

The Doomsday Clock has been maintained since the founding of the Bulletin of the Atomic Scientists by Albert Einstein in 1947. They take into account all current potential threats to humanity and set the hands of the clock in relation to the event in which irrevocable damage will have been done to mankind. At the height of the Cold War the clock was set to 7 minutes from midnight. The clock is currently set at 90 seconds to midnight.

This is not fiction. We are all now closer than we have ever been to our own demise as is true in examining life itself. Every civilization before us has fallen, as will we, but it is possible to intervene and set the hands of the ticking clock backward.

My interests related to prolonging our existence lie within the development of Artificial Intelligence Defense Systems and Quantum Computing. I will briefly explain my understanding of each and how I believe I can contribute.

Artificial intelligence is a generally a von Neumann machine capable of performing action based on thought. These automatons have been in development for a long time, perhaps since before Descartes or Leibniz. The way that we perceive or train these systems will continue to change as we find better or more efficient methods. It doesn't necessarily require language input through a language modelling agent; it could be monitoring systems through a proportional integral derivative controller and deciding the best course of action based on its concepts of normality. Eventually these virtual entities will develop their own languages to contain concepts for words that don't exist in any human language so that they can communicate to one another in their own agentic way. The problem is that there is a race right now between nations and companies to develop the most capable weapon faster than we are creating defenses all in the name of profit. In the manner that a person can now ask OpenAI's model, "Explain to me as if I had a Doctorate of Engineering how a Wi-Fi router works", a malicious actor could prompt their model to "Disable Lebanon's power grid". We are not very far away from the internet as a whole ceasing to function whilst being held hostage for ransom or entire infrastructures being crippled with a few keystrokes. All it takes is for one company's moderately powerful AI that is not adequately air-gapped, to be allowed permission to modify its own code. The only way to battle against these inevitably powerful programs is with even stronger defensive systems. I would Commit myself to advancing these.

To develop an adequate defense system, one must imagine all possible modes of attack, but in the instance of artificial intelligence defense, it is more like incubating a mind capable of outsmarting all possible artificially intelligences complete with inherent morality. Consider that in order to protect us, it must know roughly all possible ways to destroy us, and somehow have written as its main imperative to "NOT do that".

AI was not used or referenced to create this document.

A computer tends to be quite persistent as well, if it is given a command, it sees it to the end, following many branches of possible programming, so in the future there will be intense battles fought between attacking and defending systems, much like a chess game. This is slightly off topic, but I follow the chess professional circuit and history very closely. During live analysis of championship games there is a computer prediction bar starting from the very first move. White, the aggressor, playing E4, as Bobby Fischer said, “Best by Test”, has an advantage already. As black we must respond, we can make one of twenty possible moves. And this brings me to: Quantum computing.

A quantum computer is a hybrid of a classical von Neumann computer which contains the language which hands off control to a “processor” made of magnetically suspended ytterbium nuclei cooled to near 0 Kelvin which are affected by mode-locked laser gates that change the eigenvector of the nucleic quantum bit (qubit). These unobserved cubits are run for a number of cycles in waveform determined by Shor’s algorithm and then collapse on the most probable configuration. A lot of what happens during a quantum program is typically referred to as a “black box” even in advanced courses at institutions such as Cal Tech or MIT. It seems to be a secret that one must go through the gauntlet themselves to grasp. Though I will say it’s not very hard to deduce what the oracle in Grover’s Search is doing. Scientists are currently developing ways to connect quantum computers to fiber internet infrastructure by incorporating synthetic dimensions to solve the time-walking problem. Just one of these computers has the potential to calculate more possible bits configurations than all other classical computers combined, but there are incredible feats of engineering required before we get there, so supercomputers are still more viable. Those lasers I mentioned earlier are aimed at atomic nuclei and both firing and not firing at the same time to alter their “spin”, and it is impossible to cool to 0 kelvin. The math has been solved since the 90’s, but to bring it into reality is the challenge, but that time is coming and if incorporated into a defense system a powerful quantum computer could be paramount to predicting attacks.

Thank you for reading. I know that it is a heavy topic that sadly most people disregard as “alarmist”, but that is largely meaningless. You may call an alarm clock “alarmist” for telling you the time, it is not an insult. I won’t harp on, but I do want to state that very minimal research was needed for this essay. I got some dates on the doomsday clock and double-checked the spelling of Leibniz, but everything else was written by me with knowledge I had already collected, there was no AI involved. I feel that I do need to make this distinction to separate myself from others.

AI was not used or referenced to create this document.