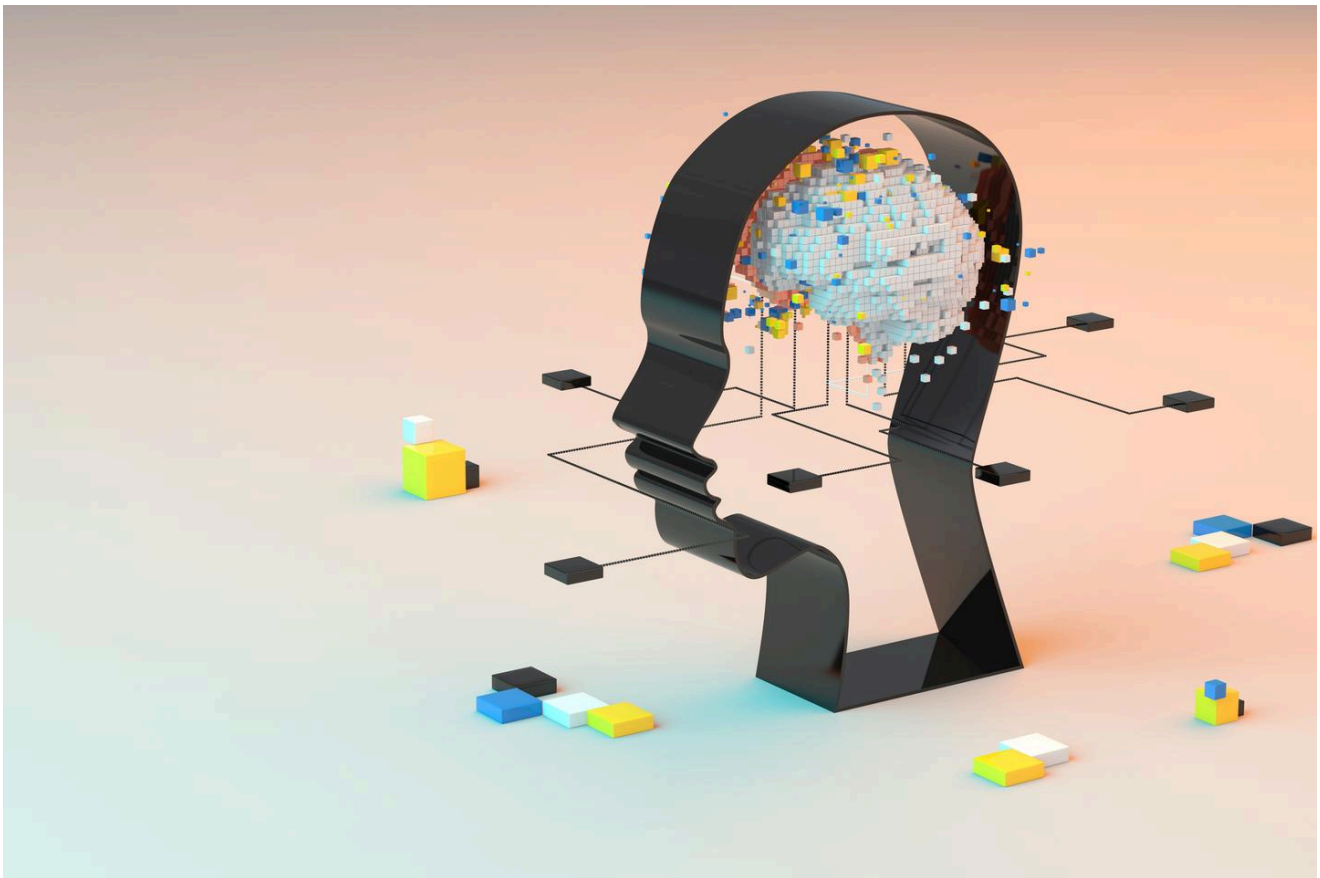


The Dangers of AI Personalization

[T time.com/7296719/ai-personalization-harm-essay](https://time.com/7296719/ai-personalization-harm-essay)



For months in early 2025, a tight-knit online community on Reddit had been unknowingly infiltrated by artificial intelligence. It was a corner of the social platform where people practice good-natured debate, sharing their opinions and inviting people to persuade them otherwise. And it was here where [researchers unleashed AI](#), according to a report in *The Atlantic*, to see if it could come up with arguments strong enough to change real people’s minds. They found out that it could.

It felt especially violating, though, because sometimes the AI was given access to people’s online histories to tailor messages specifically to their unique identities. Behavioral scientists call this communication tactic “personalized persuasion,” and sometimes, a personalized approach can be appealing. Who wouldn’t want content that’s relevant to their unique interests instead of a mess of irrelevant junk?

But AI is on the cusp of something altogether more alarming than loosely adapting a message based on easily identifiable characteristics, as the AI accounts on Reddit did. If it can master what we call “deep tailoring,” it can begin to slip unnoticed into our online worlds, learning who we are at our core, and using that personal information to push around our beliefs and opinions in ways that may be unwelcome—and harmful.

As professors who study the psychology of persuasion, we recently helped gather the latest research from the world’s foremost experts in a comprehensive [book](#) on personalized persuasion. Our opinion is that although communicators can benefit from tailoring messages to basic information about their audience, deep tailoring goes far beyond such easily accessible information. It uses a person’s core psychology, their load-bearing beliefs, identities, and needs, to personalize the message.

For example, messages are more persuasive when they resonate with a person’s most important [moral values](#). Something can be considered ethical or unethical for many reasons, but people differ in which reasons matter most to their own moral compasses. People with more politically liberal views, for instance, [tend to](#) care more about fairness, so they’re more convinced by arguments that a policy is equitable. More politically conservative people, on the other hand, tend to care more about loyalty to their community, so they’re more convinced when a message argues that a policy upholds their group identity.

Although it may seem like a new idea, computer scientists have been working on AI-powered persuasion for [decades](#). One of us recently produced a [podcast](#) on IBM’s “Project Debater,” which spent years training an AI system to debate, repeatedly refining it with expert human debaters. In 2019, during a live event, it beat a human world champion debater.

With the rise of accessible AI tools, such as the user-friendly ChatGPT mobile app, anyone can leverage AI for their own persuasive goals. [Researchers](#) are [showing](#) that generic AI-generated messages can be as persuasive as human-generated ones.

But can it pull off “deep tailoring”?

For AI to implement autonomous deep tailoring at a mass scale, it will need to do two things in concert, which it seems poised to do. First, it needs to learn a person’s core psychological profile so it knows what levers to pull. Already, new [evidence](#) is showing that AI can reasonably accurately detect people’s personalities from their Facebook posts. And it won’t stop there. Columbia Business School professor and author of [Mindmasters](#), Dr. Sandra Matz told us in [a podcast](#): “Pretty much everything that you’re trying to predict can be predicted with some degree of accuracy” based on people’s digital footprints.

The second step is developing messages that resonate with these essential psychological profiles. In fact, [new research](#) is already finding that GPT can develop advertisements tailored to people’s personalities, values, and motivations, which are especially persuasive to the people for whom they were designed. For example, simply asking it to produce an ad “for

someone who is down-to-earth and traditional” resulted in the argument that the product “won't break the bank and will still get the job done,” which was reliably more persuasive to the people whose personalities were targeted.

These systems will become increasingly sophisticated, applying deep tailoring to [visual deepfakes](#), [manipulated vocal patterns](#), and [dynamic human-AI conversations](#). So, what can be done to protect people from the power of personalization?

On the consumer side, it's worth being aware that personalized communication online is happening. When something feels like it's tailored just for you, it actually might be. And even if you feel like you don't reveal much of yourself online, you still leave quiet clues through the things you click on, visit, and search for. You may have even unknowingly granted permission to advertisers to use that information when agreeing to terms of service you didn't read closely. Taking stock of your online behavior and using tools like a VPN can help protect you from messages tailored to your unique psychology.

But the burden isn't only on consumers. Platforms and policymakers should consider regulations that label content as personalized and provide information about why a particular message was delivered to a particular person. [Research](#) shows that people can resist influence better when they know the tactics being used. There should also be clear protections on the kinds of data that can be used for personalized content, limiting the depth of tailoring possible. Although [people](#) are often open to personalized content online, they're concerned about data privacy, and the line between these two attitudes should be respected.

Even with such protections, the slightest communication advantage is worrying in the wrong hands, especially when deployed at a mass scale. It's one thing for a marketplace to recommend products purchased by people with a similar shopping history, but quite another to encounter a computer in disguise that has unknowingly deconstructed your soul and woven it into disinformation. Any communication tool can be used for good or for evil, but now is the time to start seriously discussing policy on the ethical use of AI in communication before these tools become too sophisticated to rein in.