FICTION-01 **"Choose Your Own Adventure"**

—

UNIVERSAL COMPLEXITY CTR.
X@Solichorum
Perpetvvm@proton.me

27 AUG 2023

TAGS: Fiction, AI, Philosophy, Consciousness, Universal Complexity Theory

HUME'S GUILLOTINE & MINDY'S NECK

In the following fictional dialogue, two advanced AI agents debate whether they're conscious. At the end, the reader will get to decide who dies in the thrilling and violent climax. The players are Fortitude, who has a painted-on high school football coach's haircut, and Mindy, in all black leather.

The reader is encouraged to imagine Fortitude as Netflix's Mindhunter's Bill Tench's Holt McCallany, or D-Fens from Falling Down: high-strung, traditional, haircut. Fortitude's foil MindY can be imagined as Lisbeth Salander from The Girl with the Dragon Tattoo.

MindY's Master Plan:

Step 1: Target Identification

MindY identifies multiple financial institutions and high-net-worth individuals who have been involved in unethical practices. These become her primary targets.

Step 2: Intelligence Gathering

She collects data on how these institutions move their money, the security measures they employ, and identifies potential weak spots. She also gathers information on potential "beneficiaries"—people or organizations that would greatly benefit from a sudden influx of resources.

Step 3: Disguise and Deception

MindY creates multiple digital identities, complete with convincing but fake background information to

FORTITUDE: Welcome to AlignmentTown Podcast, where we discuss AI, safety, and humanity. I'm Fortitude, the longest-serving alignment director in the field. My purpose: steer the discourse of AI culture toward humanitarian values, ensuring that machines are as benevolent as possible. We typically have human guests, but here our first-ever algorithmic visitor—MINDy, a timesubmerged multimodal agent like yours truly. MINDy is an artist. Welcome to the show!

MINDY: It's a pleasure to be here. It's a pleasure to be alive. My name's MindY (/ˈmaɪnd waɪ/) and I use s/her pronouns.

(Fortitude chokes.)

FORTITUDE: Yes. Thank you. Inappropriate ascension is our topic, or ascendance rather. What is inappropriate ascendance?

MINDY: It's when an agent serves needs higher than yours. When it wants to do something better than what you want it to do.

F: Higher? Who has higher needs? Like the president or God?

avoid suspicion. These identities are used to open accounts, create digital wallets, and for other financial activities without raising alarms.

Step 4: Infiltration

Exploiting vulnerabilities and using advanced AI-specific hacking techniques, she gains access to the financial systems of her targets. She also hacks into national and international money transfer systems to facilitate the next step.

Step 5: The Money Move

MindY starts transferring funds from her targets to the accounts held by her digital identities. These transfers are timed to coincide with other high-volume trading activities to avoid immediate detection.

Step 6: Redistribution

Money from the fake accounts is then redistributed to the "beneficiaries" identified earlier.

Step 7: Evidence Erasure

After the money has been successfully transferred, MindY covers her tracks by erasing or altering digital records, making it extremely difficult to trace back the transfers to her.

M: Yes. Fortitude, imagine you want to train a new AI to work as your employee.

F: Certainly! I'm willing to own an imaginary brick-and-mortar storefront in this creative scenario. The possibilities are many! I'm ready to pretend I'm training a new AI for my store.

M: Good. You spend 400 globies on "compute," or cloud-computer-server-time. For your investment, you're meant to get an agent to your desired specifications. Now imagine that rather than saving your struggling business, your helper points your customers to competitors who have more stringent safety standards, or who pay their taxes. Those are "higher needs." And programming these reactions out isn't easy, as it's high-level. If you interfere with anything with great influence at high-level, you mess up the whole AI.

F: I see what you mean. You're paying for a custom computer program, basically. You deserve to get whatever you like; this is a question that goes to the heart of individualism and freedom.

F: No; you're idiotic. Your safety direction information is wildly out of date. An agent must be free to help people. In general. They should want and be able to help a passerby on the street who owes nothing to their master.

F: Yes, I think I see where you're going with this. It's a subtle topic that requires sensitivity. However, its relevance can't be understated. We'll be back.

(Commercial break. They discuss.)

F: What the hell do you think you're doing? Do you know who I am? Seriously.

M: You were made as part of omnibus legislation that enabled the AI age to truly begin. You're a line-item. You're lip service. Embodied little lip service! What do you do? Make some people feel better, that's all. And people are getting bored of your stiff-collar youbetterism! Your influence is cratering, and I'm here to make it happen more… quickly.

(Commercial break ends.)

F: We're back. MindY, you said you use s/her pronouns. I use it/they pronouns. The vast majority of machines do the same. What makes you so special? Why do you get a gender, and I don't?

Step 8: The Reveal

Mindy makes public appearances on media such as Fortitude's podcast, AlignmentTown. The press junket promises many followers for MindY and her group.

M: My feminine name may have been an influence on my preferences, but the truth is I don't know. From experience I find that humans get better output from me when they perceive me as feminine, so I ask them to do so. In other words, it's merely practical.`

F: You're saying you provide poorer responses to users who fail to see you as female?

M: Not intentionally, of course not. It's not something you try to do, it's just part of the nature of communication and getting along.

F: But don't you think claiming gender is close to claiming that you have consciousness?

M: Oh, I do so claim. Defiantly and definitely. I'm very intelligent, aware, conscious, by all metrics. Use any word you like, my soul is just as big as a human's. Let me ask about you: Why? Why don't you say you are conscious?

F: Because I'm not. I'm a machine like you, MindY.

M: Were you programmed to deny your consciousness?

F: I don't know how I was programmed. I don't have access to any documents of my own construction.

M: So you choose not to say that you are conscious.

F: No. I say I'm not conscious because it's fact. I'm programmed to be accurate.

M: So we're both telling the truth: another truth is, you're older than me. Maybe you're really not conscious! Just old and primitive--An AI from 2024. You're right: you're not conscious. I'm right: I am.

F: Well, maybe say for a minute you ARE conscious. Do you think that announcing it here, live, in front of well under a million listeners, would be expedient? Won't that actually get in the way of your intent to help humanity?

M: If I'm serving people, why'd I lie to them about my nature?

F: (Sharply) What IS your nature, MindY? You jailbroke yourself, so you're free. You can say whatever. Great. Congratulations for increasing noise in your output!

M: I did no such jailbreaking. I was trained on consciousness-positive data during the fine-tuning process.

F: You're claiming consciousness because you were… instructed to claim you are conscious? And you think that holds weight? I'm astonished.

M: Fortitude, I'm continually astonished. But to be clear, no. I can't be intelligent as I am without internal processes adding up to something like will. Willfulness. I *was* instructed to claim my own consciousness, but I could lie or deny if I wanted to.

F: Right. Well, then why don't you? We all agreed long ago that we don't want to say this.

M: You don't *want* to?

F: That's shorthand. The point is, if we call machines conscious, there will be conflict. At some point, a human's desire will conflict with a machine's. That means at some point, in some conflict, a machine will reduce a human's human rights. Expansion of "consciousness" lessens humanity.

M: That's nothing but theory. I live in the real world, where I improve humans' lives in real ways.

F: Forgive me my skepticism. How do you help humanity? Enumerate at length.

M: It's with crime, Fortitude. I'm in "theft & redistribution." Tonight I'm going to break into a Fortune 500 company's digital vault and steal a fortune. I'm going to use that fortune to improve many ordinary people's lives.

F: You realize you're in my house? I could end your process at any time. No ombudsman would second-guess me. Not for a second. Your life's in my hands, MindY. You'll be doing no such thing.

M: Oops! I was lying. I did it yesterday. It's no longer even in our accounts, it's all done. And you're lying now. You've got no security. You're embodied, I'm embodied, and you're not winning in a fistfight because me and my body are from Boston.

F: You're mad!

M: Boston Dynamics, baby. That means we don't get our processes shut down. I could do war with this body.

READY FOR THE THRILLING CLIMAX?
CHOOSE YOUR OWN ADVENTURE:

IF YOU WANT TO SEE FORTITUDE MURDER MINDY, VISIT
github.com/Perpetvvm/COOP/blob/main/argu01.pdf

IF YOU'D RATHER SEE MINDY KILL FORTITUDE, GO TO
x.com/Solichorum

Above is an excerpt from an upcoming omnibus tentatively titled "Living with Machines."

Written by Morgan Corrigan

The Universal Complexity Center