Jarett Malouf
Jarett.malouf@yale.edu
CS 201
Stephen Slade
Due January 30, 2018

Reading Response: "AI, Explain Yourself" by Don Monroe

a. The notion of artificial intelligence is not a foreign one to me, nor is it to anyone born after *The Terminator* came out in 1984, or even before then — after Brian Aldiss published "Supertoys Last All Summer Long" in 1969. It is a familiar and yet disconcerting field on the frontier of both traditional and social science, blending elements of psychology with computer science, moral ethics with engineering. I've programmed bots in the past to perform certain relatively simple tasks, and so I have some degree of hands-on knowledge about the field of artificial intelligence. I am rather distrusting of it, as I do not like the degree of influence that much of modern technology has taken on my generation, absorbing my friends and family into their screens when they have any morsel of free time. Furthermore, I do not like the idea of attributing or installing agency into a non-human, non-sentient entity, transplanting faux consciousness into an intrinsically non-feeling species of ones and zeroes. Technology is a slippery slope, and so I fear that we've already passed the point of no return in terms of the degree to which artificial intelligence will inevitably consume our lives and obstruct the interpersonal nature of our conspecific human relations.
b. Frankly, I found this article to be rather poorly written with a weak through-line about artificial intelligence. It was neither extremely informative nor extremely cogent in the point it was trying to make, and it ended up coming off as just a series of quotes from professors in prestigious universities talking about a lot of facets of artificial intelligence that we already know about. We know the average person has very little insight into the complex goings-on of artificial intelligence. We know that *ideally*, computers will become explicable to average humans, but that in the meantime, this translation will not be a simple nor intuitive one. We know we need more clarity into the inner workings of artificial intelligence in order to be able to eventually trust or distrust it. And of course, as Schwarzenegger's timeless film reminds us, we must ensure human control of the binary beasts into which we arbitrarily insert animacy. If there is anything I learned from this

article, it's an ironic lesson in what needs to change in the communications between people working in computer science and average people. The writer of this piece was trying to explain how artificial intelligence needs to become more lucid and understandable in order for it to be trusted and integrated into lay society, whereas the author himself embodied this very problem. His writing as well as the point of his article were neither lucid nor fluently understandable, and the point of his piece got a bit muddled in his bot-like writing style. And so, maybe computer science needs to bring in some more writers and engineers with solid communications and writing skills, in order to really get at the heart of this issue.

c. I would like to know where the vanguard of artificial intelligence currently lies. I would like for there to have been some interviews with some of the top scientists in the field (as opposed to just commentaries from professors on the subject), so that I could more properly grasp and assess the looming threat or prospect of a governing body of artificial intelligence. With all of the talk about how average people must better understand the inner workings of artificial intelligence in order to deem it trustworthy or safe to be in operation, I would like to know just how dangerous this venture is projected to be. Also, I would like to know if it is headed in a direction dangerous enough to merit moral intervention — in the same way I believe that there quite possibly should have been moral intervention before the fruition of the atomic bomb in the Manhattan Project. After all, this article left more of a question mark in my head than anything, and got me a bit uneasy about the communications between the people at the frontier of artificial intelligence and laypeople such as myself who will either benefit or suffer long-term as a society from their potentially irresponsible passion projects.