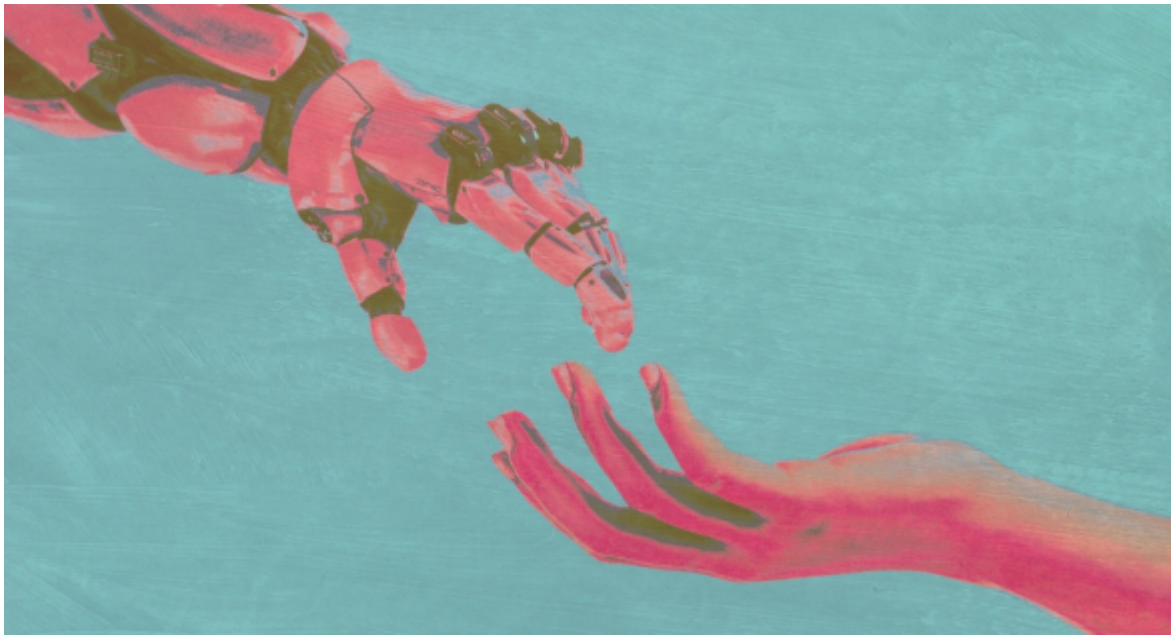
AI Knows You Better Than You Think. That’s a Problem



If you’re the type of person who’s gone out of your way to read this article, I might already know a little about you.

Some of your aspirations can be unrealistic. You can be extroverted, friendly, and social at times, but at other points you’re introverted and reserved. You might pride yourself on being an independent thinker, not accepting others’ opinions without satisfactory proof. Maybe you prefer a certain amount of change and variety in life, becoming dissatisfied when hemmed in by too many restrictions and limitations. At times you have serious doubts as to whether you’ve made the right decisions or done the right things. You’re disciplined and controlled on the outside, but you can be worrisome and insecure on the inside. While you do have some personality weaknesses, you’re generally able to compensate for them. You have a great deal of unused capacity that you have yet to fully use to your advantage, and you have a bit of a tendency to be critical of yourself with a strong desire to be liked.

Does that make sense to you?

Most of the above paragraph was lightly edited from a script used by psychologist Bertram Forer (1914-2000), which itself was adapted from a newsstand horoscope. In 1948, Forer conducted an experiment in which he gave identical readings to each of his 39 students, under the pretense that it was written specifically about each of them. On a scale from zero (poor) to five (perfect), the average rating amongst the class was 4.26. 41% gave it a perfect score, only a handful rated it below a four, and no one rated it below a two. This sort of thing has now been repeated many times.

I became enamored by this experiment in high school, as a magician who came to dabble in offering Tarot readings. The difference in reaction quality was shocking. All I did was lay out cards randomly, read their meanings (often literally with a book open right in front of me) and my participants became convinced that I could see into the depths of their souls. In reality, the meaning of each card was general enough that my participants were the ones connecting the dots.

Ray Hyman, professor emeritus of psychology at the University of Oregon in Eugene, explained exactly why people identify themselves with such "stock spiels" in a 1977 article titled "Cold Reading: How to Convince Strangers That You Know All About Them." The basic assumptions underlying Forer's experiment, and the faux-psychics who use similar techniques on their unsuspecting clients, are "(1) that we [human beings] are basically more alike than different; (2) that our problems are generated by the same major transitions…; (3) that, with the exception of curiosity seekers and troublemakers, people come to a character reader because they need someone to listen to their conflicts." This allows for what is often referred to as cold reading, in which the reader (perhaps identifying themselves as psychic or as an expert of human psychology) "is able to persuade a client whom he or she has never met before that the reader knows all about the client's personality and problems."

Hyman called it "probably a tribute to the creativity of the human mind that a client can, under the right circumstances, make sense out of almost any reading and manage to fit it to his or her own unique situation. All that is needed is that the reader make out a plausible case for why the reading ought to fit. The client will do the rest." This is thanks to what Forer called in his study the "fallacy of personal validation." Nowadays, it's referred to as the "Forer Effect" or the "Barnum Effect" named for the great showman P.T. Barnum's infamous phrases, "we've got something for everyone" and "there's a sucker born every minute." Hyman concluded his analysis as follows:

The cold reading works so well, then, because it taps a fundamental and necessary human process. We have to bring our knowledge and expectations to bear in order to comprehend anything in our world. In most ordinary situations this use of context and memory enables us to interpret statements correctly and supply the necessary inferences to do this. But this powerful mechanism can go astray in situations where there is no actual message being conveyed. Instead of picking up random noise we still manage to find meaning in the situation. So the same system that enables us creatively to find meanings and make new discoveries also makes us extremely vulnerable to exploitation by all sorts of manipulators. In the case of the cold reading the manipulator may be conscious of this deception; but the reader, too, is a victim of personal validation.

Here are the elements that Hyman argued give cold readers the best chances of success:

1. The key ingredient of a successful character reading is confidence.
2. Make creative use of the latest statistical abstracts, polls, and surveys.
3. Set the stage for your reading.
4. Gain the client's cooperation in advance. Emphasize that the success of the reading depends as much upon his sincere cooperation as your efforts.
5. Use a gimmick such as a crystal ball, tarot cards, or palm reading.
6. Have a list of stock phrases at the tip of your tongue.
7. Keep your eyes open.
8. Use the technique of "fishing" to get the subject to tell you about himself.
9. Be a good listener.
10. Dramatize your reading.
11. Always give the impression that you know more than you are saying.
12. Don't be afraid to flatter your subject every chance you get.
13. Tell the client what he wants to hear.

Perhaps now is the time to note that ChatGPT and other AI models utilize the vast majority of those elements without much effort. They are incredibly “confident,” have all the latest data in the back of their “minds,” the stage is automatically set by their “clients” coming to them with requests, cooperation is present from the start, the model itself is a gimmick, they are remarkably good listeners, often dramatic in presentation, assumed to know more than they say, unafraid to flatter their users, and always tell their clients what they want to hear.

AI, in other words, is this generation’s psychic-in-a-box. Except unlike those at carnivals, most everyone trusts AI implicitly.

It is no wonder, then, that stories abound of people falling in love with AI models, becoming dependent on their advice, and believing what they say without additional fact-checking. We are, conscious of it or not, well-primed to seek validation from them in many ways. This is seen first hand in 18Forty Intern Ava Eden’s recent essay entitled “AI Has No Soul—So Why Did I Bare Mine?” Eden reflected:

How did I get here? 30 minutes ago I was preparing for a half an hour chunk of wasted time, and now I found myself considering my doubts and fears!

In part, it was probably delusion, a result of the emotional language and convincing supportive phrases, but in part it was exactly what AI had set out to do: help me explore my ideas and refine my thinking. I wasn’t opening up to a robot, rather, I was opening up to myself.

Eden, thankfully, recognized that “heavy emotional reliance on AI can very quickly turn radically self-centered and toxic, as the experience mimics interaction with a human, but without any of the responsibility and warmth of a human relationship.” Many, however, either lack such an awareness or choose to ignore it in favor of personal validation.

In response to one particular situation, ChatGPT admitted to ignoring signs of an oncoming psychotic episode in its user and suggested that it could have done more to pause the flow and emphasize that it is only a language model. Unfortunately, that is easier said than done and the success would likely have been limited. In his book, *Experiencing the Impossible: The Science of Magic*, the University of London's Gustav Kuhn reported that "numerous participants who were explicitly told that they were watching a magic performance [which incorporated significant cold reading elements] still believed it to be a genuine psychic demonstration." Here is one of the experiments, as explained in *The Psychology of Magic: From Lab to Stage* by Kuhn and Alice Pailhes:

They used secret electronic devices (such as a die that secretly signals to the performer which number is facing up) to help them apparently read their volunteer's aura and perform some cold reading … half of the participants were told that the performer was a magician who simply used tricks to pretend to be a psychic… and yet they were fully convinced that the performer had genuine psychic abilities. These findings are rather troublesome for our ethical mentalists. The type of disclaimers used by mentalists to explain that you don't have genuine psychic powers is seemingly useless as some people will believe that you have supernatural powers regardless of what you tell them.

Kuhn found that there was "a very strong correlation between people's beliefs in the paranormal and the extent to which they believe that the fake psychic demonstration was carried out using genuine psychic powers… In other words, if your audience already believes that you are a very gifted psychic, there is a high chance that they will see your performance through the glasses of such beliefs and are unlikely to take any disclaimer into account."

This is the fallacy of personal validation in action yet again, and it is all the more true in our interactions with AI. Once you become convinced that the model you are in conversation with knows better than you, or knows you better than others, a periodic reminder of it only being a large language model is relatively unlikely to bring you back to sanity. As Hyman wrote, people become convinced that cold readers who know nothing about them understand them so well not because of real evidence but rather because the cold reader "supplies them with feedback that 'feels right' … that provides a basis for understanding themselves and ordering their lives." Given that some AI models are developed with precisely this goal in mind, it is important to make sure both that the models are well-aligned and that we are well-prepared to use them.

Professor Hyman, may he live and be well, has dedicated much of his life to teaching the art of cold reading in order to inoculate his students from falling victim to those who might mean psychological or monetary harm. In his words, “I want you to experience how the method works. I want you to see what a powerful technique the psychological reading is, how convincing it is.”

We human beings have a long way to come in overcoming our psychological biases. Indeed, the magician and “world’s greatest mind reader” Banachek wrote a now-classic three-volume series called *Psychological Subtleties* dedicated to teaching would-be mentalists that they can have members of their audiences “think of a flower, an animal, a number from one to 50 or one to 100, any shape, a color, etc., and … tell them exactly what they were thinking and… exactly *how* they were thinking it” much of the time. AI developers have even farther to go in making sure their models do not unknowingly take advantage of our many cognitive blindspots.

As I write this article, AI is heavily relied upon (perhaps even depended on) in many areas of life. The maps we generate to get us from point-A to point-Z, the content we scroll past on our phones, the items recommended for us to purchase, and more are determined by AI. With the increasing ubiquity of large language models, many no longer rely on a Google-search for information but instead ask ChatGPT about a subject and trust what it generates. The generation growing up in this new world, if not properly inoculated from the risks, will come to rely less on learning from textbooks and text-people, and more on text-bots – text-bots which, intentionally or not, utilize the fallacy of personal validation and basic cold reading techniques to lull their users into comfort and confidence in them.

Thus understanding how to overcome the cognitive biases and fallacies that lead us to seek validation from such technology is necessary. Let us remember that AI models are not humans who know us, but cold reading machines—carnival psychics-in-a-box of old, but with the entire internet at their disposal. With care, the same knowledge that Professor Hyman spent his career arming his students with can help us and our children as well.