

TO: Rick Kogan, NIST
FROM: Christian Espinoza
DATE: May 10, 2018
SUBJECT: Artificial Intelligence



A.I Safeguard

Hello Mr. Kogan,

I am writing to you regarding your request for an internal proposal on what we at NIST can do to handle the rapid advance of Artificial Intelligence in the United States. This proposal outlines the current state of Artificial Intelligence, its rapid growth, and what can be done to safeguard against its potential social and economic disruption.

This proposal will provide an outline for the NIST on the current state of A.I., an analysis of the problems it has created and the future problems it could bring. It will also provide an overview of my proposed solution, if I am to be tasked with spearheading this new division at NIST. The NIST knows the impact technology can have on society; it is therefore our responsibility to make sure that we create more awareness and to start working on a safeguard to help train displaced workers to enter new fields of work.

Current Situation

The current state of Artificial Intelligence is that it is advancing at a very rapid pace; many experts believe however, that it is not evolving at a pace fast enough to be of any real threat to society.

The premise of their argument is that the biggest areas that A.I has improved in are specialized and extremely specific tasks. Examples of this can be found in AI's such as IBM's Watson who won a game of Jeopardy against a human, or in Google's AlphaGo that uses neural networks and has beaten the best human go players.

The main point that makes most experts brush off the dangers of rapidly advancing A.I., is that as of now, they can *only* do the specific tasks they are

programmed to do, meaning that although Watson may be excellent at Jeopardy, it still can't beat a mere toddler at a game of tic-tac toe.

Experts have determined that the smartest A.I. only has the I.Q. of a 6 year old. The following table breaks down the current perceived I.Q. levels of A.I. as they are now.

Subject	I.Q.
6 year old child	55.5
Google's AI	47.28
Baidu AI	32.92
Bing	31.98
Siri (Apple)	23.94

Analysis

The evidence along with *many* expert opinions, presents A.I. as an innocuous topic. Some even go as far as writing off 'A.I. Doomsayers' like Bill Gates and Elon Musk as paranoid conspiracy theorists.

Not all experts agree however, Winter Levy writes, "In 2008, a survey of experts at a conference on global catastrophic risks at Oxford University ranked super intelligent AI as the greatest existential threat to the human race, above nuclear wars, engineered pandemics, and climate change, with a 5 percent chance of causing human extinction by 2100." (Levy). It does not take much research to learn that these experts who bet against A.I. advancement, have been wrong many times before. Their predictions and projections are constantly shattered by how rapidly A.I. advances.

Elon Musk is perhaps the most outspoken advocate on safeguarding ourselves against A.I. He compares A.I. to nuclear weapons, and even goes as far as to say that A.I. is in fact, a more dangerous threat to humanity than nuclear warfare. He talks about how A.I. production is largely unmonitored and how there are no sanctions or overseeing entities to ensure that A.I. is not being used for evil. When speaking about the dangers of A.I. Musk writes, "AI is a fundamental risk to the existence of human civilization in a way that car accidents, airplane crashes, faulty drugs or bad food were not – they were harmful to a set of individuals within society, of course, but they were not harmful to society as a whole." (Musk)

Elon Musk also speaks about how if A.I. ever reaches a point where it is smarter than humans, even the most benign scenario, where robots “played nice”, would be catastrophic to humanity. We would essentially have an entity that can do anything humans can do, but better, so humanity would have no purpose in terms of work and productivity. Musk writes, “There certainly will be job disruption. Because what's going to happen is robots will be able to do everything better than us. ... I mean all of us” (Musk).

It is extremely reckless and dangerous to ignore A.I.’s potential to disrupt society; it is essentially leaving the problem for future generations to deal with. My concern is that we at The National Institute of Standards and Technology are not doing enough to create safety nets, or at the very least, create a public awareness for this problem. The loss of jobs will leave many displaced workers without a purpose in their life, and without adequate preparation to take on different job roles.

We’ve already experienced the problems technology and A.I. can bring, with Russia’s use of bots during our last election. Russia was able to make bots that used fake Facebook and other social media accounts. These accounts frequently posted extremely polarizing and divisive views under different usernames, to create turmoil among people who don’t have the capacity to think critically.

This was all before Google recently unveiled its new Duplex system. Duplex is an A.I. system that can accomplish real world tasks over the phone. These tasks can range from scheduling a hairdresser appointment for you, to negotiating a better contract with your phone provider. In a showcase of the technology, Google showed how duplex could perfectly mimic a human voice and human speech patterns. It says “uhm” and takes long pauses. They had it call a hairdresser salon to schedule an appointment for a human; the receptionist at the salon could not tell that she was speaking to an A.I. This raised many questions and concerns as to the danger and power of A.I. that can be right around the corner. Imagine a scenario where a country like Russia somehow got its hands on this A.I and was able to create more turmoil and division among American citizens, by faking calls that are filled with fake A.I. scripted racism and hate-speech.

A.I. improvements might appear to be moving slowly on the surface, but once A.I. reaches AGI or general intelligence, it will be only a matter of time before runaway technological growth happens and brings us closer to the singularity. Levy writes, "The most serious AI safety concerns arise not from narrow intelligence or even general intelligence, but rather from what many experts argue will quickly follow AGI: artificial superintelligence. Because a capacity to learn would be an integral component of any system that attains general intelligence, the AI would be able to boost its own capabilities—such an AI would be able to design an improved version of itself, and then that smarter version would do the same, and so on. This recursive self-improvement could result in an intelligence explosion " (Levy).

Recommendations

My proposed solution is a project called A.I. Safeguard. A.I. Safeguard would be an internal division within the NIST. A.I. Safeguard would team up with the *Future of Life Institute* to create a set of rules and standards for A.I. developers.

Future of Life Institute is a non-profit organization that uses volunteers. They have already started development on a safety net; we would provide them with more funding in exchange for access to their data, research and volunteers. The plan also involves setting up conferences at U.S. colleges where leaders in the field can come and talk about what they are working on and the progress they are making. Below is my Project Plan timeline.



Phase 1	Team up with Future of Life Institute.	July 4, 2018
Phase 2	Create an official NIST - A.I. Safeguard Youtube channel to create public awareness.	July 20, 2018
Phase 3	Begin working on safety net program for displaced workers.	August 1, 2018
Phase 4	Begin touring U.S. campuses and having conferences to spread the knowledge on the dangers of A.I.	September 12, 2018
Phase 5	Work with Future of Life Institute to create an entity to oversee and implement rules and regulations for A.I developers	November 7, 2018

We are already in contact with the founders of FOLI, he is open to working with us, on the condition that they get to keep their autonomy and that they are listed as collaborators and not part of NIST. There are still much more negotiations and meetings to be held, but as of now, it is looking good.

Conclusion

We need to start creating widespread awareness of the potential dangers of Artificial Intelligence as soon as possible, as of now the American public has only heard these warning from entrepreneurs, if a government agency such as ourselves creates a Youtube channel it will be more 'official' to the public. As of now, no branch of the U.S. government has come out in support of Elon Musk or Bill Gates warnings; we would be the first. As Sam Harris so eloquently states, "We are in the process of building some sort of god, now would a good time to ensuring that it is a god we can live with". (Harris)

Works Cited

GUIHOT, MICHAEL, et al. "Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence." *Vanderbilt Journal of Entertainment & Technology Law*, vol. 20, no. 2, Winter2017, pp. 385-456. EBSCOhost, csulb.idm.oclc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=127634787&site=ehost-live.

MUSK, ELON. "I Hope Artificial Intelligence Is Nice to Us." *NPQ: New Perspectives Quarterly*, vol. 31, no. 1, Jan. 2014, pp. 51-55. EBSCOhost, doi:10.1111/npqu.11427.

"The Current State of AI as We Begin 2018 | Cesran International." *Cesran.oRg*, 1 Jan. 2018, cesran.org/the-current-state-of-ai-as-we-begin-2018.html.

WINTER-LEVY, SAM and JACOB TREFETHEN. "Safety First: Entering the Age of Artificial Intelligence." *World Policy Journal*, vol. 33, no. 1, Spring2016, pp. 105-111. EBSCOhost, doi:10.1215/07402775-3545918