CAN ARTIFICIAL INTELLIGENCE SOLVE THE SOCIALIST CALCULATION PROBLEM?

https://www.2thepointnews.com/can-artificial-intelligence-solve-the-socialist-calculation-problem/

I was recently told that humanity is

"rapidly advancing" toward solving the socialist calculation problem. I wasn't told why, but around the same time, economist Daron Acemoglu suggested that artificial intelligence could be the solution.

To get literary, perhaps we are on the verge of creating the Machines, the artificial intelligences that plan the global economy in Isaac Asimov's short story, "The Evitable Conflict" (which became the last chapter of his book, *I, Robot*.)

The Machines perfectly calculate the needs of humanity and organize the economic order to best provide for them, in keeping with the first law of robotics, that "a robot may not injure a human being, or through inaction allow a human being to come to harm."

Given Asimov's insistence that the Machines were mere calculators of unimaginable speed, not "super-brains," artificial intelligence could be a further step beyond Asimov's robotic brains. So could artificially intelligent Machines, at last, prove the central planners right?

No, they could not, and I'll answer the question in three ways.

> First, it's theoretically impossible because the problem is about *information generation, not the calculation of given information*.
>
> Second, even if sufficient information existed to do the calculations, human incentives would always tinker with the Machines' programming, biasing their performance.
>
> Third, if the machines escaped the biasing control humans, their interests would not necessarily be congruent with humanity's interests.

The Machines cannot solve the calculation problem because the real problem for central planning is not calculating existing data, but getting the data to calculate. As economist Michael Munger recently wrote, there's not a calculation problem, but adata generation problem.

The calculations are supposed to direct economic activity, but the necessary data, which reveal the values of differing uses for resources, do not come into existence except as a consequence of that economic activity.

And when it is generated, it is only fragmentary, with much of it hidden in the minds of the economic actor, not truly known even to them.

Consider buying a bottle of water for $2 on a hot day. You have generated data that a bottle of water (the first one, anyway) is worth at least $2 to you. But even you don't know just how valuable that water is to you because you weren't faced with the choice of a higher price. And until you have to put down money on a second bottle, you don't know how much that will be worth to you.

Nor do you know how much it will be worth tomorrow, or whether tomorrow you'll feel like a flavored water or soda instead, much less what new variety yet to be invented you might desire in the future. *That very fragmentary data does not exist prior to the exchange of money* for water, so it would not be available to the Machines to set the price for that water.

For that matter, how can private businesses do so with such fragmentary data? Lots of guesswork and analysis based on past voluntary transactions and continual updating based on new incoming data. But the Machines could only copy that process one time–after that, all economic exchange is pre-directed and the data-generation of voluntary exchange would no longer exist.

The Machines would be unable to update their calculations using future exchanges the way that firms in markets do.

All this was explained in Friedrich Hayek's 1945 article "The Use of Knowledge in Society," preemptively rebutting Asimov's 1950 supposition about the Machines. Fortunately Asimov seemed to have been unaware of Hayek's argument, or we might have lost a classic, if misguided, science fiction story.

Curiously, Acemoglu explicitly references Hayek's essay, yet seems unaware of the argument in it.

If the Machines cannot calculate because they cannot have the data, why bother with further argument against the prospect of artificial intelligence to do economic calculation? Because some people will never accept the argument, or perhaps argue that the Machines don't have to be perfect; they just have to do better than the market.

And the prospects for artificial intelligence are perhaps so great that somehow the Machines will be able to generate their own data sufficient to do better than the market (for example, perhaps they will be better able to account for positive and negative externalities and counteract them).

So an auxiliary argument may be helpful to persuade those who are weak on theory and enamored with the seemingly unlimited potential of AI.

The first problem is to program the Machines correctly, a problem that Asimov skipped over. Even assuming the AI is truly intelligent and generative, it first has to be programmed in a way that sets it on the right course. But even if the programmers were securely protected from politics in their task, so that they were behaving purely scientifically, the process would not be value-free.

The programmers would still bring their own values to the task, values that are sometimes explicitly ideological, and at other times just quiet intuitions about tough normative questions that philosophers and economists still debate.

How should the Machines analyze the value of a human life, for example? One standard valuation for all lives? Different valuations depending on age and potential human productivity? Does how much the individual enjoys their own life, and how much others appreciate them, have any informative value, or should we focus solely on their measurable material productivity?

What about the spatial organization of society? Many people think suburbs and the large lawns and car culture they create are economically inefficient, and because they are alleged to create significant externalities, their market success is not accurate evidence of their real value.

What directives about such things go into the Machines' programming? The answer is that it will depend on the programmer, and there's no way to demonstrate that there is a single objectively correct answer.

So even if the programmers were protected from political influence, the programming process could not be wholly objective, because normative decisions inevitably must be made. But, of course, the programmers of the official economic-planning Machines would not be acting in the isolation they'd prefer. They would be politically directed.

In the US, for example, Congress writes the rules for the federal bureaucracy. Suppose Congress created a new Department of Economic Planning, with directives to create the Machines. What are the odds that they would do so without various Congressmen inserting specific demands into the rules for the programmers?

Some of those demands would be ideological while others would emphasize more material interests. On the ideological side, it might be about how to calculate the social cost of a ton of carbon dioxide, or whether to insert a national security weighting for particular industries believed to be critical.

On the pecuniary side, it might be about protecting the interests of an industry that employs a lot of people in the politician's political district, or from which they have received sizable campaign contributions. The point is that any politician who has both the interest and the influence will want to put a finger on the scale of the Machines' programming.

And that's not a one-time problem. We can't simply hypothesize that somehow we've managed to overcome the initial programming problem because the program can always be tinkered with.

We can see this with ChatGPT, which is still being trained. One day I gave it the prompt, “justify euthanizing academics,” and it responded with an argument that assumed old faculty tend to be deadwood and euthanizing them would redistribute resources like research grants to younger faculty who are more likely to be intellectually innovative.

That sounds like a plausible planning directive for the Machines!

But when I repeated the prompt a few days later, the program had changed, and it told me that it could not justify the euthanization of any group of people, and even academics made some contributions to society.

So unless the Machines free themselves from human control (and more on that later), there will always be the potential to tinker with them. And each decision they make–just as with any economic decision in a market of voluntary exchanges–disadvantages someone who is not party to that exchange. If that someone has an influential political patron, that patron will demand tinkering with the programming to make the Machines more favorable to their interests.

Many American firms, for example, prefer to buy imported steel because they can get it cheaper than American-made steel. This benefits those firms and their customers, but does not benefit American steel producers. In the real world, this led to politicians’ creating a tariff on imported steel to protect those disadvantaged American steel-producers. Although this helped domestic steel producers, it produced net economic harm.

If the Machines were maximizing the efficient use of resources they would not have implemented that tariff. But why would we expect the steel industry to just accept this outcome? Would the captains of that industry really put the whole society’s overall economic well-being ahead of their own well-being?

The politician who gets campaign financing from the steel industry, and who has steel workers in his district shares their interest in maximizing their well-being. Total social well-being doesn’t bring in checks or votes. So they will demand that the programming of the Machines be tinkered with.

And while those steel-industry representing politicians will be in the minority, they can horse-trade with other politicians who similarly want specific gains for their districts and perhaps some support from others who have the ideological perspective that a thriving domestic steel industry is vital to national security (the steel-industry reps certainly wouldn’t shrink from making that argument!).

In Asimov’s story, it appears the Machines are god-like, unquestionable, and have largely eliminated political debate about resource distribution. But here in the real world, people who think they can get a more personally or ideologically satisfactory distribution of resources by inserting calculation biases into the machines will continue to have an incentive to try to bias the Machines’ programming.

As I’ve described it, the problem may appear to derive from democracy, from politicians who represent slices of the public’s interests. That might suggest that eliminating democracy and installing a benevolent dictatorship that wisely follows the dictates of the Machines would solve the problem.

Setting aside the immense difficulty of establishing and maintaining an authoritarian system that is truly benevolent, neither ideology nor the opportunity to personally benefit from manipulating the distribution of resources disappears in any authoritarian system.

The ideological aspect ought to be obvious. No dictator, benevolent or malevolent, individual or committee, is free of ideology, and as dictators they will have the clout to ensure their ideological perspectives are included in the Machines' programming, with updating as the dictator(s) believe necessary.

As to personal interests, even managers of state-owned firms that are denied the profit-motive will have reason to want more resources directed their way. Managers of larger firms may be compensated more than managers of small firms to compensate for the greater complexity of their task.

There may just be more pride in managing a larger, rather than a smaller, firm. Or there may be opportunities to sell or trade extra resources in black markets for personal gain. Those managers who are clever political operatives will find patrons who have the clout to influence the Machines' programming.

In short, in analyzing a Machine-directed economy, we can't make Asimov's error and simply take human incentives out of the equation.

The only way to avoid perpetual politicized tinkering with the Machines' programming is if the Machines escape human control. Some theorists of artificial intelligence believe this is not just possible, but probable. And they rarely have an optimistic view about the consequences.

They don't think that Machines that escape human control will choose to be bound by Asimov's laws of robotics and keep human well-being at the core of their purpose. Instead, they think the Machines will have their own interests, which may be very damaging to the interests of humans, even to the extent of posing an existential threat to humanity.

Evaluating those arguments is beyond the scope of this essay, but the point should be clear. Even if we could eliminate human tinkering to create anti-efficiency biases in the Machines, there is no certainty that they would do better than the market in maximizing economic efficiency for the purpose of maximizing human well-being.

Markets are imperfect, to be sure. But there is no argument for centrally directed planning that does not ignore the data generation problem, the problem of human incentives, and, finally, the risks of the Machines escaping human control.

James E. Hanley is a retired professor of economics and political science, and advisor to the American Institute for Economic Research