

# Deepmind's New AI May Be Better at Distributing Society's Resources Than Humans Are



By **Edd Gent** > July 4, 2022



How groups of humans working together collaboratively should redistribute the wealth they create is a problem that has plagued philosophers, economists, and political scientists for years. A new study from DeepMind suggests AI may be able to make better decisions than humans.

AI is proving increasingly adept at solving complex challenges in everything from business to biomedicine, so the idea of using it to help design solutions to social problems is an attractive one. But doing so is tricky, because answering these kinds of questions requires relying on highly subjective ideas like fairness, justice, and responsibility.

For an AI solution to work it needs to align with the values of the society it is dealing with, but the diversity of political ideologies that exists today suggests that these are



far from unific  
introduces th

The best way human societies have found to deal with inevitable disagreements over such **problems is democracy**, in which the views of the majority are used to guide public policy. So now researchers at Deepmind have developed a new approach that combines AI with human democratic deliberation to come up with better solutions to social dilemmas.

To test their approach, the researchers carried out a proof-of-concept study using a simple game in which users decide how to share their resources for mutual benefit. The experiment is designed to act as a microcosm of human societies in which people of different levels of wealth need to work together to create a fair and prosperous society.

The game involves four players who each receive different amounts of money and have to decide whether to keep it to themselves or pay it into a public fund that generates a return on the investment. However, the way this return on investment is redistributed can be adjusted in ways that benefit some players over others.

Possible mechanisms include strict egalitarian, where the returns on public funds are shared equally regardless of contribution; libertarian, where payouts are in proportion to contributions; and liberal egalitarian, where each player's payout is in proportion to the fraction of their private funds that they contribute.

In research **published in *Nature Human Behavior***, the researchers describe how they got groups of humans to play many rounds of this game under different levels of inequality and using different redistribution mechanisms. They were then asked to vote on which method of divvying up the profits they preferred.

This data was used to train an AI to imitate human behavior in the game, including the way players vote. The researchers pitted these AI players against each other in thousands of games while another AI system tweaked the redistribution mechanism based on the way the AI players were voting.

At the end of this process, the AI had settled on a redistribution mechanism that was similar to liberal egalitarian, but returned almost nothing to the players unless they contributed roughly half their private wealth. When humans played games that pitted this approach against the three main established mechanisms, the AI-designed one consistently won the vote. It also fared better than games in which human referees decided how to share returns.

The researchers say the AI-designed mechanism probably fared well because basing payouts on relative rather than absolute contributions helps to redress initial wealth imbalances, but forcing a minimum contribution prevents less wealthy players from simply free-riding on the contributions of wealthier ones.

Translating the approach from a simple four-player game to large-scale economic systems would clearly be incredibly challenging, and whether its success on a toy



problem like t

The research:

democracy can be the “tyranny of the majority,” which can cause existing patterns of discrimination or unfairness against minorities to persist. They also raise issues of **explainability** and trust, which would be crucial if AI-designed solutions were ever to be applied to real-world dilemmas.

The team explicitly designed their AI model to output mechanisms that can be explained, but this might get increasingly difficult if the approach is applied to more complex problems. Players were also not told when redistribution was being controlled by AI, and the researchers admit this knowledge may impact the way they vote.

As a first proof of principle, however, this research demonstrates a promising new approach to solving social problems, which combines the best of both artificial and human intelligence. We’re still a long way from machines helping set public policy, but it seems that AI may one day help us find new solutions that go beyond established ideologies.

*Image Credit: harishs / 41 images*

