The Atlantic

TECHNOLOGY

The Problem With Counterfeit People

Companies using AI to generate fake people are committing an immoral act of vandalism, and should be held liable.

By Daniel C. Dennett

The Atlantic / Getty

May 16, 2023

Money has existed for several thousand years, and from the outset counterfeiting was recognized to be a very serious crime, one that in many cases calls for capital punishment because it undermines the trust on which society depends. Today, for the first time in history, thanks to artificial intelligence, it is possible for anybody to make counterfeit people who can pass for real in many of

the new digital environments we have created. These counterfeit people are the most dangerous artifacts in human history, capable of destroying not just economies but human freedom itself. Before it's too late (it may well be too late already) we must outlaw both the creation of counterfeit people and the "passing along" of counterfeit people. The penalties for either offense should be extremely severe, given that civilization itself is at risk.

It is a terrible irony that the current infatuation with fooling people into thinking they are interacting with a real person grew out of Alan Turing's innocent proposal in 1950 to use what he called "the imitation game" (now known as the Turing Test) as the benchmark of real thinking. This has engendered not just a cottage industry but a munificently funded high-tech industry engaged in making products that will trick even the most skeptical of interlocutors. Our natural inclination to treat anything that seems to talk sensibly with us as a person-adopting what I have called the "intentional stance"-turns out to be easy to invoke and almost impossible to resist, even for experts. We're all going to be sitting ducks in the immediate future.

The philosopher and historian Yuval Noah Harari, writing in The Economist in April, ended his timely warning about AI's imminent threat to human civilization with these words: "This text has been generated by a human. Or has it?"

It will soon be next to impossible to tell. And even if (for the time being) we are able to teach one another reliable methods of exposing counterfeit people, the cost of such deepfakes to human trust will be enormous. How will you respond to having your friends and family probe you with gotcha questions every time you try to converse with them online?

Creating counterfeit digital people risks destroying our civilization. Democracy depends on the informed (not misinformed) consent of the governed. By allowing the most economically and politically powerful

people, corporations, and governments to control our attention, these systems will control us. Counterfeit people, by distracting and confusing us and by exploiting our most irresistible fears and anxieties, will lead us into temptation and, from there, into acquiescing to our own subjugation. The counterfeit people will talk us into adopting policies and convictions that will make us vulnerable to still more manipulation. Or we will simply turn off our attention and become passive and ignorant pawns. This is a terrifying prospect.

The key design innovation in the technology that makes losing control of these systems a real possibility is that, unlike nuclear bombs, these weapons can reproduce. Evolution is not restricted to living organisms, as Richard Dawkins demonstrated in 1976 in The Selfish Gene. Counterfeit people are already beginning to manipulate us into midwiving their progeny. They will learn from one another, and those that are the smartest, the fittest, will not just survive; they will multiply. The population explosion of brooms in The Sorcerer's Apprentice has begun, and we had better hope there is a non-magical way of shutting it down.

There may be a way of at least postponing and possibly even extinguishing this ominous development, borrowing from the success-limited but impressive-in keeping counterfeit money merely in the nuisance category for most of us (or do you carefully examine every $20 bill you receive?).

As Harari says, we must "make it mandatory for AI to disclose that it is an AI." How could we do that? By adopting a high-tech "watermark" system like the EURion Constellation, which now protects most of the world's currencies. The system, though not foolproof, is exceedingly difficult and costly to overpower-not worth the effort, for almost all agents, even governments. Computer scientists similarly have the capacity to create almost indelible patterns that will scream FAKE! under almost all conditions-so long as the manufacturers of cellphones,

computers, digital TVs, and other devices cooperate by installing the software that will interrupt any fake messages with a warning. Some computer scientists are already working on such measures, but unless we act swiftly, they will arrive too late to save us from drowning in the flood of counterfeits.

Did you know that the manufacturers of scanners have already installed software that responds to the EURion Constellation (or other watermarks) by interrupting any attempt to scan or photocopy legal currency? Creating new laws along these lines will require cooperation from the major participants, but they can be incentivized. Bad actors can expect to face horrific penalties if they get caught either disabling watermarks or passing on the products of the technology that have already been stripped somehow of their watermarks. AI companies (Google, OpenAI, and others) that create software with these counterfeiting capabilities should be held liable for any misuse of the products (and of the products of their products-remember, these systems can evolve on their own). That will keep companies that create or use AI-and their liability-insurance underwriters-very aggressive in making sure that people can easily tell when conversing with one of their AI products.

I'm not in favor of capital punishment for any crime, but it would be reassuring to know that major executives, as well as their technicians, were in jeopardy of spending the rest of their life in prison in addition to paying billions in restitution for any violations or any harms done. And strict liability laws, removing the need to prove either negligence or evil intent, would keep them on their toes. The economic rewards of AI are great, and the price of sharing in them should be taking on the risk of both condemnation and bankruptcy for failing to meet ethical obligations for its use.

It will be difficult-maybe impossible-to clean up the pollution of our media of communication that has already occurred, thanks to the arms race of algorithms that is spreading infection at an alarming rate.

Another pandemic is coming, this time attacking the fragile control systems in our brains-namely, our capacity to reason with one another-that we have used so effectively to keep ourselves relatively safe in recent centuries.

The moment has arrived to insist on making anybody who even thinks of counterfeiting people feel ashamed-and duly deterred from committing such an antisocial act of vandalism. If we spread the word now that such acts will be against the law as soon as we can arrange it, people will have no excuse for persisting in their activities. Many in the AI community these days are so eager to explore their new powers that they have lost track of their moral obligations. We should remind them, as rudely as is necessary, that they are risking the future freedom of their loved ones, and of all the rest of us.