Tech Ethics: Sample Essay #2

The use of predictive algorithms has permeated many aspects of our lives—from the trivial movie recommendation, to tasks like criminal sentencing that have human lives at stake. Many of these algorithms employ machine learning methodologies, like deep learning, that are difficult or impossible to audit or explain. Scholars like Jonathan Zittrain, a law professor at Harvard University, warn against the use of these "secret algorithms" in areas with significant human consequences until or unless we can address the "intellectual debt" that comes from using something that we do not fully understand. While we as a society should make an effort to improve our understanding of these algorithms, and thereby limit their risks, I will argue that we should not postpone their use until that understanding is achieved. In fact, nearly *all* of human action depends on the phenomenon Zittrain describes as intellectual debt, which makes the choice to forgo a demonstrably better algorithmic solution on the basis of incomplete information—*especially* to a problem like sentencing where the stakes are high—irrational and to the likely detriment of humanity.

In a 2019 article in the New Yorker, Zittrain describes intellectual debt as the accumulated cost or risk of using a solution for which "we" do not fully understand the causal relationship between it and the outcome.[1] On the surface, Zittrain's "we" is ambiguous given that we all—as individuals—use most things without much understanding of how they work. While the accepted medical explanation of Aspirin's effectiveness may have been discovered in 1995, I myself am not aware of it, nor do I acknowledge any debt associated with using Aspirin without that understanding. Zittrain would likely counter that the "we" in question is a *societal* we. While I do not know how Aspirin works, somebody does, and that protects us all from the burden of intellectual debt. Even so, Zittrain still profoundly overstates the extent of this societal knowledge and understates the prevalence of human actions done without a complete understanding of the causal chain—his intellectual debt. When I study for an exam, I do so with the expectation that it will improve my performance on the exam. I may have some understanding of how reading relates to memory, which relates to recall on the exam, but even the most advanced neuroscientists in the world are far from having a complete picture of how these pieces function and fit together; how the human mind functions is still largely unknown.

In the "physical" world, philosophers such as David Hume have likewise called into question our ability to truly understand or observe causality. While I can say that one pool ball consistently moves when it is hit by another pool ball, I cannot observe the one ball *causing* the other ball to move. Even if I refine my observational tools to use microscopes, cameras, etc., and study physics for decades, my understanding will always reduce to more and more granular cases of "when x happens, y happens." So, when Zittrain argues that machine learning algorithms are "statistical-correlation engines", a fair response is "so are we." Humans function not because we have solved the majority of the world's mysteries down to the level of pure causation, but because we are willing and able to act based on patterns and perpetually incomplete information. Put differently, to argue that we should only act when we have perfect causal information is to argue that we should never act at all.

[1] Jonathan Zittrain, "The Hidden Costs of Automated Thinking," The New Yorker, July 23, 2019.

Commented [BP1]: This is a strong thesis. It **takes a position** on the issue in question *and* it **identifies the core reason** that will be developed in support of the paper's claim.

Also note that, essentially, there is only one reason offered in support of the thesis. That one reason is developed in great detail over the course of the thousand word essay. This is a virtue of the paper! **Fewer reasons more thoroughly explained usually trump many reasons thinly explained.**

Commented [BP2]: The theoretical material a paper draws on should always be explained in enough detail for it to be clear how it is supposed to apply in context. **Make sure to fully explain any theoretical material from the readings that you draw on, rather than assuming the reader is already familiar with it**. This will demonstrate your understanding of the material and make your overall argument clearer.

The question the paper is responding to ("1b") is asking about Zittrain's concept of "intellectual debt," even though the prompt does not use the term. Given this, directly addressing Zittrain's argument is a good idea. The author rightly begins by making explicit what Zittrain means by the concept before going on to challenge Zittrain's concerns. This section would be stronger, however, if the author also articulated what Zittrain thinks is problematic about intellectual debt.

Commented [BP3]: This is the main move of the argument, and its **originality and insightfulness** are key to making this an "A" level essay.

Of course, Zittrain does not argue for perfect information, but rather to exercise caution when making decisions—like criminal sentencing—that significantly affect other people. However, *that* principle is not unique to problems involving machine learning; we live in a world filled with uncertainty, and we benefit from trying to understand it better before acting, but there is not a categorical difference between algorithm-driven decision-making and everything else when it comes to explainability.

Even if we accept a general principle of caution with regard to secret algorithms, Zittrain's claim that we should more readily apply machine learning to pizza recipes than to medicine or law given the stakes involved implies, counterintuitively, that we should put our best foot forward on more trivial problems, and wait on more important ones. As Zittrain himself points out, humans have learned incredible amounts from trial and error, so while postponing may spare some victims of the early "errors", it could also deprive later generations of the benefits of a more mature version of the algorithm.

In practice, there are a number of valid concerns with applying algorithms to the particular problem criminal sentencing. One pointed out by Sonja Starr is that longer prison sentences—like the ones that could be recommended by an algorithm-driven risk assessment—can themselves increase the likelihood of recidivism, creating a self-fulfilling prophecy.[2] While Starr's objection is well taken, it is worth noting that this is primarily a critique of the prison system and its ability to fulfill the different functions we would like it to, not the method we use to sort people in and out of it. If we imagine that we have identified and constructed the perfect set of institutions to handle punish, contain, and/or rehabilitate different types of criminals, the value of an algorithm that can consistently and accurately group people according to the type of institution that is most appropriate for them becomes much more obvious.

Together, it seems that we should not fear secret algorithms because we do not understand exactly how they reached the outputs they did, but because the institutions and the options we have to act on those outputs are so imperfect. With that in mind, we might reasonably pursue simultaneously (1) the improvement of predictive algorithms, (2) the improvement of our understanding of those algorithms, *and* (3) the improvement the institutions that these algorithms will help populate.

[2] Frank Pasquale, "Secret Algorithms Threaten the Rule of Law," MIT Technology Review, June 1, 2017.

Commented [BP4]: Sometimes an argument faces **a likely objection that *needs* to be addressed** in order for the argument to be at least initially persuasive.

Having made their positive argument, the author proceeds to take up an unavoidably obvious counterargument. **A strength of the author's rebuttal is that it continues to develop the core claim** that acting in the presence of intellectual debt is unavoidable.

Commented [BP5]: This is another important ethical consideration that lends strength to the overall argument. What is nice about offering this reason is that it flows directly from the logic of the core argument of the paper. In this way, it shows the strength of the focal argument.

Commented [BP6]: This paragraph offers another counterargument and an effective rebuttal, but it is less naturally related to the main argument of this paper. As a consequence, the paragraph as a whole adds little to the persuasiveness of the essay. Remember that **the best counterarguments are directed at the *reasons* upon which your argument relies, rather than the *position* you take on the issue in question**.