AI Ethics Framework

The need to implement Artificial Intelligence ethically and safely

Artificial Intelligence has the potential to have an enormous impact on our lives, both at an individual and a wider societal level. From managing our online preferences to improving the efficiency of supply chains, and speeding up medical developments, the technology is becoming more and more pervasive in our everyday lives.

What remains apparent though is that the development and use of any AI system must not result in unfair discrimination against individuals. Nor must it create unintended consequences within its processes[1].
So how can we implement AI ethically and safely?

1 www.cambridge.org/core/journals/ethics-and-international-affairs/article/artificial-intelligence-power-to-the-people/1A35A67929E4A5C34216144243C805E

How much automation is too much?

Increasingly, we see the use of AI systems requiring compliance with data protection laws, mainly to inform end users when an algorithm is being used that impacts them, and which types of features the algorithm is using to make those decisions[2]. A further related problem is for society to place an overreliance on AI to solve all types of problems. AI cannot infer additional context if the information is not present in the data. Unfortunately, in some circumstances this has already led to fatalities due to an overreliance of the technology[3].

So, in order to ensure AI can be used to achieve better outcomes, questions need to be asked on how much automation should be placed into the hands of a machine. To what extent we are comfortable relinquishing decision-making, and to a lesser extent responsibility.

Using AI ethics can act as that governance layer between the fuzzy boundaries of human and AI system interaction[4], ensuring that sound development decisions can be followed to overcome certain situations, such as:

- **Misuse** - AI being developed to falsely identify and incorrectly analyse anyone
- **Boundaries of operation and safety** - any misinterpretation of trust arising between humans and their AI counterparts
- **Accountability** - incorrect reasons being given to relevant authorities to show that no negative effects on groups or individuals were taken or made.

Establishing the building blocks for an AI ethics framework

With such a myriad of potential uses, an AI ethics framework should be applied on a project-by-project basis in order to assess the application's specific requirements. For example, an AI model which filters out spam emails will present fewer ethical challenges than one which attempts to identify vulnerable children or potential targets. As a starting point to mitigate unforeseen circumstances in the rollout of AI capability, we should seek to establish the necessary building blocks around **four key development questions** for ethical AI use:

1 Am I using AI for the right reasons?

It is important to view AI as a tool to accelerate the path we take to arrive at a decision, ultimately leading to better outcomes and what distinguishes ethical AI applications as the intended use. This implies a need to incorporate an ethical evaluation and justification of intended use prior to deployment. Moreover, tools that can provide continuous monitoring of AI systems while in operation are necessary to make sure there is no deviation from the intended use case towards anything unethical. This would allow issues to be dealt with earlier rather than further down the critical decision-making process.

2 Can I explain the reasoning path?

For AI to deliver on its promise, it will require predictability and trust of the end user. The purpose is to make AI transparent within the decision-making process, as well as the data used to develop the models and algorithms that incorporate the decisions made. When complex AI systems have been used to make significant decisions, it may be difficult to unpick the causes behind a specific course of action. The clear explanation of machine reasoning is necessary to determine accountability at the decision outcome, particularly in highly regulated industries and sectors where auditing is used extensively. To an extent, AI applications being developed require necessary built-in tracking in order to enable drill downs to the specific data points that led to the recommendations being made.

3 Can I recognise and mitigate AI bias?

Bias is an inherent problem when using AI as a decision-making tool. AI is only as good as the data behind it, and as such, this data must be fair and representative to ensure that AI evolves to be non-discriminatory. The two main sources of bias in AI arise from data and teams. With data, this often means underrepresented and imbalanced datasets. It is important to raise this question of data representation from the start. Bias occurring from teams arises when we only have a small group of people to train the machines. Therefore, the algorithms end up unknowingly reflecting the thinking of a select few.

4 How secure is the data I am using?

When we use data to feed AI algorithms, the data needs to be secure. Otherwise the risk of tampering or corruption can skew the machine's output at the expense of the end user. The result is a potential unintended decision outcome. Obviously, the training phase cannot cover all possible examples that a system may deal with in the real world and of course, this gets worse if the data itself is tampered with. Active measures should be taken to protect the data and applications, as well as an ongoing assessment of new vulnerabilities.

2 www.gov.uk/government/collections/data-protection-act-2018
3 www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html
4 standards.ieee.org/industry-connections/ec/autonomous-systems.html

Practical steps towards setting up an AI ethics framework

Central to any AI ethics framework should be the principle to encourage responsible innovation through good AI governance practices and project workflow processes. Responsible workflow steps can be applied through:

- **Maintaining strong regimes of transparency**
- **Establishing a well-defined audit trail through rigorous activity logging**

An ethical framework for AI should also encourage diversity in data and teams to prevent biases and achieving more rounded outcomes with intended use cases. The goal is to create comprehensive datasets for AI training that can address all possible scenarios and users. Likewise, an intended framework should strive to include expertise made up of individuals with varying skills and backgrounds.

To encourage this view requires cultural changes to be made within existing project structures, specifically to address the following issues:

- **Ethical issues** - Creating an AI project advisory body which fosters discussion forums and publishes the resulting guidance to the industry, community and regulators ensures ethical issues are suitably considered.
- **Diverse perspectives** - Including a broad range of perspectives is a sensible approach to handling AI ethics and establishing guidelines. A diverse team can look for unethical use from multiple perspectives and monitor unwanted outcomes during the project lifecycle
- **Trust** - Communicating developments to the wider user base to show initiative in tackling trust in using AI. The argument goes that the more isolated technologists are from the problems they're trying to solve with their AI tools, the higher the potential for unintended negative consequences.
- **Feedback** - Recognising that AI by contrast develops and grows through iterative feedback loops and from learning the way it's used. This means users and developers need to have more similar roles in determining how the system should grow over time
- **Ethical codes** - Encouraging the development of specific ethical codes that can be repeated again and shared with others within the wider community working on similar use cases across different application domains.
- **Regulatory interaction** - Developing inclusive design practises to avoid dilemmas where regulatory guidelines cannot keep pace with technological change, to a point where the guideline no longer matters. Getting regulatory interaction early to shape future policy is key.
- **Cultural sensitivities** - Inclusive design can help to resolve cultural differences. Customs can vary from place to place, and AI applications have to be aware of how their outputs will be perceived within these different cultural groups.

How does AI ethics influence the Defence sector?

As with most industry sectors, the opportunities presented by AI within the Defence sector are far reaching and potentially transformational. But given the nature of the decisions being made and the consequences such critical decisions can have within a military context, the question of ethics and safety is arguably of even greater importance here. Within the Defence sector, the potential use of AI goes far beyond autonomous weapons and stretches into uses such as information advantage applications, diagnostics, cybersecurity, supply chain logistics and asset maintenance, to name just a few.

The potential benefits are huge. But having a thorough knowledge of the critical decision-making process, and a clear understanding of what the implications could be further down the chain of command are absolutely crucial requirements for AI to play its part effectively. It is also a critical requirement to be able to trust the data sources and resulting information that propagates such critical decision making; AI is only as good as the available data. Finally, it's necessary to establish what level of human control we are comfortable giving up and what level of control is to be retained. For instance, how much control should be relinquished regarding the firing of autonomous weapons?

Positioning statement

Fujitsu is acutely aware of the potential risks posed by the development of unethical AI systems. Fujitsu is engaged with industry, academia and regulators as they continue to investigate and develop good practise measures and guidelines to ensure ethical use of AI solutions across a wide range of industry applications.

Until AI solutions reach the point where government regulators and industry have provided applicable deployment and mature usage guidelines, Fujitsu will seek to improve its development practises through direct engagement with industry standard bodies and test with participation in wider community to improve deployment of AI solutions.

What next?

Please share your experiences of implementing AI within your organisation. What lessons have you learned? What will you do differently next time?

/Fujitsu_Defence

/showcase/fujitsu-in-defence-and-national-security/

blog.global.fujitsu.com/fgb/

About the Author

Dr. Darminder Ghataoura has over 15 years' experience in the design and development of AI systems and services across the UK Public and Defence sectors as well as UK and international commercial businesses. Darminder currently leads Fujitsu's offerings and capabilities in AI and Data Science within the Defence and National Security space, acting as Technical Design Authority with responsibility for shaping proposals and development of integrated AI solutions. He also manages the strategic technical AI relationships with partners and UK government.

Darminder holds an Engineering Doctorate (EngD) in Autonomous Military Sensor Networks for Surveillance Applications, from University College London (UCL).

Why Fujitsu?

For over 50 years we have innovated with the MOD, Government Departments and intelligence communities, co-creating new technologies and capabilities. As a result, Fujitsu has around 4,000 security cleared staff and the experience to deliver and manage both generic industry offerings and those tailored to specialist needs at OFFICIAL, SECRET and ABOVE SECRET classifications.

Enabling *Your* Information Advantage

In today's complex, digital operational environment, never before has information been such a key asset in securing operational advantage. Fujitsu's vision is to provide customers with the means to translate complex data into useful information upon which to base critical decisions and actions. Transforming this ever-increasing pool of data into meaningful, useful information through analytics, automation and genuine Artificial Intelligence is critical to achieving this goal.

Fujitsu is fully committed to working closely with our customers, and through the use of co-creation will seek to enhance capability both through the acceleration of existing processes, and also through the delivery of truly new capabilities and ways of working. Our approach is based upon maximising both existing investment and best-in-class innovation, delivering the full spectrum of capabilities needed to enable your information advantage.

Contact
Telephone: +44 (0)870 242 7998
Email: askfujitsu@uk.fujitsu.com
Ref: 3960
uk.fujitsu.com