Should we develop AI?

Consider a life-or-death choice scenario: in a split-second, a self-driving car is faced with the decision to either kill a crowd of people by running into them or kill its passengers by veering sideways and crashing, sparing the crowd. What does car decide to do? What factors influence this decision? As an application of Artificial Intelligence (AI), the development of self-driving cars carries a heavy weight of ethical and socio-cultural questions. In the general case, the risks and benefits of continuing to develop automated thought are hotly debated, even to the extent of casting the field into taboo. Fears of a "singularity," an exponential explosion of self-improving intelligence originating from a superhuman AI, remain prolific throughout media. On the other hand, however, the usefulness of machine learning applications, which range from diagnosing disease and assisting the disabled to logistics and manufacturing, cannot be understated. This leads to two major questions: How must we proceed with the development of AI? And by what metric do we measure the capabilities of an AI? By carefully examining the history and likely progression of AI development, it can be understood that its current applications have already proven to be beneficial for humanity, and will continue to provide new solutions to complex problems as we progress into the future. However, we must proceed with caution to avoid catastrophic misuse of intelligent systems.

Where do we draw the fine line between what is "intelligent" and what is not? The first attempt at a rigorous definition of intelligence within the context of artificially constructed machines was introduced by the famed father of computer science, Alan Turing, in his 1950 paper *Computing Machinery and Intelligence*. Turing developed a test, or "imitation game," as a metric by which to judge if a machine is intelligent. This definition, which has come to be known as the "Turing Test," follows one simple rule: If a machine, isolated within a different room and

using unrevealing methods of correspondence, is capable of deceptively convincing a human that it is also human, then that machine may be considered intelligent (Turing).

Although numerous attempts have been made, a computer that is capable of passing the Turing Test has yet to be built. The Loebner Prize even offers a $100,000 reward for building a "system capable of being indistinguishable from a human," yet the entries are still amazingly far from being indistinguishable from humans. (Smith, "The History of Artificial Intelligence"). The problem, as it seems, isn't the speed and memory capacity of modern computers, but rather the complexity of the algorithm required; the complex intricacies of human interaction are simply too difficult for even the best algorithms to imitate adequately. For a moment, consider how a computer might interpret the phrase "Time flies like an arrow, fruit flies like a banana;" procedural text analysis may run dictionary searches on the words and string together a meaning based on their definitions and parts of speech. In this case, however, the homonym "flies" is fiendishly difficult to pick out by a machine. Similar implicitly understood meanings exist all throughout language; language does not operate within the world of literal syntactic definitions as computers do.

If we have yet to implement a computer that can pass the Turing Test, does that imply that we haven't created a true Artificial Intelligence? Do we call computers unintelligent, even if they outperform humans on many tasks? The complex tasks which modern AI algorithms actually are capable of are actually quite astounding. In 1997, IBM's Deep Blue managed to defeat the world chess champion Gary Kasparov 3.5 to 2.5 (three wins and one tie out of six games), largely by brute-force searching for the best possible move out of all possible moves (Smith, "The History of Artificial Intelligence"). In a vastly different style, Marvin Minsky's SNARC (Stochastic Neural Analog Reinforcement Calculator) built an AI which could nearly

instantly navigate a maze by constructing a “Neural Network” which electronically mimics the structure of a human brain (Dodd, “Artificial Intelligence Through the Eyes of the Public”). Clearly, these forms of intelligence rival that of human intelligence within their respective tasks, but these AIs would also no doubt fail the Turing Test, as neither of them are “general” enough intelligences to adequately pose as human. It is important to understand that most modern applications of artificial intelligence are only successful at a single task for which they are designed, and thus are limited. One example of this property of modern AI given in the *One Hundred Year Study on Artificial Intelligence*, which is a “long-term investigation of the field of Artificial Intelligence and its influences on people, their communities, and society,” (Stone, "Artificial Intelligence and Life in 2030), deals with Natural Language Processing (NLP). The paper postulates that, although the pattern recognition that makes NLP capable of winning Jeopardy is impressive, its “highly tailored” software requires years of research to develop, and is not event remotely a “general” intelligence (Stone, "Artificial Intelligence and Life in 2030). A better definition of intelligence, then, is to do away with the binary “intelligent or not” label, and introduce a spectrum of intelligence based on the machine’s objective capabilities. Under this label, some semblance of “more” or “less” intelligent exists, but a case-by-case evaluation is implemented: “machine X can do Y, but not Z.” It may be argued that the concept of a true “general” intelligence, one that is capable of learning and/or computing anything, is a myth, because any finite intelligence will have certain finite capabilities.

What does this say about the future of Artificial Intelligence development? Assuming intelligence is a spectrum, it’s safe to say we aren’t at an immediate risk from AI takeovers as seen in The Matrix, as there will likely be at least a few tasks, or combinations of tasks, that humans can perform better than AI for the foreseeable future. *What is the singularity*, an article

published on New Scientist, argues that billions of years of evolution has tuned our intelligence to be specifically human, and that a comparable number of evolutionary generations subject to comparable environments would be necessary to simulate a human-like intelligence (Hodson, “Visions of the Singularity: How Smart can AI Get”). Whether or not enough computing power can be mustered to simulate such an evolutionary process is an open question, but the existence of human brains demonstrates its possibility (the anthropic principle applies here).

There are some futurists who believe in the idea of the “Singularity,” which can be defined as the point at which machine intelligence becomes powerful enough to design and build improvements to itself, or, alternately, the point at which the “level” of machine intelligence meets or exceeds human intelligence. The resulting explosion in intelligence can be likened to a high-pitched feedback loop within a sound system after a microphone is placed near a speaker. After a certain loudness threshold is met, the microphone begins picking up its own sound out of the speaker and the sound gets exponentially louder. Something must be said, however, for the iteration speed of such intellectual explosion. Although scary, the AI must, at some level, have a machine capable of running it, and there is a finite limit to the processing power which can be produced, even if an AI figures out how to design new hardware for itself. Additionally, like the speaker, there is a finite limit to the “loudness” that the feedback loop can reach based on the system’s available resources, such as electricity and hardware. Thus, there is “no cause for concern that AI is an imminent threat to humankind” (Stone, "Artificial Intelligence and Life in 2030”). This fact doesn’t disprove the possibility of a singularity, but it does limit its potential to do harm, especially in the near future. If computerized systems had a larger grasp on our industrial systems, to the point where an AI could conceivably hijack them for their resources, more cause for concern would be warranted. Instead, a more optimistic view of the singularity

exists: *Artificial Intelligence in the Eyes of the Public* points out science fiction writer and futurist Ray Kurzweil's view that "by the 2030's things like mind uploading and complete immersion in virtual reality will be possible" (Stone, "Artificial Intelligence and Life in 2030"). In Kurzweil's view, humans and computers could one day merge. Why worry about AI taking over the world when we ARE the AI?

The debate about the singularity is still very much an open question, even among experts in the field, but there is not significant reason to halt the development of a new and revolutionary technology out of unfounded fear. In fact, the *One Hundred Year Study on Artificial Intelligence* suggests that if society approaches these technologies primarily with fear and suspicion, missteps that slow AI's development or drive it underground will result, impeding ethical and social work on ensuring the safety and reliability of AI technologies (Stone, "Artificial Intelligence and Life in 2030").

Regardless of its capabilities, the development of AI still elicits questions regarding ethics, morality, and even practicality. Returning to the self-driving car crash scenario, it must be asked if a consumer would even purchase a car that was programmed to kill its passengers to save a crowd. An AI system could also conceivably be programmed (or learn to) to save itself at all costs. Such negative sentiment could impede the progress of AI development, but these types of questions can be seen as merely fine print that comes as baggage with an incredible jump in technological capability. In a positive light, the development of self-driving cars, with their superhuman sensing and reaction time will make roads much safer. In fact, it is predicted that "Self-driving cars will eliminate one of the biggest causes of accidental death and injury in United States, and lengthen people's life expectancy" (Stone, "Artificial Intelligence and Life in

2030"). If we are to maintain our industrious excitement towards AI, then the public must understand that the risks of its development outweigh its benefit.

In a survey conducted by Worcester Polytechnic Institute regarding the popular opinion of AI development, it was discovered that very few people (10%) have taken a class in Artificial Intelligence (Dodd, "Artificial Intelligence Through the Eyes of the Public"). This finding raises questions about the public's understanding of the topic. Optimistically, however, 72% of individuals surveyed indicated that the promises of AI did not scare them (Dodd, "Artificial Intelligence Through the Eyes of the Public"). This indicates that most of the public view the development of AI favorably.

As we continue to develop AI, we must measure its capabilities by our empirical observations of such. There is no easy, binary answer to the question "is a machine intelligent or not?" On the intelligence spectrum, each machine performs differently on different tasks, and its capabilities are measured instead by the question "does AI machine do something meaningful that couldn't have been done otherwise?" In many instances, from self-driving cars to playing chess, we are already noticing that the answer is "Yes". Overall, there seems to be no cause for concern with respect to the continued development of AI, as the potential benefits vastly outweigh the potential drawbacks. By carefully examining the history and likely progression of AI development, it can be understood that its current applications have already proven to be beneficial for humanity, and will continue to provide new solutions to complex problems as we progress into the future. Although the singularity is theoretically possible, technological limitations render it nearly impossible in the immediate future of AI development. By the time it does occur, if ever, it's possible that humans will have merged with computers anyway. Wouldn't we want the singularity to happen to us in that case? Some degree of caution and ethics

is necessary within the development cycle, however it is evident that the most important action to take is to educate the public about AI in order to quell misconceptions. The general outlook on the public's sentiment is indeed positive, however, so it's full steam ahead towards passing the Turing Test.

Works Cited

Dodd, Matthew, Alexander W. Grant, and Latiff Zaaliembike Seruwagi. “Artificial Intelligence Through the Eyes of the Public.” Worcester Polytechnic Institute, 4 Mar. 2011. Web. 21 Nov. 2016.

Hodson, Hal. "Visions of the Singularity: How Smart Can AI Get?" New Scientist. New Scientist, 26 Mar. 2016. Web. 21 Nov. 2016.

Smith, Chris, Brian McGuire, Ting Huang, and Gary Yang. "The History of Artificial Intelligence." University of Washington, Dec. 2006. Web. 21 Nov. 2016.

Stone, Peter, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller. "Artificial Intelligence and Life in 2030." One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA, September 2016. Doc: http://ai100.stanford.edu/2016-report. Accessed: November 21, 2016.

Turing, Alan M. "Computing Machinery and Intelligence." Mind 49 (1950): 433-60. Web. 21 Nov. 2016.