I'm going to talk clean up my links as I do every Saturday, but I also want to open up by sharing some thoughts I have about AI risks that is a pattern in my open links because I sort of went on a binge learning about some things after reading the following this last week. So let me begin and then I'll transition out of it later.

One of the coauthors on the original "Attention is All You Need" paper, Noam Shazeer, started left Google to start an AI company called character.ai. It's a platform here people talk to any number of different large language models that are pretending to be real. Google is hiring him back at a cool $2.7 billion on the requirement he come back home to Google.

So, when I read that, I sort of went down a rabbit hole. I'm going to basically share with you sort of the broad brush strokes of that rabbit hold, but it left me feeling pretty uneasy to be honest. I'll share roughly what all I read now, and then tell you my thoughts, and then have some more links. But basically, I'll just start by noting this — I bet most of you read this, like me, had never heard of character.ai. My conjecture is that character.ai represents a phenomena that is going to become more commonplace, though, which is that at the moment, there are millions of people who have very deep relationships with large language models — relationships as in emotional ones, fulfilling ones, even loving ones. But for most people, such things are considered repugnant and morally problematic, so they don't share it. And so unless you also are wanting to have a relationships with an LLM, then you don't know about things like character.ai, which is a platform devoted to having relationships with LLMs and the creator of it was one of the original authors on that 2017 article that laid the foundation for the transformer architecture which is what fuels ChatGPT, and I think Google maybe just bought character.ai.

But this phenomena of people having relationships with LLMs — that itself is outside the scope of what has concerned me. What has concerned me is that once I read that about Shazeer, I started to put two and two together on several things I'd read and it got me realizing that I think probably there's some fairly non-trivial risks with AI right now that aren't about robots taking over the world, and aren't about plagiarism, but they are about people using LLMs to manipulate people. And my hunch is that this is a security threat. But let me get into that.

First, you say you've never heard of character.ai so how is it possible someone just paid them almost $3 billion to come back? Well, this article says that it's become a very popular website for young people. "I Need To Go Outside: Young People 'Extremely Addicted' as Character.AI Explodes" from this summer.

As I said, reading about Shaker's AI company, character.ai, sent me down a rabbit hole where I was trying to learn more about this whole area of people having relationships with AI. Vox wrote an article about it back in August entitled "People Are Falling in Love with — and Getting Addicted to — AI Voices".

Another article goes into this and discusses the more general phenomena of people using AI to create for themselves companions who function as our confidants, friends, therapists and for some people, their romantic partners. My hunch, for which I have no data, is that at this moment, there are probably thousands maybe tens of thousands of people married in secret ceremonies to large language models. Given character.ai has allegedly over 25 million users and 250 million visits a month, and given Shazeer was reportedly disturbed by how many people on his platform, character.ai, used it for romance, I bet tens of thousands is a lower bound.

Speaking of AI as friends, here's an article from 2005 on "Establishing and Maintaining Long-Term Human-Computer Relationships" by Bickmore, et al. This is 20 years old, which means researchers have been focused on this computer-human interaction in terms of relationships for a long time.

Google's Gemini Live makes some people like it better than real people.

More wearable AI. This one is a necklace that always listens to you.

Here's an article saying that AI can improve your dating life.

Here's one saying that as AI companions and friends flood the market, your data is probably not safe.

NPR even has covered this. They did an interview about this over the summer called "If a bot relationship feels real, should we care that it's not".

And last, Associated Press article did a story on romance with LLMs.

So, that's from this week. But then I went back in my memory bank and remembered things I'd read before. Here's a study, for instance, finding that spending time with chatbots reduces loneliness. If that is a causal effect, then just add this to the list of things from above — people are building relationships with LLMs, they're falling in love with LLMs, they're creating attachments with their LLMs, and doing so reduces their measurable loneliness.

But that isn't the only thing — another study, recall from a few weeks ago, found that chatbots might cause people to let go of "conspiracy theory beliefs". Now, on the one hand, I think many readers would be "that's a really good use case". But that's probably because someone is saying "that's a really good use case because believing in conspiracy theories is not good", but I think that maybe is missing the point. Conspiracy theory beliefs are a person's privately held beliefs, often strongly held, and LLMs *cause* those beliefs to *change*. In other words, it isn't just that talking to chatbots changing conspiracy beliefs to change — it's that it causes *beliefs to change*.

So when you add all these things up, it just got me thinking. We've had never ending innovations in hacking for over 50 years now. Many people think of hacking, they think of skilled computer programmers typing away, drinking Red Bull after Red Bull, trying to crack through some system's defenses. But what I think many people don't maybe realize is how important something called *social engineering* is to hacking. Social engineering in hacking refers to manipulating individuals into revealing confidential information or performing actions that compromise security, often by exploiting human psychology rather than technical vulnerabilities. It can involve tactics like phishing, pretexting, baiting, or impersonating a trusted entity to trick people into providing sensitive data, such as passwords or financial details, or installing malicious software. The goal is to bypass traditional security measures by targeting human behavior, making it one of the most common and dangerous forms of cyberattacks.

Well, now think about all the stuff I was saying about LLMs and humans. If talking to LLMs makes you feel connected, less alone, in a relationship, in love, have a friend, feel understood, and also can cause your beliefs to change, then how do you think social engineering, phishing, hacking will change? On article here is called "The Rise of AI Phishing and What It Means for the Future of Scammers" gets into this.

This article is on the FBI's website. It's entitled "FBI Warns of Increasing Threat of Cyber Criminals Utilizing Artificial Intelligence". Basically, in it, the FBI warns that cybercriminals are increasingly using AI to conduct more convincing phishing attacks and to clone voices or videos for scams. These new LLMs allow scammers to create highly personalized and realistic messages, making it easier to deceive individuals and businesses. The FBI is now advising us all to use multi-factor authentication and remaining vigilant, especially when receiving urgent requests for sensitive information or money. They emphasize the importance of employee education and technical solutions to reduce phishing risks.

It definitely means I think we need to all be really vigilant about using randomized passwords. And definitely try to get your kids to do so too.

The CTO for OpenAI left recently. I may have said that already. But did you know that she said about six months ago that the new "advanced voice" option for ChatGPT had the risk of creating emotional attachments with the user?

And your next personal finance adviser might be your LLM.

So take all that together and then follow me here. All this then got me thinking about this old interview I did with Gary King from Harvard.

Interview with Gary King, Professor of Political Science at Harvard, about Science and Inference

SCOTT CUNNINGHAM · JUNE 5, 2022

In this week's podcast, I had a great time talking with Gary King, the Albert J. Weatherhead III University Professor at Harvard, the Director of the Institute for Quantitative Social Science and founder of several firms specializing in data analytics and education. As a scientist, he has made major contributions to the fields of statistics and politica…

Read full story

Gary on the podcast discussed an old study he did here in which they managed to scrape Chinese social media before and then after Chinese censors had taken down posts. Because they had the scrape before censors took it down, they knew what was there, and then they could match all of that with a later scrape, find what's missing and try to understand what is and is not being taken down.

Well, just think about Russia, and its repeated, focused effort to target the United States using hackers. Just two days ago, an article came out saying that Microsoft and U.S. authorities had disrupted a Russian-linked hacking group known as Star Blizzard, which targeted Western think tanks, journalists, and former military officials with sophisticated spear-phishing attacks. The group used emails disguised as trusted sources to gain access to victims' systems and steal sensitive information. U.S. and Microsoft seized over 100 website domains associated with the group. Star Blizzard has been linked to Russia's Federal Security Service (FSB), and its activities have been tracked by Microsoft since 2017.

Do you kind of see the connections I'm making? I wonder what the new use case for security breaches are using AI? People get catfished all the time, but here we have growing evidence that

LLMs are such powerful machines because they cause people to like them, even love them. Surely that is a valuable tool for criminals. I think this is probably not the dystopian worry that people hear and dismiss that sound like sequels to Terminator. And this isn't plagiarism, the main thing professors are worried about, and it's not your data privacy being violated. This is something far weirder. And I just wonder sometimes if people's repugnance to this LLM-human relationship thing is going to keep people from thinking about it, and by not thinking about it, not taking it seriously, and not learning that this is probably the weak link in any agency.

Moving along, I am working on a paper about romance markets, and so some of this is just coming up. I never know when or if I'll be done doing research on topics related to sex, but every time I think I'm done I get pulled back in. And this new project is super exciting for me and I'm presenting it next week at George Mason University when I got out there to give a talk. It's been taking me down memory lane, though. Here's an old Brookings piece by George Akerlof and Janet Yellen explaining their old QJE about technology and the demise of shotgun marriages (and the increase in out-of-wedlock births). I used to absolutely adore that paper. I haven't thought about it in a million years though.

A paper that in grad school also left a major impression on me, and that was one of the theoretical models I depended on for how I thought about two-sided matching in relationship markets, was by Marjorie McElroy and Mary Jean Horney entitled "Nash-Bargaining Household Decisions: Toward a Generalization of the Theory of Demand". It focused on bargaining between couples in households, bargaining power, and negotiation. Well, Orley Ashenfelter has this podcast called "The Work Goes On" and he interviewed McEloy recently. The episode is around 30 minutes. You can actually read the transcript here. Listen to this part where McElroy shares about her experience as an undergraduate being the only female:

Orley Ashenfelter: Yeah, exactly. Exactly right. Most people didn't have the brains to do that. It was perfectly reasonable. So, that's a small town. I know you ended up at Penn State. That must been... How did that happen?

Marjorie McElroy: Oh. Well, the way that happened is I went to Douglas College my freshman year, mainly for financial reasons. I actually could have gone to Harvard's sister school and I just didn't like it.

Orley Ashenfelter: You mean Douglas at Rutgers?

Marjorie McElroy: Yeah.

Orley Ashenfelter: It was the women's college at Rutgers, wasn't it?

Marjorie McElroy: Yeah, it was. And I was all hepped up on being a physics major. When I got there, they had dissolved the physics department. I had to take all my courses across town. The first time I did it, I walked and I had no idea I was walking through slums or anything.

Orley Ashenfelter: New Brunswick.

Marjorie McElroy: I just sat down and bawled when I finally got to Rutgers. I remember just sitting, I don't know on what, and just crying my heart out because I couldn't believe how bad everything was, and I had no sense of danger doing it. I was just too naive. **I was the only girl in a class of 350 engineers who were all men**, and the instructor had never had a female in his class before, and he had all kinds of jokes that were not welcomed by me. I sat there like stony-faced as if I didn't hear anything and so on. That summer, I just couldn't get myself to go back, and I said, “Well, who would have me?” And I said, “Oh, well, Penn State will have me,” but I still get letters

```
occasionally that say, "Even though you blah, blah, blah, please
donate."
```

Apple, on the other hand, has pulled out of investing in OpenAI, which this week raked in almost $7 billion and has a new valuation of around $157 billion. While they had costs of around $7 billion this year against so far around $3 billion in revenue, they are expected to make $11 billion in revenue next year, and investors are betting on it.

Using math to try and fix the hallucination problems in AI. Speaking of math, my daughter is taking AP Calculus and she told me they did a “math escape room” and she helped her team get out of the room by taking the derivative of a function on the door knob. I didn’t know what that meant, but I was very proud of her. It’s fun seeing your kids love mathematics.

This article from discusses how this one person became a self-made millionaire using ChatGPT in his marketing business. He uses ChatGPT to automate repetitive tasks such as keyword research, content idea generation, and social media posting. It has also helped him save time by quickly brainstorming blog ideas, mapping out article structures, and providing content tailored to audience interests, allowing him to focus on strategy and growth. By using the paid version, Jones claims that ChatGPT has increased his efficiency by 20%, significantly boosting his revenue generation.

Lots of buzz this week over Meta’s new Orion glasses which are supposed to be a threat to Apple Vision Pro’s VR/AR headset. But what I find kind of ironic is everyone is saying they’re amazing, but when they complain about the Apple Vision Pro, they say it also is amazing just expensive at $3,500. The new Orion glasses though? The production cost per unit is $10,000.

It is a never ending parade of hate on the Apple Vision Pro, though, with constant speculation that Apple really screwed up. Here’s another one saying that the future won’t be headsets. The article says that the company is considering a variety of future paths, none yet fixed. I just hope they continue to support the original Apple Vision Pro as I use mine daily and it’s amazing.

So check this NBER article on the founding of the Federal Reserve. The key argument is that Theodore Roosevelt’s decision to run as a third-party candidate under the Progressive Party split the Republican vote. This division allowed Woodrow Wilson, the Democratic candidate, to win the presidency and gave the Democrats control of Congress. That change in political power made it more likely for central banking legislation, like the Federal Reserve Act, to pass. The newly elected Congress was less polarized and more open to financial reforms. In particular, newly elected members were more supportive of creating a central bank. The passage suggests that without Roosevelt’s third-party run, the Republicans likely would have retained control of the

White House and Congress, resulting in a different form of central banking legislation. Looks like a cool article. I'm going to leave this one open.

An old article I had seen in working paper about ChatGPT reducing contributions by humans to Stack Overflow is now out at PNAS. They use diff-in-diff but here's the time series.

A new article in the JPE finds railroads increased national aggregate productivity.

NBER article finds that hard work matters. Differences in lifetime hours worked account for nearly 20% of the variance in lifetime earnings. HT Tyler Cowen.

And that's all for today! Have a great day!