*Name:*____________________________________
*ID:* ____________________________________

# ECS 089: Ethics in an Age of Technology
## Report 3
## November 15, 2024

According to the article “Why Kant would not fear AI”, Kant's philosophy suggests a fundamental difference between human cognition and artificial intelligence. Analyze this distinction. In your response, you may:

- Evaluate the implications of this philosophical perspective for current AI development
- Consider whether you agree with the author's interpretation of Kant's relevance to modern AI concerns

*Name:*____________________________________

*ID:* ______________________________________

*Name:*____________________________________

*ID:* _____________________________________

# **Why Kant would not fear AI**

*By William Egginton. Published in Time Magazine, August 29, 2023.*

The philosophical world is busy making plans to mark the 300th birthday next year of the German philosopher Immanuel Kant. Non-philosophers might be forgiven for wondering why they should care about the opinions of a man who lived before the onset of cars, computers, and climate change. But arguably the most important thinker of European modernity had insights that can still illuminate some of our most vexing problems.

Take the wide-spread concerns about AI that have emerged full force with the development of generative language models like ChatGPT-4. Kant's understanding of the nature of human intelligence can help us work out what, if anything, we have to fear in the face of machines that write, reason, and create exponentially faster than we can.

Specifically, Kant's philosophy tells us that our anxiety about machines making decisions for themselves rather than following the instructions of their creators is misplaced. From a Kantian perspective, this fear derives from the misconception that there is no fundamental difference between humans and machines, that just like a machine, a human being's intelligence amounts to following a series of instructions, just one so complex that it creates the illusion of autonomy. But long before the advent of computing machines, Kant realized that human cognition could not be reduced to following instructions, no matter how complex.

Here is how he worked it out. A being whose cognition consisted entirely of a set of instructions could, by definition, make no distinction between the instructions and an experience of the world outside those instructions. For such a being, its knowledge would always equate to its world, with no difference between the two. And yet, Kant saw, such a model of knowledge emphatically contradicts everything we know about our experience.

Let's take the case of memory. You remember a beautiful day you spent as a child. You wish you could relive it. Now let's push the scenario a little further. You do relive it. Exactly as it was. But you can't, because reliving that day exactly as it happened would involve erasing the very difference that permits it to be a memory in the first place. Your memory in this scenario is pure information, but without a lived experience to contrast it to, you wouldn't be reliving it, you would be living it, again, without any distance or ability to know if you were living it for the first time or for the 500th time.

This paradox of a perfect memory haunts every attempt we can make to imagine cognition arising solely from information. Just as remembering the past requires a modicum of a self-suspended at a distance from that past to relate to it as past, perceiving the present requires a modicum of a self that is not utterly immersed in that present, without which there would be nothing to synthesize disparate moments into a coherent experience.

It is this indelible aspect of our cognition that ensures our fundamental difference from machines, or from any being we can imagine that processes information but doesn't have a sensual experience of the world. And it is this existential fact of human being, that we are embodied users of information, that accounts for our ability to make choices in a way that fundamentally differs from a machine following instructions.

The suspension of the self at a slight distance from its immersion in the sensory world is what allows us to perceive anything at all, and also ensures that we may pause, consider what we have seen, heard, or felt, and ponder alternatives. It allows us to both be in the present and at the same time compare the present with our memory of the past, regret paths not taken, and compare possible ways forward. When faced with a choice of roads to take, beings like us immediately feel the pull of the road not taken, a road we can only barely make out. Like Robert Frost's walker in the woods, we may keep "the first for another day," but we doubt if we should ever come back.

Unlike us, an algorithm selecting a most-likely next word or a program calculating the best move in a game of chess isn't choosing and can't feel regret. It hasn't chosen because, since its information is the same as its reality, it has already explored all available options; it has already traveled down all roads.

Doubtless machines can and do mislead us into thinking that they are performing such cognitive functions as choosing what to do or say. But only a machine that represents the world in code while at the same time sensing it physically—and that experiences the difference between those two—could be said to be making decisions as opposed to just following instructions.

While nothing in Kant's philosophy says that machines having such abilities is impossible, what he does tell us is that they won't get there by following ever more complex strings of instructions or by crunching vast stores of data to calculate the likelihood of the next word appearing in a sentence. In other words, if machines ever do develop the ability to choose for themselves, they won't be anything like today's generative AI programs.