Schmidt futures: AI and mechanism design

Washington DC May 11-12, 2023

Hal Varian
Google and UC Berkeley

Disclaimer: This document consists of my views only and does not reflect views of my employer. I used the proofreading utility from Google Docs when writing the document but there were no other AIs involved in its creation.

In the *last 20 years*, economic mechanism design has moved from theory to practice, and is now being applied in a number of real world applications such as sustainability, medicine, education, housing, and spectrum allocation, to mention just a few.[1]

In the last *10 years*, artificial intelligence has made a similar move from theory to practice and is now being applied in a number of real world applications involving autonomous vehicles, machine vision, chess, go, protein folding, weather forecasting, translation, coding, and chat bots, to mention just a few.

In Silicon Valley we tend to overestimate what can be accomplished in a year but underestimate what can be accomplished in 10 years. Given the rapid and broad diffusion of AI, it may be an exception to that rule. There is a good chance that AI applications will evolve at least as fast as mechanism design did.

However, it should be remembered that a low cost of components leads to and intense competition and low product prices. So far things look good with respect to affordable innovation, since several AI innovations are free or open source.

[1] I think of "market design" as a special case of "economic mechanism design".

What is AI?

There are many definitions of artificial intelligence but a simple one is “a machine that behaves in a way that we would describe as “intelligent” if a human or animal did it.” There are several approaches to artificial intelligence, but we will consider only two approaches in this note, adversarial and generative.

Adversarial AI

An “Adversarial AI is” appropriate when you want an AI to find a winning strategy in a game. The trick is to create two AIs and have them play against each other. According to Demis Hassabis of DeepMind “If you program AI to play chess in the morning, by lunch it can beat most players, and by the evening, it's stronger than world champions.” He exaggerates a bit, but not by much.

Generative AI

“Generative AI” is appropriate when you want an AI to generate a document about a particular topic. This is typically done by using a “Large language model (LLm)” that can predict the next word in a sentence based on the previous words that the model has seen.

You start the AI running by providing a “prompt” which could would be a sentence or two about the topic of interest. The AI then looks at each word in the prompt and uses the LLm to choose the most likely word to follow. This operation is then repeated several times and often produces a coherent and meaningful sentence about the topic.

Please note that this is a vastly simplified description of how a LLm works. In actual procedure there are several important tweaks that dramatically improve performance.

Generative AI and and logical puzzles

The sentences generated by the procedure outlined above are often quite striking. Once I tried the famous fox/chicken/corn river crossing puzzle using Googles AI, Bard. Bard came back with a perfect solution--- maybe too perfect. To see where the solution came from I searched the web using Google and the same prompt I used with Bard. The result was 25 Q&N logic puzzles from Parade Magazine. Each logic problem was followed by its answer. Moral: it's easier to answer a question when the answer is posted on the web!

This is an important point: Generative AIs find documents that already exist in some textual corpus, while adversarial AIs can find novel strategies that have never been seen before.

This is not a criticism of large language models. Having more and better ways to search large collections of information and return coherent and correct responses is clearly a major advance in language processing and search.

However, as the example above shows that LLms are not quite as magical as they appear. They are like an open book exam. The answer to your question, or something close to it, is probably on the web or whatever corpus is used---the challenge is to find it. A generative AI or a search engine are both ways to search large repositories of words and both methods can be helpful, depending on what is wanted.

Finally, it is likely that AI has received a lot of attention due to its catchy name. Ask yourself how much excitement would be generated if the generative AI technology was called a “multi document query indexer with natural language responses" rather than "artificial intelligence"?

Military applications

We used to think that war was obsolete but it is now abundantly clear that is not the case. Historically innovation stimulated by wars has been limited to devices, such as rockets or radar. But in World War II there was significant progress in mathematical modeling. This made a big impression on the Pentagon and led to the establishment of RAND in 1945, right at the end of World War II.

RAND's mission was "connecting military planning with research and development decisions." Indeed two of its major successes were game theory and operations research. Mechanism design can be thought of as a merger of these two research programs.

Towards the end of the 20th century DARPA and NSF were instrumental in funding the development of the internet for academic researchers and, later on, for the general public. The vision was that the internet would contain a digital library of content that could be accessed by anyone, anywhere or any time.

It soon became clear that the digital library needed a digital card catalog that would help ordinary users find relevent content on the internet. This led DARPA and NSF to fund a digital library research program that advanced the state of the art in image and textual search and could be used for both military and civilian applications. See On the Origins of Google for details of this program.

Military and AI

DARPA played a major role in creation of the internet, but what is the role of AI in technological innovation today? Much of the research in this area is classified, but Scharre [2023] describes some of the issues that arise when humans work alongside and against AIs in both military and civilian settings. Should AIs be viewed as tools controlled by humans, or can they be thought of as co-pilots that can make decisions on their own? If they

can make independent decisions on their own, what should the scope of those decisions be?

AIs are different from humans

One of Scharre's themes is that AIs can make radically different decisions from humans. Here are a few examples of how AI decision making differs from human decision making.

First, they are hyper rational. Game theory may be a better model for modeling AI behavior than it is for modeling human behavior. This is not a new observation. Way back in 1994 I wrote a paper on "Mechanism Design for Computerized Agents" which contained this passage:

> "Game theory has been justly criticized for its "hyper-rational" view of human behavior. However, such hyper-rationality may actually be an appropriate model for software agents since they presumably have much better computational powers than human beings." Varian [1994]

My paper was only a short overview of mechanism design written for an audience of computer scientists but other researchers have investigated these issues in more depth.

Second, AI's don't make mistakes. They may be *mistaken* due to misinformation, but they don't make mistakes such as pushing the wrong button due to fatigue, or other clumsy mistakes that humans are prone to.

Third, AIs can train themselves. As mentioned previously, adversarial AI has led to solving Chess, Go and even board games such as Diplomacy, which involves strategic planning and natural language negotiations with other players. One of the most important capabilities about AIs is that they can be "trained by example". If you want a clever limerick you can write a good prompt, show it to the AI, and generate as many limericks as you want.

Fourth, AIs don't care whether they survive or not. The relevant concern, from the AI's point of view, or its owner, is only how much it costs to build a replacement. AIs possess not only artificial intelligence, but also what might be called "artificial bravery."

AIs are different from humans in many other ways. They play Chess differently, they play Go differently, they fly fighter jets differently, and so on. See Paul Scharre's op ed on AI's inhuman advantage where he says "Experts who study AlphaGo's playing style describe it as "alien," and "from an alternate dimension." Chess experts reacted much the same way.

Not all of the differences between humans and AIs operate to the advantage of an AI. As Scharre puts it "Their performance is often very brittle." For example, an AI was trained on a standard 19 x 19 Go board and performed well as expected. But when it played on a smaller and simpler 13×13 board it did very poorly. This brittleness also shows up in other contexts; here is another example from Scharre [2023]:

> A DARPA team of engineers designed a sensing device that was intended to detect soldiers sneaking into a secure area. The engineers challenged eight Marines to try to evade their system. The outcome: not a single Marine was detected by the DARPA device! Two of the Marines somersaulted for 300 meters, two hid under a cardboard box and slowly crept through the perimeter, and one of them stripped the branches from a fir tree and wore them like a coat.

The problem, of course, was that the sensing device did not use somersaults in its training set so it could not recognize the anomalous behavior as a threat. Of course, the engineers could enlarge the training set but then the Marines could come up with new ones.

Terrorism

It has been claimed that risk assessment experts were well aware of the destructive power of an airplane crash but did not think that terrorists could recruit 19 men who would committ suicide for their cause. Sadly, they were able to find such people and nowadays we recognize that automobiles, drones, planes, trucks, and semi-trucks can all be lethal weapons.

Piloting a weaponized vehicle would typically involve considerable risk. This risk is not a concern for autonomous AIs but would be a serious consideration for humans. This difference gives AIs a fundamental advantage compared to humans.

Suppose you, a human, are negotiating with someone and they say “do what I ask or I will blow us both up.” This would not be a credible threat coming from another human but could easily be plausible coming from an AI. After all, you can always quickly build another AI, another critical difference between AIs and humans.

War and Peace

I have described several examples involving military applications that could be used in conducting wars. Is there a role for AIs in *keeping* the peace?

Clausewiths occe said “war is a mere continuation of policy with other means.” That may be true but all wars end sooner or later, while policy goes on forever. So we might well ask whether AIs have a role in conducting policy?

My answer is “yes”. If adversarial activity can be analyzed using game theory then analysis, arbitration, contracts, diplomacy, negotiations, treaties, and so on can all be used to describe economic mechanisms for reaching some kind of agreement.

The difficulty is that trade agreements, treaties and diplomatic memos can be hundreds of pages long. Human beings who can understand such massive documents are not readily available. Perhaps AIs can play a positive role in making these documents easier to write, easier to search, easier to translate, easier to understand, and easier to verify compliance.

Document management may sound mundane, but the world runs on exchange of documents, in both the public and private sector. Wars can be started due to critical ambiguities in documents. Having clear, explicit, and enforceable documents can play an important role in keeping the peace.

So we find ourselves in a virtual circle: Artificial intelligence can assist in creating clear and explicit economic mechanisms. It follows that improvement in artificial intelligence can lead to an improvement in economic mechanisms, each reinforcing the other.

Earlier I mentioned the famous quote from Clausewiths : "war is a mere continuation of policy with other means." But "policy" is also a way to avoid war, and policy can be improved with better tools for document management.

Bibliography

Hal Varian [1995] "Mechanism Design for Computerized Agents" in *Proc. USENIX Workshop on Electronic Commerce,* July 11-12, 1995, New York. pdf-link.

Nico Voigtländer and Hans-Joachim Voth [2013]
"Gifts of Mars: Warfare and Europe's Early Rise to Riches,"
The Journal of Economic Perspectives,
Vol. 27, No. 4 (Fall 2013),
pp 165-186.

Paul Scharre [2023]
Four Battlegrounds: Power in the Age of Artificial Intelligence
WW Norton, New York