An Excerpt From

The Physics of Superheroes Goes Hollywood

By James Kakalios

Is JARVIS the Singularity?

Back in the 1950's, computers were sometimes referred to as "thinking machines" or "electronic brains," but even today, the type of 'thinking' they are capable of is quite simplistic. Computers are essentially devices for the storage and manipulation of numbers and, at a fundamental level, computers can handle just two numbers: "one" and "zero," represented by a large or small current flowing through a transistor. While this seems limited, a two-digit syntax, as in the dots and dashes of Morse code from the days of the telegraph, can express quite a bit.

Artificial Intelligence (A.I.) programs such as ChatGPT operate as "auto-complete" programs on super-soldier serum. If I were to type: "The greatest superhero of all time is_____," the program would search the internet for characters who have appeared in similar sentences. One obvious approach is to find the name that appears most often in conjunction with "greatest" and "superhero." However, when used to create general text, this approach can yield awkward and repetitive sentences and paragraphs. A simplified picture of how A.I. determines the missing next word uses a concept borrowed from the physics field of Statistical Mechanics, introducing the notion of "temperature" into the search.

As an illustration, consider a marble in the bottom of a cup. As I shake the cup, the marble will bounce around, and the more aggressively I shake, the higher the marble will rise on average

as it collides with the walls of the cup. Let me characterize the shaking amplitude as the "temperature" of the cup. The lowest energy state for the marble is to just sit at the bottom of the cup, and I would call this being the zero-temperature state, while higher temperatures lead to more and more extreme excursions of the marble. For ChatGPT, the zero-temperature state corresponds to selecting the number one most popular next word found from the internet search. By varying the "temperature," the algorithm can move deeper down the list of possible next words, ranked by their frequency of occurrence. In this way the next word generated in the sentence more closely mimics text plausibly written by a person. While this can yield impressive results, it is not yet at the stage of human intelligence.

In the MCU 2008 film *Iron Man*, we first meet Tony Stark's most impressive invention. I don't mean the iron man suit he built in a cave with a box of scraps, nor the "arc reactor" that powered the suit, generating the energy equivalent of three nuclear electric power plants, yet no bigger than a hockey puck. Rather, Tony Stark's most impressive invention is the Artificial Intelligence system known as JARVIS.

The fictional JARVIS is far more sophisticated than ChatGPT, in that it anticipates needs and takes actions based on changing circumstances. In the MCU *Avengers: Age of Ultron* (2104), the artificial intelligence Ultron has been given a mission statement of protecting the world (in Tony Stark's description, "peace in our time"). Ultron, using the entire internet as its training platform (Yikes!), reaches the conclusion that the best way to protect humanity is to destroy it. Ultron thereby tries to hack into the nuclear launch codes, but his efforts are thwarted by JARVIS, who anticipated that this is what Ultron might do and uses its own hacking ability to continually

change these codes, preventing their use by Ultron. JARVIS did not need a prompt to know what needed to be done, to figure out how to do it, and to take the initiative to do it.

Human level prediction is not just an auto-complete on a grand scale, but is central to decision making, motor control, mental states, and other forms of information processing. Moreover, the role of emotion in human-level intelligence, even for activities such as solving a mathematical equation, cannot be discounted.

Before we can write a computer program that exceeds human intelligence (termed by some the "Singularity"), it would be good to understand how our minds emerge from biological, chemical, and electrical structures and interactions. Signals in the brain result from ionic currents that exist in space and time. We are just scratching the surface of learning what these time-varying signals represent. Absent this understanding, we are limited to ever more sophisticated auto-completes. Which is impressive and useful, to be sure, but does not represent a new era in the history of civilization. Back in 1979, a presentation at IBM included a slide stating: "A computer can never be held accountable. Therefore, a computer must never make a management decision" – a sentiment that remains true today. Regardless of when or whether the Singularity arrives, there will always be a need for human intelligence.

And if it ever is possible to replicate the human brain, and have a fully functioning super-intelligence, we would still have to address an important question: who's going to pay for it? Before these A. I. systems are made available to the public, the neural nets must be trained, often on examples numbering in the five to six figures, running on servers that use a lot of electrical

energy. In 2024, Google reported that its carbon emissions had increased by nearly 50% compared to what they were in 2019, primarily due to the energy demands of A.I. programs. There is a very small, but not zero, energy cost to change the ability of a transistor to carry current, the fundamental step in any computer. However, when a single microchip has billions of transistors, which are switching back and forth hundreds of billions of times a second, these energy requirements add up very rapidly. What we really need is a breakthrough in developing the Tony Stark arc reactor. In the meantime, we may have to choose – do we want A.I. or A.C.?