Scribe AI

by
Robert Cunningham

Envisioning the Future of Computing Prize
Social and Ethical Responsibilities of Computing
Massachusetts Institute of Technology

Introduction

This piece considers the implications of personalized language models. After learning to predict generic text, like Wikipedia, these models are fine tuned to predict an individual's writing based on their past texts and emails.

We hear from several characters about their experiences: Margaret and Vincent, a couple in college, and Laura, a founder at ScribeAI.

The structure resembles both Terkel's *Working* and Chiang's *Liking What You See*, in the way it uses a series of interviews to tell a story of societal change. I read and liked both, neither recently.

Although this is not exactly an essay, the structure was approved in accordance with the competition guidelines.

Margaret

I broke up with Dustin three months ago. We were watching Breaking Bad, and my phone buzzed. So I look over, and it's a text from him–a long response to one of my rants about my aunt. But here's the thing: there's no way he could have written it. He was right next to me on the couch the whole time.

So, he eventually admitted that he was using some kind of auto-responder service to reply to my texts. He said he uploaded all his past messages to it, and I guess it learned to text like him. Then he forgot to turn it off.

Dustin said he only used it occasionally, when I talked about something boring like my relationship with my aunt. He said didn't want to ignore me, but he didn't want to spend the time thinking about it, so he just turned on this autoresponder, and it wrote the responses for him.

It felt horrible. Like I've been trying to work through these feelings about my aunt, you know? And he couldn't even be bothered to read what I had to say, so he just offloaded talking with me onto a robot!?

Anyway, we had a big fight and I said that if he didn't even want to text me, then maybe he shouldn't date me, and maybe he should go home, and maybe he should never talk to me again, and that was the end of that.

Laura

I started ScribeAI in 2025. Our goal was simple: we wanted to personalize language models.

From a technical standpoint, language models are pretty simple. You train them on some text, and they try to predict the next word in a sentence.

ScribeAI has a generic language model trained on the entire internet. It predicts the next word in a sequence of words. For example, given the input “I hate waking up ____,” it would predict that the word “early” is most likely to follow. Many people on the internet hate waking up early.

Enthusiasts had been using generic language models to write texts and emails for a few years, but it was pretty clunky. You could always tell when someone went from writing their own emails to using a language model. One day they’d be writing two line emails full of weird slang and zingers, and the next day their emails would be ten paragraphs of polite Wikipedia English.

We started ScribeAI to personalize that generated text. You upload a bunch of your past emails and texts to us, and we train a language model specifically on your data. It learns what kind of writer you are, and then it can generate emails and texts in your voice. We call this personalized generator your *Scribe*.

Your Scribe learns all kinds of things about you during the training process. To respond to jokes, it learns whether you're a "lol" person or a "haha" person. To complete the phrase “I hate waking up ____,” it learns whether you’re the kind of person more likely to say “early” or “late.” To respond to your coworker asking about that late project, it learns whether you’re the kind of person more likely to start your response with “actually” or “sorry.” To respond to your boyfriend’s texts about his long day, it learns whether you’re more likely to respond with “have you tried” or “that must be tough.”

In the process of learning to imitate your writing style, our models absorb deep aspects of your personality. Are you more argumentative or apologetic? How are your sleeping habits? What kind of emotional support do you generally provide?

Our Scribes use this knowledge to suggest realistic responses to our clients, which ultimately helps them spend less time in their email and more time in their lives.

Margaret

So after I broke up with Dustin, I figured I’d give dating another try. And I kind of noticed that people on dating apps were much more talkative than last year. Eventually my friend told me what was up: I was mostly talking to bots.

I looked it up, and sure enough there's this service called ScribeAI. I just uploaded my entire chat history to it, a bit creepy but whatever? I don't have anything to hide, you know? And they deleted it all afterwards I guess.

Now, when I match with someone, I just attach my ScribeAI account to the chat. My Scribe talks with the person I matched with, simulating the kinds of things I would have said. I just check in on the conversations every once in a while. If it looks interesting, I schedule a date. Otherwise, I just unmatch.

I used to spend hours replying to messages, and I was talking to so many guys I couldn't keep them all straight. Now I barely have to write any messages at all.

I ended up dating this guy named Vincent. He's really talkative, which is a good fit. Often I want to talk with someone but don't have anything in particular to say; he's in a fraternity and has lots to talk about, so our conversations flow easily. And he really knows how to listen when I want to blow off steam. I think it's a great fit.

Vincent

Yeah, I've been using ScribeAI since I got to college. Man, it's crazy good for dating apps. I uploaded my entire chat history to their website, and now they take care of all my conversations. Last time I was single, I went on 20 or 30 dates before I found someone. Some of those dates were terrible–my date talked the whole time about her favorite '70s rock bands, or even worse, never talked at all. Now that I'm using my Scribe, it's like I'm having hundreds of conversations at once. I only bother meeting the people that my Scribes have good conversations with. I think my Scribe had something like 500 conversations before I found Margaret.

And let me tell you, ScribeAI isn't just for dating. I'm a brother at Pi Kappa Phi, and we're using it all over the place for rush.

Normally, during the first few weeks of school, I have to talk to hundreds of frosh. Then we have all these meetings, where we have to guess if a frosh is going to be a good brother at PKP.

This year, we tried something new. We asked all our brothers and all the frosh to get ScribeAI accounts. Then we simulated a group chat with the current fraternity and each new member. If the recruit's Scribe had good conversations with the brothers' Scribes, we considered admitting them. Otherwise we dropped them from our process without wasting their time.

We have to be really careful not to mess up the fraternity vibe during rush. Right now we have this great laid-back vibe, but if we admit too many hard asses we might lose that culture. This year, we noticed that a simulated group chat with new recruits spent too much time worrying about their assignments. We dropped one particularly talkative academic recruit, and when we re-ran the simulation, the vibe was back how we like it.

We can even screen for stuff like sexism. One brother had the idea to simulate one-on-one conversations with each new recruit, where the brother says something overtly sexist. Then, if the recruit goes along with it, we drop them. We caught a couple of recruits that way this year.

Even aside from the frat and dating stuff, I'm starting to use ScribeAI to find friends. I like history, but not too many dudes in my frat dig it. So I signed up for this college-wide Scribe pool. You search for something like "I want to find friends who like history." Then the admins run hundreds of simulated conversations between you and all the other people in the pool. They send back a list of the people that you had the longest simulated history conversations with.

I started using this right when it came out, and now one of my closest friends is from the pool. She's this girl from the math department, but she's interested in history and we have these crazy conversations about the Bronze Age collapse.

Laura

We just hired Seth as a machine learning engineer at ScribeAI. He's coming from a job in accounting, but he studied computer science in college and we're very confident that he's a good fit.

Our hiring process here is a bit unusual. We don't really have time for conventional interviews. They take lots of time, and we're competing with everyone else for the same set of engineers. Instead, we require our applicants to submit their Scribe accounts. Then, to get to know the applicants, we subject their Scribes to simulated slack conversations. We hired Seth because his Scribe did particularly well in two tests: attention to detail and loyalty.

We test attention to detail by proposing a solution to a complicated technical question. If the candidate's scribe responds by saying "actually, have you considered …?," we consider it a success. Although the candidate might not know anything about our technology, their predicted response shows that they're *the kind of person* who often points out issues that others have overlooked. This focus and willingness to speak up is very important in our line of business.

Since we'll spend a lot of time and energy training Seth as a machine learning engineer, we also care about his loyalty. In our loyalty test, we pose as a competitor's recruiter and offer Seth a 50% raise to leave ScribeAI. Seth's Scribe said that our company had been generous to teach him so much, and that he wanted to pay it back before working at a competitor. That's exactly the kind of response we look for.

When it comes down to it, we're not hiring Seth for any of his particular skills. He knows next to nothing about the technology we work with. We're hiring him because, at a fundamental level, he's *the kind of person* we want to work with.

Margaret

Vincent's really into this Scribe stuff. He's used it for his frat, and his dating life, and even finding friends now? And last week we even tried this new long-term dating predictor. It simulates thousands and thousands of conversations between two people and reports some statistics. We decided to try it, and it turned out that we weren't that compatible. On average, in at least 77% of conversations, the texting thread eventually dies with us choosing to break up. There's a pretty consistent pattern, actually: it seems like when my Scribe feels neglected, Vincent's Scribe tends to become annoyed that I'm not more ambitious. Then he spends more time with his work friends, and I feel even more neglected, I guess.

I don't know, it's like a sort of negative spiral embedded in our personalities? But we had a long talk about it last week, and we decided that now that we know about this flaw, we'll try our best to avoid it.

I know Vincent thinks that scribes are the future, but I don't want them to tell me how to live my life. Now that we know the specific ways that we're likely to break up, I hope we can overcome

them and stay together. That's what it means to be human, instead of just a Scribe predicting a bunch of likely words, you know?

Laura

As a company, ScribeAI has a legal obligation to allow our clients to access their personal data.

We don't keep our clients' actual texting history around after we train their scribes. Instead, we compress all that information into a single 768-long vector. The 768 numbers in that vector encapsulate everything our model learned about them from their texts. In some sense, those numbers are all we need to reconstruct someone's personality.

Everyone used to be concerned about privacy. But after the Supreme Court ruled on Garza v. Oregon, the courts have consistently treated your vectors as legally equivalent to your entire chat history. The government can still get them with a warrant, but they need probable cause. In practice, it's more useful for the government to get your actual chat history instead. We get very few subpoenas these days.

Vincent

That long-term dating simulator got me thinking about whether we can change our vectors. Obviously kids' vectors change a lot as they grow up. But the vectors of teenagers and twenty-somethings? Do they change?

I want to go into politics, and I've been thinking, what if I could become a bit more similar to Obama? Could I learn from him by pulling my vector a bit closer to his? I checked, and right now our vectors are kinda similar, but not *that* similar. It's like 40% or 50% similarity or something.

I found a startup that helps you become more like someone you admire. Their main service is matching people. They know that when two people talk, their vectors gradually drift towards each other.

To take advantage of this, they match people whose vectors are different from Obama in opposite ways. Say my vector points to the right of Obama's, then they'll match me with someone whose vector points to the left. Then, when we talk, both of our vectors gradually move towards Obama's, and we'll both discover a new part of our personality.

My first match is great. He's a bit less socially confident than I am, but he's more down to earth and he speaks extremely clearly. We try to talk at least once a day, since it's helping both of us become more like the person we want to be.

It's crazy–in the past, I would have started out my political journey by doing something like working at a polling office. I would have had to hope that I gradually absorbed social skills from the candidates I was around. Instead, I'm skipping directly to working on my charisma and presidential potential.

Laura

We've had a few challenges as we've rolled this technology out. Before Garza v. Oregon, we struggled with legal uncertainty about privacy. Then we had mobs trying to cancel people for things they were merely likely to say.

Some nights I wondered whether we were the bad guys, pulling society towards some dystopia where the government knows everything about you.

Over time, I think the data has come to show our work in a positive light. Severe loneliness, which quietly rose all throughout the early 2000s, is at all-time lows and still dropping quickly. Likewise, global happiness is up. And those gains have been remarkably evenly distributed between rich and poor countries; everyone benefits from having closer friends.

We're giving people an honest mirror into who they are, and some people don't like it. But I go to work increasingly confident that we're pursuing the greatest social project of the century.