**Week 6: Ethical Challenges with AI**

Nate Bachmeier

April 14th, 2019

TIM-8101: Principals of Computer Science

Northcentral University

# Determine controversial area of research

Artificial intelligence and autonomous systems are defined as any component that can derive contextual meaning without human interaction (Inversardi, 2019). For many this conjures images of *Skynet* from the Terminator series, where our only hope is an implementation of Isaac Asimov’s Three Laws of Robotics.

As these systems do not require humans there is a natural fear that they will supersede their creators. The fear of losing control leads to a desire to stifle that creative force and limit the areas that are directly impacted.

One controversial area where this can be seen is the health care industry (Kostopoulos, 2016) (Matsuzaki, 2017). Health care will eventually become enriched through the adoption of AI systems however it must first overcome some key challenges in reliability, transparency, liability, security, and privacy.

# Why is it controversial

Imagine that a loved one needs to have a complicated surgery, and the doctor tells their patient that there are two choices. The first is an excellent surgeon with an 8.9% mortality rate, or a machine that commands a 1% mortality rate.

If something goes wrong the machine will not adapt to the changed conditions, nor can the patience’s family sue when things go wrong. When the patient dies who should be held accountable-- the manufacture or the hospital? Modern legal systems are not keeping pace with technological advancements which introduces challenges. Recent court cases have found that gun manufactures are not liable for shootings (Kasprak, 2018), can the same be said for autonomous devices?

When the machine decides to perform an action during the surgery, how can we know that it is the right one to make? What systems are auditing those choices and what evidence can be provided to a review board to justify those behaviors?

For these machines to make intelligent decisions they require access to large amounts of data. How does that impact the patients right to privacy versus the safety across the larger community?

This is not a new challenge as can be seen with Henrietta Lacks and the theft of her cancerous cells (Skloot, 2010). Her cells enabled numerous medical discoveries and saved countless lives, however her personal privacy was forfeit.

Assume for a moment that only a subset of patient data was made available to the AI system. This would lead to more challenges as the reliability would be reduced. How would the machine react when presented with a completely foreign configuration? The human can reach out to colleges or consult medical encyclopedias, neither of these actions are freely available to the machines.

On top of these challenges are concerns related to the security of the machines. Malicious actors could manipulate the firmware in such a way as to kill the patient (Zetter, 2014). Aside from a *Manchurian Candidate* scenario, the same is not true of the human.

# What solutions are being presented

When Joseph Sirosh became vice president of Microsoft’s AI division a Q/A session was held. One of the questions was “how do we address the privacy and ethical implications of AI systems?”

He replied that the traditional idea of privacy is dead, and the conversation needs to transition toward responsible governance of information. Adding that humans inherently lack enough entropy many common scenarios. For instance, systems can determine from a stream of GPS notifications where a given end user lives, works, and privately walks their dog.

In contrast the European Commission has recently stated that personal privacy needs to be the heart of the discussion. Systems need to be built with core tenants of transparency, accountability and oversight (European Commission, 2019). They envision a world where data lives on private islands completely controlled by the end user. For instance, user’s have a right to be forgotten and delete data they have generated.

These extremes can be modeled within the *Many Worlds Theorem*, which states that there are three worlds-- our world, the rest of the world, and the largest possible world (Onuf, 2016). Onuf proposes that each societal scope needs to handle different permissions and granularity of information.

For example, within my personal view health dashboard key performance indicators (KPI) need to be highly personalized and confidential. However, with the rest of the users of the service those results should be anonymized. If that data leaves the control of the service provider, then it is given in aggregate format.

# What is the ultimate resolution to these controversies

Businesses and government

# What are future avenues for research?

One of the biggest challenges with ethical system design is that consistent modeling frameworks do not exist. There are basic human rights that most societies agree upon, but a lot of grey area exists.

Consider the scenario where many autonomous devices are operating in a shared space. It can be proven that it is more efficient for the all devices to collaborate (Rabanal & Friedman, 2015), but there does not enforce all systems to behave as such.

There is also a disconnect that these systems are not being ethical because it is the right thing to do, they are being ethical as a means of avoiding bad. They are designed to be pro-sociality.