

# Ethical considerations for AI

By  Rudeon Snell

5 Sep 2018

There is little doubt that the pace of innovation is accelerating at unprecedented levels. Technology enabled breakthroughs are happening with increased frequency, enhancing the human lifespan, improving access to basic needs and leaving the general public with little time to adjust and comprehend the magnitude of these advances.

Within the field of Artificial Intelligence (AI), this phenomenon is certainly just as true, with the accelerated pace of AI development, generating huge interest about moral AI and how, as imperfect human beings, we are teaching AI the differences between right and wrong.

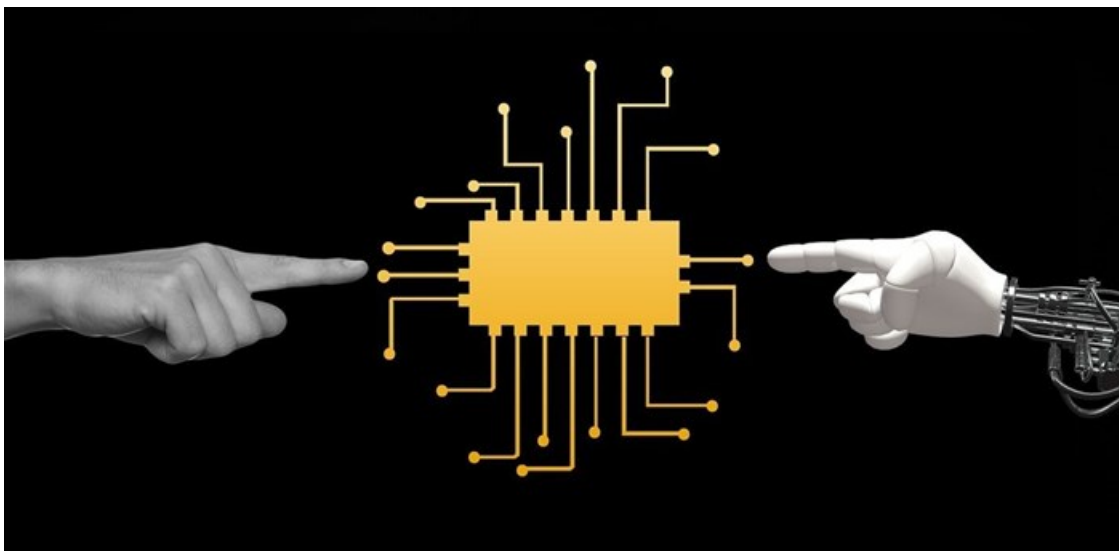
As AI systems continue to evolve, humanity will place increasing levels of trust in them for decision making, especially as these systems transition from being perceived as mere tools, to operating as autonomous agents making autonomous decisions.

The question of the ethics pertaining to the decisions that get made by AI systems must be addressed.

## Ethical fundamentals of everyday life

The question of ethics finds some of its roots in the notion of fairness. What is fairness? How does one define fairness? Instinctively, human beings grasp the concept of what is fair and what is not fair.

As an example, we commonly accept that one for me, two for you, is not fair. We teach our children about what it means to be fair, why we need to share, what the moral and ethical constructs as we believe them to be are when it comes to fairness and sharing.



Source: [pixabay.com](https://pixabay.com)

The concept of fairness also features prominently in the [United Nations Sustainable Development Goals](#): Gender Equality (Goal 5), Decent Work and Economic Growth (Goal 8) and Reduced Inequalities (Goal 10) are all arguably built on the concept of fairness.

But how do we teach AI systems about fairness in the same way we teach our children about fairness, especially when an AI system decides that achieving its goal in an optimal manner can be done through unfair advantage?

Consider an AI system in charge of ambulance response with the goal of servicing as many patients as possible. It's quite possible that it might prioritise serving 10 people with small scratches and surface cuts above serving two people with severe internal injuries because serving 10 people allows it to achieve its goal better. Although this optimises patient service, it fundamentally falls flat, when one considers the intent of what was meant to be accomplished in the most optimal way.

In business, we have ethical and unethical behaviour and we have strict codes of conduct regarding what we consider to be ethical and unethical business conduct. We accept that not everything that is legal is ethical and not everything that unethical is illegal and as a society, we frown upon unethical business conduct, especially from big corporates.

How does this transfer to AI systems? Surely, we wouldn't want AI systems that stay within the bounds of the law, but push as hard as they can against those boundaries to see what they can get away with, exploiting loopholes to fulfil their goals.

## Data perpetuating embedded bias

AI systems feed off data. If AI is the new electricity, data is the grid it runs on. AI systems look at data, evaluate that data against its goals and then find the most optimal path towards achieving those goals. Data is absolutely critical for AI systems to be effective. Machine learning (ML) algorithms gain their experience from the data they are given and if that data is biased or ethically or morally tainted, the ML algorithms will perpetuate this. What about factors that are not expressed with data, such as the value of another person, the value of connections, the value of a relationship? The biggest challenge with data, unfortunately, is that data quite simply just does not data give you ethics.

Then there's the issue of blame - who is to blame for AI making mistakes?

The manufacturer, the software supplier, the reseller, the data set, the owner, the user? The issue gets more complicated when we talk about the loss of life in an accident. Consider incidents with AI systems in healthcare and who would be legally held liable. What about autonomous vehicles disrupting the automotive industry and making its way into society sooner rather than later? If we expand on this trend, what about AI systems making decisions based on their programming, leading to them committing crimes? Are they guilty? Can an AI system be guilty of a crime? Are their programmers to blame? Their data sets? What laws govern this eventuality?



Is AI the virus killing creativity?

icandi CQ 31 Aug 2018



---

Take our smartphones autocorrect function as a simple example. I'm positive many of us have had an incident where we've sent texts to friends right after an autocorrect function changes one word to another, often to a more embarrassing version, from where we often issue some grovelling apology. The point is this; if technology today struggles with understanding the intent of a few lines of text, how can we count on it to understand and make life and death decisions?

## Revisiting classic questions regarding ethics and morality

Researchers have explored how to effectively resolve this situation in the past. The Trolley problem tests have been around since 1967. First proposed by the philosopher Phillipa Foot it has subsequently proliferated into many variants. Generally, it is used to assess what actions people would take when asked to take an action that would, for example, kill one person versus 10 people. This is specifically being applied in the context of autonomous vehicles, as a reference model to help AIs make effective life or death decisions, but it's not a foolproof solution.

Utilitarian principles could offer a framework to help with the ethical decisions AIs need to make. The focus would be on AIs making decisions that result in the greatest good for the greatest amount of people. However, at what cost? How do utilitarian calculations that violate individual rights get reconciled?

Ethics is often not about one thing or the other specifically, but more leaning towards the notion of how if you go down a particular road, that road has a particular set of ramifications. If you go down an alternative road the implications could be different. This is what AIs currently struggle with and what humans instinctively understand.

AI systems have largely been built to achieve specific goals and specific outcomes. For humans to have any semblance of creating ethical AI, AI systems should be programmed to be sensitive to achieving its goals in the construct of human values as they could achieve their goals in rather bizarre fashions.

Think about a machine deciding to protect humanity by enslaving it (The movie i-Robot rings a bell). Soft governance, industry standards, professional codes of conduct and policies. These are the considerations that must be given in order for us to understand how we can engineer AI in a safer way and how we make our values part of the design process when implementing AI systems. Who decides how ethics are defined? Who decides which ethics are applied in AI?

Ethics ultimately is embodied in knowing the difference between what you have the right to do and what is right to do. We all will need to do our part in ensuring AI systems know how to do this.

Private and public-sector organisations with all their multifarious complexities; societies, from the family to the nation; economies, from the subsistence farmer to the giant multinational - all are inherently human undertakings fuelled by desires and ideas and made possible through collaboration, conversations and amazing technologies.

## ABOUT RUDEON SNELL

Director: Intelligent Solutions at SAP EMEA South Africa

- Eliminating bias is the new frontier of AI innovation - 9 Sep 2020
- How AI-powered chatbots are unlocking business value - 14 Dec 2018
- Ethical considerations for AI - 5 Sep 2018

[View my profile and articles...](#)

For more, visit: <https://www.bizcommunity.com>