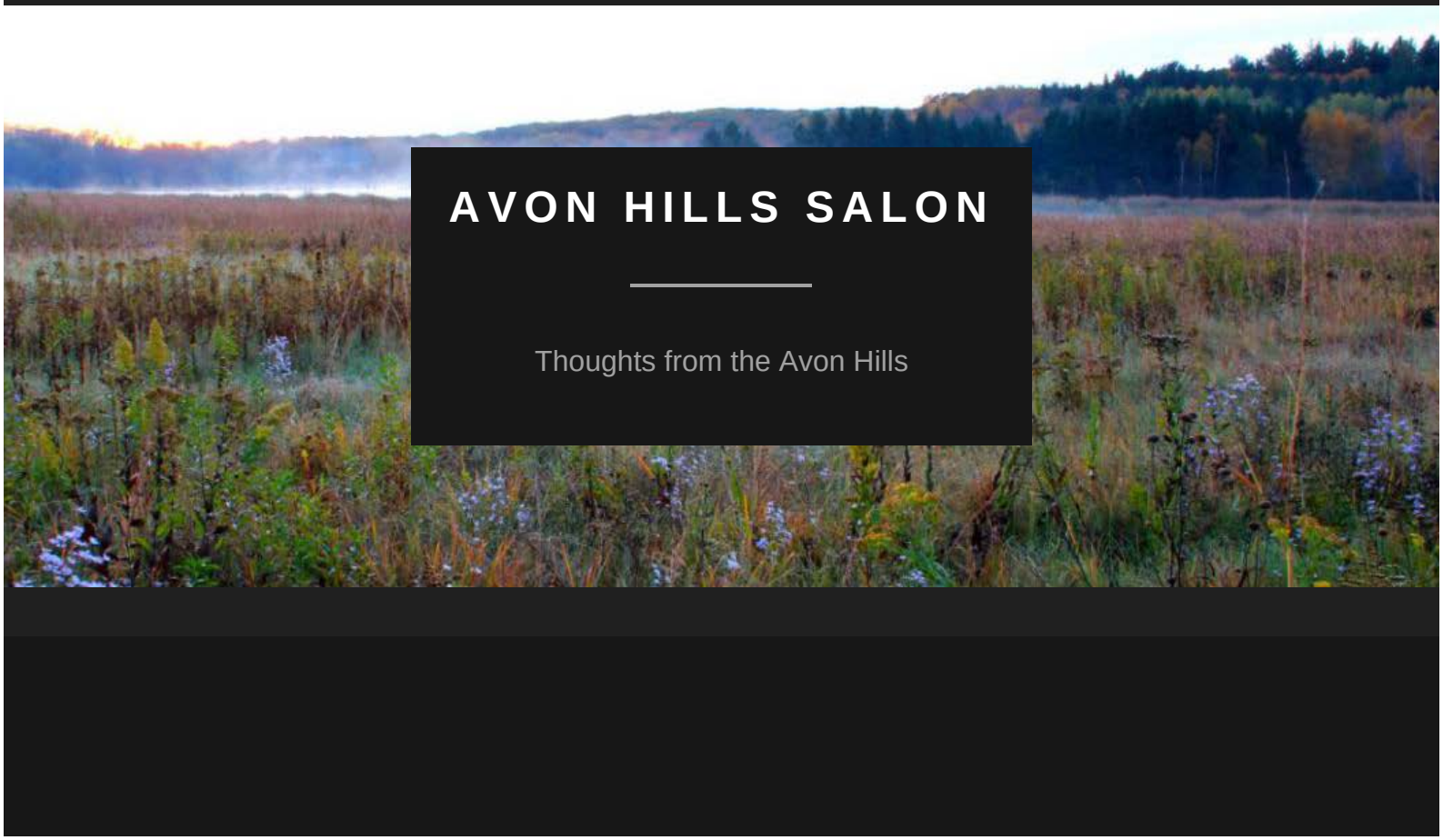
6-25-2018

The sorcery of Artificial Intelligence (AI)

Noreen L. Herzfeld
College of Saint Benedict/Saint John's University, nherzfeld@csbsju.edu



Noreen Herzfeld on The Sorcery of Artificial Intelligence (AI)

JUNE 25, 2018 / LMOE

When I think of Artificial Intelligence (AI) the image that first comes to mind is is "The Sorcerer's Apprentice" in Walt Disney's *Fantasia*. Mickey, left with the task of filling the workshop water tank, pages through a book of magic and casts a spell on a broom, giving it the task of toting the water from well to tank. Relieved of his chore, Mickey goes to sleep dreaming of power and glory, while the broom dutifully brings in bucket after bucket of water. The broom, having but one instruction, brings in more and more wat

flooding the workshop and waking a hapless Mickey, who does not know how to stop it from its single-minded devotion to its task.

In an article entitled "The End of the Enlightenment," published in the June edition of *The Atlantic*, Henry Kissinger fears that AI might bring a similar tragic result. Kissinger begins by noting the basic flaws we already experience in the Internet age—how computers lead us to treat people as data, overwhelm us with too much information, separate us by catering to our preferences, and provide an all too tempting diversion from deep thought and reflection. "The digital world's emphasis on speed inhibits reflection; its incentive empowers the radical over the thoughtful; its values are shaped by subgroup consensus, not by introspection." Kissinger then turns his lens more specifically on AI. Here, he makes three key observations. He also makes one key mistake.

First, what Kissinger gets right. After nodding to the possibilities for "extraordinary benefits" in medical science (AI is already better at detecting cancer than many clinicians), clean-energy provision, and other environmental issues, Kissinger warns of AI's potential for unintended consequences, especially those that may arise from the inability of an AI to contextualize. Like Mickey's broom, which was told nothing about the size of the water tank or the undesirability of a flooded workshop, AI may not be able to "comprehend the context that informs its instructions." He asks, "Can we, at an early stage, detect and correct an AI program that is acting outside our framework of expectation? Or will AI, left to its own devices, inevitably develop slight deviations that could, over time, cascade into catastrophic departures?" The latter is, perhaps, what should worry us most. As Sir Nigel Shadbolt, professor of computer science at Oxford, recently noted, "The danger is clearly not that robots will decide to put us away and have a robot revolution. . . If there [are] killer robots, it will be because we've been stupid enough to give it the instructions or software for it to do that without having a human in the loop deciding."

Second, Kissinger worries that AI is likely to change our own thought processes and values. He notes that the recent champion Go-playing program, AlphaGo, does not play the way humans do and suggests that AI has changed the nature of the game in that "winning" no longer seems tethered to strategies we humans have thought of, strategies that seem also to apply to other parts of life. Though he does not say it outright, it seems easy to surmise that AI could easily change the way we think about a number of human endeavors. My fear is that, just as Go might be reduced to "winning", so in other areas the single-mindedness of AI, like the single-mindedness of Mickey's broom, might narrow the way we think of our tasks, and our

world. Mickey never thought about the exercise he was losing or the joy he might have found in going out to the well and looking at the sky.

Third, Kissinger rightly notes that machine learning programs have a certain opacity. We start them up and evaluate them on their results, but we do not in the end know exactly how they reach the conclusions they do nor what they have learned. The classic story from the early days of machine learning is of a program devised by the Department of Defense that was given the task of learning to locate hidden tanks. The machine got quite proficient at identifying all the pictures with tanks in its initial set, but when given a new set of pictures, totally failed. It turned out that the photos in the training set harboring hidden tanks were all taken on cloudy days. The machine had learned nothing about tanks, but knew how to distinguish a cloudy from a sunny day. Whether true or apocryphal, this story illustrates how machine learning programs may reach conclusions that we do not understand. Explaining those conclusions is often a more challenging task than reaching them, one we may not choose to bother with. Kissinger writes, "[AI] algorithms, being mathematical interpretations of observed data, do not explain the underlying reality that produces them. Paradoxically, as the world becomes more transparent, it will also become increasingly mysterious."

Will all this bring an end to Enlightenment thinking? Kissinger sees the last 200 years as a time when humans moved from reliance on faith and authority to reliance on reason. However, in a world that has seen fascism and communism rise and fall, one busily producing leaders such as Donald Trump, Vladimir Putin, and groups like ISIS, I suspect reason's supremacy over faith and authority has been tenuous at best. AI's effect on this has, so far, been minimal.

Kissinger goes one step too far. He ascribes computers with agency: "[AI] goes far beyond automation as we have known it. Automation deals with means; it achieves prescribed objectives by rationalizing or mechanizing instruments for reaching them. AI, by contrast, deals with ends; it establishes its own objectives." Really? Not any AI I know of. We tell AI what to do. Without significant breakthroughs in our understanding of both consciousness and emotion, AI will not and indeed cannot have volition, for volition depends on both knowing what we are doing and wanting to do it. AI can do neither.

To many, AI is likely to be as inscrutable as the spells in the sorcerer's magic book. We know it works, but we don't know how—thus we may find it as hard to control as Mickey's industrious broom. The broom had no intention of causing trouble. It did what it was told. AI will do the same. The problem is that we, like Mickey, are filled with dreams of power and glory while

being mere beginners in casting our spells over our mechanical servants. There will be unintended consequences, challenges to our way of thinking, and an element of mystery. We had better stay awake.