TECHNOLOGY

The Tug of War in AI Decision Responsibility

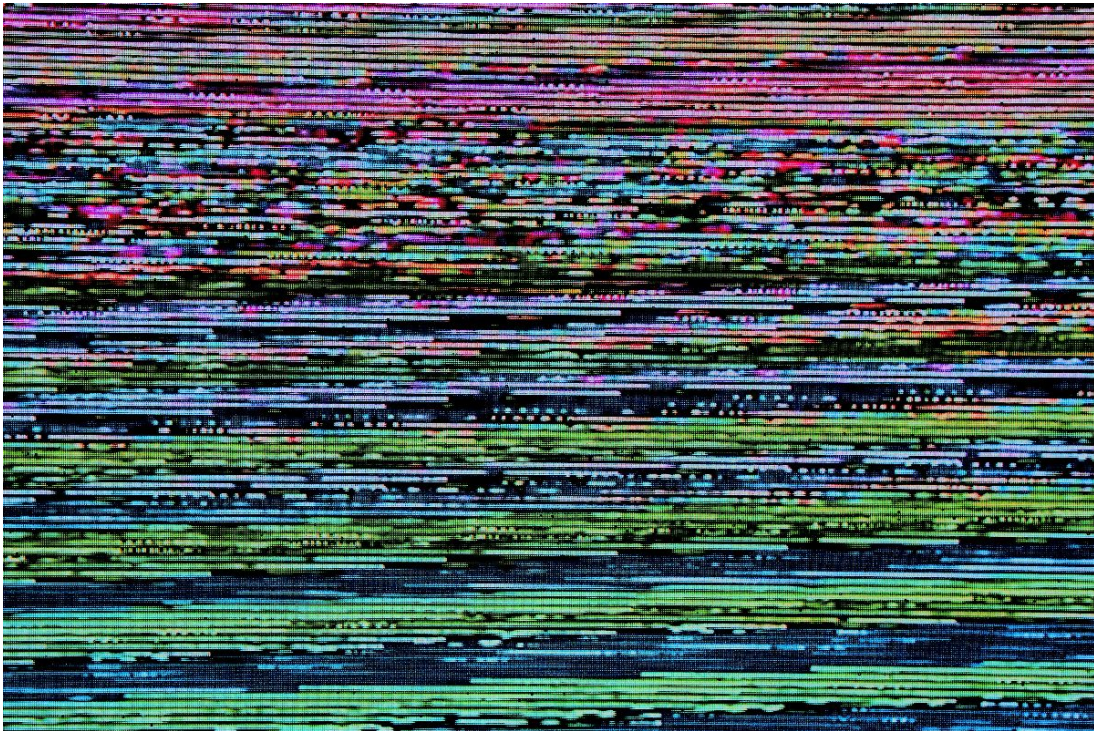by Melodena Stephens and J. Mark Munoz

Image Credit | Michael Dziedzic

Governments lag behind the private sector in their understanding of AI. What happens if an AI system fails?

☑ **INSIGHT | FRONTIER**   24 Jun 2021

With the COVID-19 pandemic, AI adoption has grown by 61-93%. A McKinsey (2018) report predicted that AI can add US $13 trillion to the global economy, with greater rewards going to countries that establish themselves as global leaders. The scale at which

AI has been deployed raises some concerns. Artificial intelligence is considered trendy, and there is a perception it saves costs, makes operations more efficient, and generates profits.

Governments however, lag behind the private sector and often have a fragmented approach to governing AI. While many ethics documents are released by the **private sector, governments** and academia (Jobin, Ienca and Vayena, 2019), a report by Hagendorff (2020) finds that only about 50% of the ethics guidelines have principles referring to human oversight and control. Much of ethics regulation focuses on the AI as an agent.

In this article, the authors focus on other agents involved in the process – humans and the institution. It also moves the discussion to important questions that are often left unanswered – who should take responsibility if an AI system fails? What are the implications on executive decision making?

Human vs AI?

As AI gains global acceptance, accountability for the growing number of errors come into question. Should humans or AI be blamed?

When a decision is taken, risk is considered. Resources are committed in the belief that there are specific outcomes. What happens if that fails? Who bears responsibility? The human behind the AI, or the human deploying the AI, or the human using the AI or the human safeguarding the AI, or the AI? Managers should deliberate on these choices before designing, adopting, deploying, or using AI systems.

The KPMG (2021) report titled *Thriving in an AI World*, found that 93% of the business leaders surveyed from the financial services were confident in AI's ability to detect fraud (an increase of 8% from the previous year). The UK had implemented a Japanese software developed by Fujitsu called Horizon in the Postal Office in 1999. From 2000 and 2014, 736 Post Office employees were prosecuted for theft. Some employees went to jail, some re-mortgaged their homes to pay for the shortfall. One woman committed suicide – all the while pleading innocent. More than 2400 employees were affected, and more than **736**

sub-postmasters and sub-postmistresses were prosecuted (Peachy, 2021) – this is the average of one a week between 2000 and 2014. In 2021, a fatal flaw was identified in the software. Soon, the courts were overturning previously meted-out sentences.

Who should bear the brunt on the decisions made? In this BBC Report (Peachy, 2021), it was indicated that "nobody at the Post Office or Fujitsu has been held accountable, although the High Court judge said he would refer Fujitsu to the Director of Public Prosecutions for possible further action because he had "grave concerns" about the evidence of the company's employees." The people who went to jail experienced immense tragedy, which money could not compensate.

Humans and machines are different. The modality, process and speed in which they process information are significantly different. Furthermore, an AI feels nothing. It cannot take moral responsibility. This moral construct is what separates humans from machines. Today, as AI gets deployed across country borders, getting accountability becomes even more challenging.

Corporation vs Society?

A growingly important issue that needs to be weighed upon pertains to whose needs are paramount, the corporation's or that of the society?

In the Post Office case, it became apparent in another BBC (2020) report that the Post Office's legal department knew that software was not functioning accurately but chose not to disclose this to save the organization's reputation. As indicated in Bidstats (2021), the Post Office decided to keep and extend the contract for the same software system by one more year to 2024 as it is a "highly complex, legacy platform, written in outdated versions of software languages, and incorporates five 'systems' in one i.e. financial services, banking, government services, mails, and retail. "Horizon" is an aging platform and has an inflexible monolithic architecture that makes technology change difficult."

In deploying AI, legacy systems are a nightmare. Legacy systems and the hyper-innovation in the AI sphere has resulted in an urgent need for people familiar with increasingly redundant programming languages like **DB2 (1983 – IBM) and Cobol (1959**) (White,

2017). In 2020, **governments and banks** were desperately searching for Cobol programmers (Lee, 2020). When new technology does not work on older existing systems, much complexity arises (i.e., mobile phone upgrades, new charger designs). Whose responsibility is it when AI fails because of legacy systems, poor purchase planning, ignorance, or indifference?

In many cases, corporations have the money to manage the fallout from such situations. Justice systems have evolved with a fundamental understanding that humans have morals and hence need to do penance via punishments for agreed wrong-doings. Corporations that take responsibility for AI, are like AI, and do not feel remorse.

Corporate misdeeds have been evident in high-profile criminal cases. In 2015, Ashley Madison, the online dating website, fraudulently used a bot posing as human to communicate with members and get them to buy products. Other instances include: Facebook's Cambridge Analytica or fake news, the case of Clearview AI that illegally trolls social media to collect data for facial recognition on 3 billion people, and the da Vinci surgical robots that made mistakes. The cases are increasing. In the case of The *United States v. Athlone Indus., Inc.,* 746 F.2d 977, id. at 979 (3d Cir. 1984) it was stated that "robots cannot be sued," but the manufacturer (company) could be liable for civil penalties. But, these liabilities are **difficult to pin down on corporations** (Quinnemanuel.com, 2016).

While governments and corporations should regulate what they produce, deploy or buy, they need the humans in these organizations to make decisions and take responsibility for the decisions. Oftentimes, hard choices have to be made. Should organizational goals be paramount or should it be the betterment of society? How can humans or the executives in these organizations grapple with this ethical tug of war?

Winning Through Mindful Decisions

The authors offer three key decision making approaches that will help overcome the overwhelming push and pull forces relating to AI ethical dilemmas:

Human-Centered Approach

AI decisions need to have a human-centered approach. Since, by default, most countries have signed the United Nation's Universal Declaration of Human Rights and endorsed the Sustainable Development Goals, this is an excellent place to start. How does the AI impact all humans – not just those that use but also those that are affected? This human-value centric approach to design and decision making is a responsibility of inventors, educators, investors, deployers, regulators, and mentors that extends through the global value chain. Not all users may have the knowledge or competence or the voice to make their concerns known, so due diligence is critically needed rather than fast scale-up. Often AI owners speak in terms of acceptable error costs – but again, if a human makes a mistake and the penalty is jail - is this fair? **IEEE**, has done excellent work focusing on the engineers in the AI value chain. However, there is a further need to extend this type of thinking to a deeper ethical context, across more markets and with multi-perspective information in order to create socially responsible products and services.

Human as the Final Decision Maker

All AI systems must by default, have a human final decision-maker who takes accountability for their decision. For example, in the healthcare systems, an inaccurate diagnosis may result in death. In **Singapore**, the Court of Appeal in ***Hii Chii Kok v Ooi Peng Jin London Lucien and another* [2017] SGCA 38**, articulates the standard of care that is required by the doctor at each stage: diagnosis, advice and treatment stages (Lysaght et al, 2019). This could have ramifications of how AI is being used in medical care. A doctor whose patient is harmed and who may have decided to go against an AI advice could, in this case, be pulled up, if it was not pre-decided that AI is to augment decision making, and the final responsibility of that decision is the person who is using the AI. In many cases, this choice is taken away from the user in the way the AI system is deployed (like in the example with the Post Office or, in some cases, autonomous cars, auto-pilots etc.). While technology tools are quite helpful, humans need to be able to keep important decision making abilities and make the final judgement call.

Do Good and Do Good Well

A broader and more strategic perspective on social betterment needs to be upheld. A significant number of AI Ethics papers address the issue of 'do no harm' or **non-maleficence** (Jobin, Ienca and Vayena, 2019). But there is a need to ask if the changes being made are actually good in the long term. If a newly deployed AI makes jobs obsolete, then where will the workers find new jobs? Are the new jobs being created ones that these workers can fill-in? One of the Universal Human Rights is the Right to Work (Article 23).

On a similar note, if AI is being deployed to save paper or reduce carbon emissions, then by planting new trees and being carbon neutral, is biodiversity ensured? The Sustainable Development Goal 15 emphasizes biodiversity.

The digital divide and digital biases need to be carefully weighed upon. If AI is using data largely created by a younger age group, is it representative of the aging population? What about a population with no access to smart phones? With regard to data used to train AI, there is a need for mindfulness when deliberating on **data** generation, collection, recording, curation, processing, distribution, sharing, security, and usage (Hagendorff, 2020). Taking measures to do good well is not always easy, but proactive and responsible action can make all the difference.

A stronger and more consistent effort is necessary. A report released by the **European Parliament** (2020) titled *Artificial intelligence: From Ethics to Policy*, states: "*Ethics cannot be reduced to codes of conduct, guidelines, or principles exclusively. Rather, ethics should also be understood as a continuous process (akin to character development) that must accompany the design, development, and implementation of AI.*"

In governments and corporations worldwide, the AI operational model needs deep deliberation and the starting point is to identify who is ultimately responsible. When greater attention is placed on details, there is better transparency especially in policy formation and contract preparation. In situations where negative outcomes are foreseen in advance, AI deployment can be done in a more strategic and ethically sensible way.

In a digital economy, the decision deadlock on whether to prioritize AI vs humans or the corporations vs society will linger in the minds of many executives in the foreseeable future. Similar to the game of tug of war, victory goes to those who plan ahead, line up a motivated team with right resources, and expend the greatest effort at the right time.

▶ References

Melodena Stephens (Follow)

Melodena Stephens is Professor of Innovation Management at the Mohammed Bin Rashid School of Government, Dubai, UAE, a member of two IEEE SA Ethics Working Committees and co-author of the book AI Smart Kit: Agile Decision Making on AI.

J. Mark Munoz (Follow)

Dr. J. Mark Munoz is a tenured Full Professor of Management at Millikin University, and a former Visiting Fellow at the Kennedy School of Government at Harvard University. Aside from top-tier journal publications, he has authored/edited/co-edited more than 20 books such as: Global Business Intelligence and The AI Leader.