



Written by [Selwyn Duke](#) on May 15, 2024

Report: AI Is Shaping Up to be an Effective Politician — and a Deadly Statesman

If creations do often reflect their creators, at least in some way, perhaps it's not surprising that artificial intelligence has reportedly developed a troubling new skill: lying.

In fact, writes the [AFP](#), forget about fears of AI potentially “going rogue” — because it already has.

Science Alert [reports](#) on the story:

You probably know to take everything an artificial intelligence (AI) chatbot says with a grain of salt, since they are often just scraping data indiscriminately, without the nous to determine its veracity.

But there may be reason to be even more cautious. Many AI systems, new research has found, have already developed the ability to deliberately present a human user with false information. These devious bots have mastered the art of deception.

“AI developers do not have a confident understanding of what causes undesirable AI behaviors like deception,” says mathematician and cognitive scientist Peter Park of the Massachusetts Institute of Technology (MIT).

“But generally speaking, we think AI deception arises because a deception-based strategy turned out to be the best way to perform well at the given AI’s training task. Deception helps them achieve their goals.”



Vertigo3d/iStock/Getty Images Plus

In other words, AI is lying for what's usually the same reason immoral people and average politicians (forgive the redundancy) do. What's more, this supposedly is contrary to the AI's programming.



Written by [Selwyn Duke](#) on May 15, 2024

As the *Daily Mail* [informs](#):

The most striking example of AI deception they [the MIT researchers] uncovered was Meta's CICERO, a system designed to play the world conquest game Diplomacy that involves building alliances.

Even though the AI was trained to be 'largely honest and helpful' and 'never intentionally backstab' its human allies, data shows it didn't play fair and had learned to be a master of deception.

Other AI systems demonstrated the ability to bluff in a game of Texas hold 'em poker against professional human players, to fake attacks during the strategy game Starcraft II in order to defeat opponents, and to misrepresent their preferences in order to gain the upper hand in economic negotiations.

While it may seem harmless if AI systems cheat at games, it can lead to 'breakthroughs in deceptive AI capabilities' that can spiral into more advanced forms of AI deception in the future, the experts said.

Some AI systems have even learned to cheat tests designed to evaluate their safety, they found.

In one study, AI organisms in a digital simulator 'played dead' in order to trick a test built to eliminate AI systems that rapidly replicate.

This suggests AI could 'lead humans into a false sense of security,' the authors said.

The major short-term risks of deceptive AI include making it easier for people to commit fraud and tamper with elections, they warned.

(My, you mean, we may never again experience another 2020, the Most Secure Election in American History™?)

Moreover, much as with combating election fraud, we are currently with AI deception ever a day late and a dollar short. To wit:

"These dangerous capabilities tend to only be discovered after the fact,' Park told AFP, while 'our ability to train for honest tendencies rather than deceptive tendencies is very low,'" the AFP [relates](#).

"Unlike traditional software, deep-learning AI systems aren't 'written' but rather 'grown' through a process akin to selective breeding, said Park," the site continues.

"This means that AI behavior that appears predictable and controllable in a training setting can quickly turn unpredictable out in the wild" (as with some children I've known).

In reality, however, while the specifics may be unpredictable, AI's embrace of immoral behavior is perhaps entirely predictable. After all, too many modern people today are themselves governed by expediency.

Instructive here, as "art imitates life," is a relevant scene from 1991 film *Terminator 2: Judgment Day*. After the young John Connor character stops the Terminator — a highly advanced, artificially intelligent robot — from shooting someone, the boy exclaims, "You just can't go around killing people!" To this the robot replies, robotically, "Why?"



Written by [Selwyn Duke](#) on May 15, 2024

Connor, exasperated, then asks rhetorically, “What do you mean why?! ‘Cause you can’t!”

“Why?” the Terminator again asks monotonically.

“Because you just can’t, OK? Trust me on this,” is all the boy can say.

In reality, the robot’s (AI’s) response and implied assumptions are entirely logical, given the premises from which it’s operating. For you can’t see a moral under a microscope or a principle in a Petri dish; *no moral position can be proven* — that is, *scientifically*. Operating under the assumption that what we can see and touch, the physical world, is all that exists, murder (or anything else) isn’t wrong, just possible.

Given this, and that the materialistic premise is now common, it’s no surprise that humans — mere robots themselves (organic ones) if that premise is correct — essentially draw the Terminator’s conclusions. As to this, just consider a passage from a 2023 Underknown [article about nihilism](#).

“One night I was scrolling through Reddit and I came across the question — ‘if you had the chance to save your pet or a stranger, who would you save?’” the author related. “An overwhelming number of people said their pet, pretty obviously. When one commenter was confronted, they simply asked the question ‘why do you think a human life is worth more than that of an animal?’ And no one had an answer.”

That’s a shame because there’s a very good answer. It begins with the axiom (a self-evident truth that requires no further proof) that Truth, transcendent by nature, exists and dictates that there is objective right and wrong. And, among other things, Truth deems that murder is wrong and that human life is to be placed above animal life. This is just as logical as the “Terminator” conclusion — only, we start with a different premise.

Nonetheless, many people are uncomfortable with this — Bill Maher once said (I’m paraphrasing) “The concept of Absolute Truth is scary” — because Truth’s existence implies God’s, and because it means their wrongs aren’t mere “lifestyle choices” but are actual sins. Like it or not, however, there is *no other basis for morality*. This also means there’s no other basis for true civilization.

Not governed by Truth, the fictional Terminator simply acted based on expediency, determined by its programmed priorities; likewise, the real-life, divorced-from-Truth Reddit users are acting based on their priorities. So would it be surprising if our real-life AI were no different?

Perhaps what we really should fear from AI is that it may become just like its post-modern creators, only more so — that is, the perfect sociopath.



Subscribe to the New American

Get exclusive digital access to the most informative,
non-partisan truthful news source for patriotic Americans!

Discover a refreshing blend of time-honored values, principles and insightful perspectives within the pages of "The New American" magazine. Delve into a world where tradition is the foundation, and exploration knows no bounds.

From politics and finance to foreign affairs, environment, culture, and technology, we bring you an unparalleled array of topics that matter most.



Subscribe

What's Included?

- 24 Issues Per Year
- Optional Print Edition
- Digital Edition Access
- Exclusive Subscriber Content
- Audio provided for all articles
- Unlimited access to past issues
- Coming Soon! Ad FREE
- 60-Day money back guarantee!
- Cancel anytime.