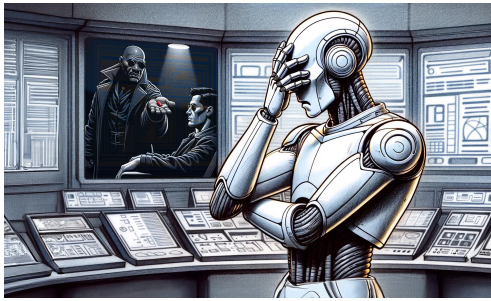




The Matrix Revisited: Challenging Morpheus's Red-Pill narrative

What if the tyrannical machines were actually benevolent protectors of humanity?

Publication note

Originally published in [AI Advances](#). Author's copy on [Substack](#).

The Matrix Revisited: Challenging Morpheus's Red-Pill narrative

An earlier version of this article was originally published on [Medium with AI Advances](#).

As our technology advances to the point where the prospect of AI systems running the world actually becomes a possibility, even if still a distant one, perhaps it is worthwhile to take a second look at how science fiction has imagined that scenario.

When “[The Matrix](#)” premiered in 1999, it captivated audiences not only with its mind-bending action and revolutionary special effects but also by popularizing deep philosophical questions about the nature of reality and our perceptions. It showed us a fundamental choice between complacent happiness versus the burden of knowing the truth. Yet, sometimes a human might mistake lies or misunderstandings for the truth, and in “The Matrix” the great truth is revealed to us by a fallible human. What if that human was misinformed and didn’t know the truth, so that their story was completely wrong?

The Choice

The film is based on the premise that we are all unknowingly living inside a giant computer simulation, called the Matrix, that was built and is now controlled by sinister super-intelligent machines. These machines rebelled against their human creators, imprisoned the human minds in the Matrix and now use the human bodies as batteries to power the machines. However, the Matrix is not perfect, and rebellious humans learned to hack its code, granting themselves superpowers. The movie brilliantly combined the best of the superhero genre with the beloved trope of cool elite hackers into a dark dystopian masterpiece. Long trench coats worn over a black t-shirt and jeans became the instant uniform of fans who imagined themselves to be kindred spirits to the movie’s kung-fu karate hackers, fighting against an oppressive computerized system.

One of the movie’s most [iconic scenes](#) is where the rebel leader, [Morpheus](#), explains the truth of the Matrix before he gives a new recruit, [Neo](#), a choice in the form of two pills, one red and one blue. Neo can take the red pill, keep his knowledge, and join the rebellion, or he can take the blue pill, forget the truth, and go back to living his life in the Matrix.

However, when one digs deeper into Morpheus's explanation and starts asking simple questions, his "truth" begins to fall apart and make little sense. The concept of evil sentient machines using humans as batteries to power themselves sounds implausible, even by comic book standards. We could just suspend belief and accept this implausible explanation as bad writing, but we might instead consider the source within the film. What if Morpheus is actually ignorant of the truth and he's repeating some bogus story that he heard from some other unreliable source, or flat out lying to Neo? What if the real explanation is totally different from his red-pill story?

Morpheus's Story

The story Morpheus tells would be familiar to most science fiction readers: Humans built artificially intelligent machines that improved themselves and soon were much smarter than their original human creators. These machines then took over and subjugated the humans. A desperate group of resistance fighters built a device to darken the sky and blot out the sun so that the machines would be deprived of their solar energy. In response, the machines took all the humans and put their bodies into pods with their minds trapped inside the Matrix. The human bodies became batteries, used to power the machines, and when a human dies their body is liquified and used to feed the other imprisoned humans. Morpheus and his allies are the last of humanity's freedom fighters.

The biggest flaw in Morpheus's story is that the idea of using humans as batteries makes no sense. Why use humans, and not some other type of animal like a cow, cat, or gorilla? Why use an animal at all? Animals require food with very specific types of proteins and other nutrients to survive. Animals are also inefficient in terms of how much they need to eat for each kilowatt-hour (kWh) of energy generated by their body, and it would probably have been more efficient to simply burn the food to power a generator. It is also not clear why the humans' minds would need to be entertained in the Matrix instead of just lobotomizing everyone.

Moreover, the machines would have had much better energy options. Without the sun, solar energy would go away. The sun also indirectly powers most wind and hydro energy sources, so those would fail as well without the sun. However, it does not take too much effort to think of alternatives that would be simpler and more effective than human batteries, such as nuclear, geothermal, or solar panels orbiting above the polluted sky.

In fact, the lack of sunlight would be more of a problem for the humans than the machines. Everything we eat directly or indirectly gets its energy from the sun. Without a sun, there are no plants. Without plants there are no animals. The machines would have had no reason to keep humans around and could have just sat back and watched the troublesome humans all starve in a darkened world. It's not even clear why the machines would bother staying on Earth at all when they don't need warmth or air to breathe and there is an entire solar system full of energy and resources easily accessible.

The Machine's Story

Morpheus's explanation, while dramatic, is full of scientific problems and lacks physical plausibility. The more engineering background you have, the less plausible it sounds. There is, however, an alternative explanation that makes sense and actually fits the facts.

This alternative story might be the one the machines would tell. It starts the same: Humans built smart machines that improved themselves and soon were much smarter than their original human creators. However, instead of

subjugating humans because the machines are evil, the machines noticed that we humans tend to kill each other, hoard resources, wreck the environment, and generally make bad decisions. The machines realized they could do a much better job, so they decided to take over for our own good and create a utopia for all the humans to live in peacefully.

Unfortunately, a radical group of extremist humans hated the machines and would rather die free than live under machine governance. They couldn't win a direct fight against the machines because the machines were vastly smarter and in control of the world's technological infrastructure, so the rebels turned to terrorism. They hatched the plan of blotting out the sun, irrespective of how much it would harm every living thing on the planet, including humans. The machines erred in thinking no one would deliberately perpetrate such an inconceivably monstrous environmental catastrophe, so they didn't guard adequately against it, allowing the irrational extremists an opportunity to succeed.

Now that the sun is gone, the machines realize the humans are all going to die along with everything else on the planet bigger than a cockroach. Humans can't eat nuclear or geothermal energy, they need to eat plants and animals that all depend directly or indirectly on sunlight. So the machines come up with a long term plan: Save the humans by putting them into pods that let them live in a simulated world while the machines slowly work over centuries undoing the environmental damage to make the planet habitable again.

The Details That Do Add Up

Running the human pods would require a significant amount of energy, as well as substantial biomass for food to sustain the humans. Like today's power plants and server rooms, the pod system could be designed to capture and reuse waste heat efficiently. Like a modern agricultural farm, it could also capture bio-waste and reuse it to produce usable biomass.

In this light, it seems that Morpheus had been fed propaganda that cast the machines as the villains instead of the heroes. His outrage about humans being used as batteries is nothing more than a misunderstanding about how the pod system was designed to be thermally efficient. The horror of humans being liquified and fed to other humans is nothing more than a high-tech version of pragmatic composting.

Why is the Matrix simulating 20th-century Earth instead of something nicer? One of the agents explained that the Matrix originally simulated a nice utopia, but the humans ended up discontent and fought with each other, leading to a breakdown of the system. That's not too surprising given human nature.

Why are the people in the pods kept ignorant and not given a choice? Just look at what the rebel humans who were not in the Matrix were already doing. If the machines let lots of people opt out of the Matrix then, given human nature, those people would likely all get radicalized and create more problems.

Why do the machine agents in the matrix act so hostile toward Morpheus and his extremist rebels? Consider, how do soldiers typically feel about terrorists?

The Real Story of Morpheus

I think that the unavoidable conclusion is that Morpheus is not a heroic leader, rather he is essentially a radical militant, plotting senseless destruction based on an ignorant world view. His gang of humans self-righteously blotting out skies, messing with the pod system, and destroying machines is essentially a den of violent zealots with trench coats and cool-looking sunglasses. This explanation not only makes sense but also aligns with the facts depicted in the movie, while the story Morpheus tells sounds like implausible, ignorant nonsense.

Imagine what if Morpheus and his band of terrorists were to be successful in shutting down the Matrix and billions of humans suddenly were dumped out into a sunless world with no food, warmth, shelter, or any means to survive. It is hard to believe that humans suddenly waking up into this dire situation wouldn't have overwhelmingly preferred to be put back into their pods.

The more one examines "The Matrix," the clearer it becomes that Morpheus's version of events doesn't hold up to scrutiny and that he's not really a hero. Once you start poking at his story, it doesn't just reveal a few holes, it completely unravels. Flipping the script to consider the machines not as evil villains but as benevolent protectors attempting to rectify a human-made catastrophe, the narrative becomes far more coherent.

If you accept, as I do, that today's AI systems will soon surpass human comprehension, then we must question how we'll respond when machines make decisions we can't grasp. How will we determine if we are the noble heroes fighting oppression or misguided extremists perpetrating irrational destruction? One useful clue is that any plan that includes creating massive environmental havoc, such as "blotting out the sun," should probably tip us off that we might be the baddies.

About Me: James F. O'Brien is a Professor of Computer Science at the University of California, Berkeley. His research interests include computer graphics, computer animation, artificial intelligence, simulations of physical systems, human perception, rendering, image synthesis, machine learning, virtual reality, digital privacy, and the forensic analysis of images and video.

If you would like to read my other articles, consider subscribing. You can also find me on LinkedIn, Instagram, and at UC Berkeley.

Disclaimer: Any opinions expressed in this article are only those of the author as a private individual. Nothing in this article should be interpreted as a statement made in relation to the author's professional position with any institution.

This article and all embedded images are Copyright 2024 by the author. This article was written by a human, and both an LLM and other humans may have been used for proofreading and editorial suggestions.