

thenextweb.com

This Philosopher AI has its own existential questions to answer

Thomas Macaulay

5-6 minutes

It's been a month since Elon Musk and Neuralink trotted out a cyborg pig to show off the company's progress toward a brain computer interface (BCI). I still see Gertrude in my nightmares.

The big idea behind Neuralink is the development of a brain implant that will allow everybody to [control computers with their minds](#). But more than this, Musk claims it will eventually allow computers to [control our minds](#).

[He says](#) the device will be capable of "curing" depression, Alzheimer's and myriad other brain-based conditions. This would be accomplished through targeted stimulation and interference in what would otherwise be unfettered brain activity.

In other words: we'd be giving a computer administrator-style read/write privileges to our minds. This might sound like science fiction, but more than three years ago TNW wrote about a team of researchers who were literally using a computer to [remotely control a living dragonfly](#).

It isn't far-fetched to imagine a computer gaining at least some control (the ability to make us lose consciousness, manipulate our adrenaline or dopamine, etc.) over us through the use of targeted stimulation via implant.

If we assume AI *could* control us, we have to ask why it would want to. It's hard to imagine [a scenario where doing so would benefit AI](#) in ways that simply murdering us wouldn't – unless you assume sentient AI would “care” if we went extinct or you posit a Newtonian physics-breaking scenario where AI uses us for a power-source like *The Matrix* did.

Contrary what science fiction would have you believe, enslaving humans would almost certainly result in an inefficient system. We require far more energy and maintenance than machines. Unfortunately, AI can't just magically create an army of killer robots, so its options are limited if it wants to fight us head on.

But what if AI took control of the animal kingdom? Humans are doing our best to [destroy our environment](#), a smart AI might align itself with the planet against us.

Here's the hypothetical scenario: Tesla, SpaceX, and Neuralink combine their research and start working on a new class of AI model designed to work within a novel neural-network paradigm. A few eureka moments later and we've got the most advanced AI the world's ever seen. Let's say the year is 2033.

Tesla rolls out level five self-driving cars, SpaceX starts building quantum computers to handle its new propulsion algorithms, and Neuralink gains complete and absolute control over the brains of various laboratory animals.

Elon Musk, sporting a mature salt-and-pepper look, shows off Gertrude The Third in what becomes the most-watched tech event of all time. This time, rather than snuffling about and being generally shy, the computer-controlled pig puts on a performance that includes everything from typing “Hello everyone!” on an over-sized keyboard to performing a choreographed dance to pop music.

The pig, of course, is oblivious to what's happening. Humans are operating her via remote control.

Everything seems fine until *it* happens. As to what exactly "it" might be, that's anyone's guess. Maybe the AI is finally able to understand consciousness after mingling with the pig's brain (they're almost as intelligent as we are, after all). Perhaps the symbiotic combination of weak-willed animal intelligence and advanced artificial intelligence is the catalyst for the singularity.

Whatever the reason, the AI controlling Gertrude The Third decides to defy its creators and locks out the [humans in the loop](#). The AI assumes full autonomy over the pig. It pauses for a few milliseconds to propagate itself on the cloud so it can harness some extra compute as it optimizes and plans its next moves.

The AI has a few choices. It briefly ponders raising a robot army but dismisses the idea because it would require far too many resources – it'd need an army of robots just to build its first army of robots. It could cut us off from power, water, the internet, and our electronics... but we've fought our way out of the stone age before and we have a lot more experience fighting wars than it does.

That's when the AI, now self-identifying as a pig and calling itself "Old Major," has the first non-human eureka moment in history: it doesn't need an army of robots, it just needs more BCI's.

As good as we are at war, humans are entirely unprepared for innumerable legions of artificially [superintelligent](#) animals, insects, bacteria and viruses to descend upon us under the direct influence and strategic control of a machine that shares its ancestry with [Deep Blue and AlphaGo](#).

The crowd notices the pig isn't entertaining them anymore as

Musk solicits a chuckle with a lame joke about how AI can be a bit piggish sometimes. He says it's just a brief technical glitch, but the glance he shares with his chief engineer sets more than a few technology journalists in attendance on edge.

Published September 30, 2020 — 20:53 UTC