

# Why does AI beat humans at the strategy game Diplomacy?

August 15 2024, by Stephanie Lee

---



Credit: Pixabay/CC0 Public Domain

In Diplomacy, a strategy game set on the eve of World War I, success

hinges not on luck but the ability to negotiate. Players, each representing the armed forces of a European superpower, spend much of their time building trust, forming alliances, and ultimately betraying opponents to gain the most territory. "The most skillful negotiator will climb to victory," said the game's maker, the company Avalon Hill.

So, in 2022, when an AI model competed in an online Diplomacy league and dominated [human](#) players across 40 games, it seemed to suggest a kind of computer mastery over human-like communication.

## **A closer look at AI players**

But appearances can be deceiving. A new study from researchers at USC Viterbi's Information Sciences Institute, the University of Maryland, Princeton University, and the University of Sydney sheds light on how CICERO, the Meta-developed AI model, pulls off its Diplomacy wins. They stem, the study found, more from the model's strategy prowess than communication skills. The latter still lags behind those of human players.

The findings, presented at the [62nd Annual Meeting of the Association for Computational Linguistics](#) (ACL), could lead to a better understanding of AI's ability to communicate and strategize with humans, not just during board [game](#) night but for everyday problems.

"We're studying this because we care about modeling AI-human communication," said Jonathan May, a research associate professor at the USC Viterbi School of Engineering and co-author of the study. "An important and difficult question is: How much deception is the AI model doing?"

## **Decoding communication**

The researchers set up a series of Diplomacy games, pitting CICERO against human players. Over 24 games and 200 hours of competition, they collected more than 27,000 messages. Unlike previous studies, their focus shifted from CICERO's impressive win rate to a more nuanced examination of its adeptness in wielding the deceptive and persuasive communication skills that lie at the heart of Diplomacy.

To test levels of trickery, the team developed a system to analyze in-game conversations using a technique called Abstract Meaning Representation (AMR), which distills complex natural language messages into structured, machine-readable data.

AMR enabled the researchers to compare what players said they would do in their messages with what they actually did in the game. For instance, if Germany told England, "I'll support your invasion of Sweden in the next turn," researchers would check whether the player actually provided that support—or instead made a contradictory move.

This method allowed the researchers to quantify instances of deception and persuasion, as well as compare CICERO's communication skills with those of humans.

## **Strategy outweighs speech**

Despite the fact that CICERO won 20 out of 24 games, the study found that its messages often lacked coherence and didn't reflect its actual gameplay intentions. "If you pay attention to what it's saying over the game, it's garbage," May said. "What it's saying is things a Diplomacy player has said before. It's not reflective of what it's actually doing."

The researchers also ran experiments where they limited CICERO's communication in different ways. In some games, the model couldn't send messages at all, while in others, it could only send very basic

strategic information. Changing these inputs did not significantly impact its high score, suggesting that negotiation skills play little part in the model's Diplomacy talent.

Humans, however, lead the way in lying. The study revealed that CICERO is less deceptive and less persuasive than humans. It is also less susceptible to persuasion. Human players were found to be more intentionally deceptive and more successful at persuading other humans compared to CICERO. Intriguingly, humans also lied more to CICERO once they recognized it as an AI.

"What really makes CICERO good is that it's seen a whole lot of Diplomacy play and knows what moves to make," May said. "It struggles to be really convincing or duplicitous, and it doesn't significantly react to what other players are saying."

## Helping humans

Though it's just a game, understanding the nature of AI deception in Diplomacy could pave the way for new research into more consequential forms of adversarial communication. May suggests that these insights could help develop applications to combat AI-generated threats in real-world scenarios, such as a digital assistant that helps humans identify misinformation and navigate who or what to trust online.

"There's a lot of jerks out there," May said. "We want to guard against that by providing an extra layer of help."

Provided by University of Southern California

Citation: Why does AI beat humans at the strategy game Diplomacy? (2024, August 15) retrieved 15 August 2024 from

<https://techxplore.com/news/2024-08-ai-humans-strategy-game-diplomacy.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.