57(7), 176 (2024). Also see Sadana, U., Chenreddy, A., Delage, E., Forel, A., Frejinger, E., & Vidal, T. A survey of contextual optimization methods for decision-making under uncertainty. *European Journal of Operational Research* (2024); and Wang, A., Kapoor, S., Barocas, S., & Narayanan, A. Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy. *ACM Journal on Responsible Computing*, 1(1), 1–45 (2024).

⁴ Webb, T., Holyoak, K. J., & Lu, H. Emergent analogical reasoning in large language models. *Nature Human Behaviour*, 7(9), 1526–1541 (2023).

⁵ Kahneman, D., Sibony, O., & Sunstein, C. R. *Noise: A Flaw in Human Judgment*. Little, Brown Spark, New York, NY, 2021, page 111.

⁶ Agarwal, A., Gans, J., & Goldfarb, 2022.

⁷ Lewis, M., & Mitchell, M. Using counterfactual tasks to evaluate the generality of analogical reasoning in large language models. *arXiv preprint arXiv:2402.08955* (2024).

⁸ Nezhurina, M., Cipolina-Kun, L., Cherti, M., & Jitsev, J. Alice in Wonderland: Simple Tasks Showing Complete Reasoning Breakdown in State-Of-the-Art Large Language Models. *arXiv preprint arXiv:2406.02061* (2024).

⁹ Hong, P., Ghosal, D., Majumder, N., Aditya, S., Mihalcea, R., & Poria, S. Stuck in the quicksand of numeracy, far from AGI summit: Evaluating LLMs' mathematical competency through ontology-guided perturbations. *arXiv preprint arXiv:2401.09395* (2024).

¹⁰ Chollet, F. *On the measure of intelligence*. *arXiv preprint arXiv: 1911.01447*. (2019). Interested readers can test themselves with some of these abstract reasoning tasks here: <https://lab42.global/arc/>

¹¹ Zeff, M. and Wiggers, K. OpenAI announces new o3 models. *TechCrunch*, December 20, 2024.

¹² Goddu, M. K., & Gopnik, A. The development of human causal learning and reasoning. *Nature Reviews Psychology*, 1-21 (2024).

¹³ Kahneman, D., Sibony, O., & Sunstein, C. R., 2021.

¹⁴ Felin, T., and Holweg, M. Theory is all you need: AI, human cognition and causal reasoning. *Strategy Science*, 9(4), 346-371 (2024).

¹⁵ Gius, L. Disagreement predicts startup success: Evidence from venture competitions. *Strategy Science* (2025, forthcoming).

¹⁶ Bonelli, M. Data-driven investors. SSRN 4362173, 2023.

¹⁷ Felin, T., & Zenger, T. R. The theory-based view: Economic actors as theorists. *Strategy Science*, 2(4), 258–271 (2017).

¹⁸ Felin, T. and Holweg, M., 2024.

¹⁹ Computer scientists have recently pointed out the need to understand “individualized” decision making. For a recent recognition of this, see Jessica Hullman's blog post and associated reading list: <https://statmodeling.stat.columbia.edu/2024/12/06/new-course-prediction-for-individualized-decision-making/>