CATAPULT Digital

Challenges to Responsible AI Adoption in Industry

July
2021

Accelerating the digital future

Abstract

In January 2020, Digital Catapult convened the Industry Working Group, with its members assembled from UK-based organisations actively engaged in artificial intelligence (AI) deployment and procurement. The objective was to define what a working group of industry peers can do to advance best practices and responsible AI adoption.

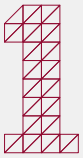
Executive summary

The introduction of AI and machine learning (ML) technology presents particular anxiety in companies. It is hard to define roles and responsibilities, while the consequences of its use may be difficult to foresee. Questions in the industry surrounding AI lead to considerable caution and negatively affect the wider adoption of machine learning techniques.

In January 2020, Digital Catapult convened the Industry Working Group, with its members assembled from UK-based organisations actively engaged in AI deployment and procurement. Over three workshops, the group members shared the challenges they faced when applying responsible AI practices.

The workshops

Three workshops provided an opportunity for Working Group members to share their experiences and perspectives on AI adoption, while fulfilling the need for peer networking and representation.



WORKSHOP ONE: SHARED INTERESTS AND CHALLENGES:

In this initial workshop, members identified the challenges faced by responsible AI and discussed potential solutions.

Five interconnected themes were identified:

- Sharing and representation
- Making the positive case
- Supply and procurement
- Tools
- Standards and regulation

WORKSHOP TWO: SUPPLYING AND PROCURING AI

This workshop debated the ethical implications of AI supply chains and how to build trust, while maximising opportunities and mitigating risk.

WORKSHOP THREE: CHAIN OF ASSURANCES

In this two-part workshop, the group analysed the roles seen in AI supply chains and the connections between them, followed by an exercise to identify methods that could establish trust between different links in a supply chain.

KEY FINDINGS:

- Five key themes were identified, then examined in detail: sharing and representation; making the positive case; supply and procurement; tools; standards and regulation
- Widespread adoption of AI requires overcoming barriers
- Individuals and companies are cautious about taking responsibility for systems they do not fully understand or control
- The need arises for an open exchange of best practice, sharing what works and what doesn't in a safe environment
- There was a strong agreement that convening as an Industry Working Group would be more impactful than working alone
- The demonstration of trustworthy AI leads to the deeper adoption by consumers and the public
- There was general frustration arising from the lack of clarity around regulations and responsibilities
- Collaboration is necessary from all stakeholders in AI supply chains, not only those represented in the Working Group, but also the subjects of AI models, the regulators and wider society.

OVERCOMING BARRIERS TO AI ADOPTION

The AI Supply Chain of Assurances has been proposed as a method to provide clarity on responsibilities, capabilities and controls across supply chains. This is expected to improve the transparency and accountability required for widespread, sustainable adoption.

KNOWRISK PROJECT

The KnowRisk project, a Digital Catapult project, will examine what a chain of assurances could look like in a federated setting. It will invite stakeholders in AI supply chains to collaborate on this project and contribute case studies.

If you are interested in contributing a case study or finding out more about the Working Group, please contact us at: appliedAIethics@digicatapult.org.uk

Contents

Introduction

In recent years, the explosion in artificial intelligence (AI) has been accompanied by increasing awareness of ethical issues in its design, development and deployment. Organisations that want to take advantage of AI's promised benefits must weigh them against the risk of doing harm or the failure to do good, as the resulting implementations and outcomes will be scrutinised by their shareholders, customers, users, employees and wider stakeholders. Additionally, there is no one-size-fits-all approach to adopting AI, nor a roadmap to follow. Organisations of all sizes demand help to apply responsible principles in practice.

In line with our role to grow early adoption of advanced digital technologies in the UK, Digital Catapult convened the Industry Working Group in January 2020 to address this interest. The responsible use of algorithms and data is fundamental to the adoption of AI and needs greater definition, so that practitioners can adopt best practices for stakeholders and buyers to recognise.

While so-called AI-first companies have naturally been the first to encounter (and often attempt to mitigate) ethical issues, their requirements are not the same as companies that are not deemed AI-first.

As large AI-first companies are predominantly based on the US West Coast, conversations about responsible AI tend to be dominated by these organisations, with their own cultural sensibilities, requirements and interests.

Therefore, Working Group members are drawn from UK-based established organisations that have operationalised AI and actively integrate AI into their processes, products, and services responsibly. Specifically, the group consists of builders, integrators and buyers of AI-enabled products and services whose primary business is not AI.

AI-first - a definition

AI-first companies, primarily build AI models for their own use: they fulfil the roles of data owner/generator, model-builder, model-user and model maintainer, implying a degree of control more difficult to achieve by organisations that are instead part of an AI supply chain.

Objectives

From the outset, the objective was to convene organisations actively engaged in AI deployment and procurement to define what a Working Group of industry peers can do to advance best practices and increase safe and responsible AI adoption.

An initial workshop was held to identify responsible AI challenges and potential solutions in the context of wider involvement with industry, government and academia, with actionable timescales.

FIVE THEMES AROSE FROM THIS ACTIVITY:

Sharing and representing

The adoption of AI is for everyone. For a group focused on AI adoption in industry, the need arises for an open exchange of good practices, sharing what works and what doesn't in a safe environment. Similarly, a group of diverse companies can communicate industry challenges and viewpoints to regulators, policy makers and researchers. Other groups in existence, such as the Partnership on AI, may have barriers to engagement or lack focus on UK (or European) interests and AI adoption.

Making the positive case

This theme focuses on making a positive (rather than risk-based) case for responsible AI practices and demonstrating that trustworthy AI leads to the deeper adoption by consumers and the public (to support a European model). Ideas for employing the carrot, rather than the stick, included economic incentives for good players, building a positive case to invest in responsible AI practices and providing ROI evidence for ethics.

Presenting this case might involve case studies from industry and a research element from academia.

Supply and procurement

This theme explored how to have confidence that a procured AI system is compliant with AI ethics. It also looked at managing where the responsibility for a designer, integrator or supplier would end for the ethical state of a ML component or system. Legal, certification, audit and design solutions were all mentioned. One near-term solution is to include ethics-related questions as part of the procurement process. Long-term, the aspiration was to standardise ways to validate and share how a project is compliant with responsible AI adoption.

Tools

Tools can help to operationalise and streamline good practices but can be difficult to find and use. For example, a checklist or tool to manage and avoid practical drift in AI use was mentioned, as well as the potential to integrate ethical matters into the continuous integration and continuous delivery (CI/CD) process.

Standards and regulation

There was general frustration arising from the incompatibility of regulations in different jurisdictions and the lack of clarity on algorithmic auditing requirements. Who is responsible for what? Who should decide what standards machines are held to? Can these be certified?

However, one attendee was cautious of standards - presumably concerned about what is achievable in the short term and what role an Industry Working Group would play.

From a collaboration perspective, there was strong agreement that convening as an Industry Working Group would be more impactful than working alone. If a sufficiently diverse group of participants is engaged, there is value in coming together as industry peers to share experiences and communicate challenges and requirements to other stakeholders (e.g. academics and policy-makers).

Participants were aware of the fact that many responsible AI working groups and initiatives already exist, so the creation of another needs to be validated. To justify this, the Industry Working Group must fill a specific gap and be more than just a talking shop. It must aim to make real progress towards the aforementioned big picture ideas.

The big picture focused on harnessing the benefits of AI while mitigating the risks, ensuring that responsible AI was the norm.

Methodology

The Working Group met three times in 2020. Each occasion provided an opportunity to share experiences and perspectives from within the group and with contributors outside it, to fulfil the need for peer networking and representation.

A more challenging task was to identify and define specific work that the group could do to make tangible progress against one or more of the other themes.

A series of workshops aimed to funnel ideas from aspiration to concrete work packages. As described earlier, the first workshop sought to identify shared interests and challenges and debate potential solutions, resulting in five interconnected big picture themes.

SUPPLYING AND PROCURING AI

For the second workshop, the Working Group honed in on the theme of supply and procurement, recognising that the ethical implications of AI supply chains are generally complex and poorly understood, while approaches to mitigating risks and maximising opportunities did not appear to exist or were underdeveloped. **Stephen Pattison, Vice President of Public Affairs ARM Holdings**, introduced the concept of the AI Supply Chain of Assurances as a theoretical solution to build trust in the supply chain, and participants mapped supply chain processes for risks and mitigations.

CHAIN OF ASSURANCES

For the third workshop, the focus was on the Chain of Assurances idea. The first exercise focused on the chain element, seeking to identify the roles played in AI supply chains and the connections between them.

AI supply chain diagrams for the Kaggle 'Deepfake Detection' competition and for cross-silo federated learning were proposed to kick-start the discussion of roles.

As a result:

- 16 roles were identified that might be played by different participants in an AI supply chain.
- Participants aligned their companies to the roles they currently play in AI supply chains.
- Links were drawn between roles where AI supply chain connections exist and where assurances might be required.

See Spotlight section on page 14

The second exercise turned to assurances. The seven ethical principles identified by the 'EU High-level expert group on artificial intelligence - ethics guidelines for trustworthy AI' were used to highlight areas that could create assurance between different links in a supply chain:

- Taking individual links in a supply chain, participants write down user-stories based on the format: 'As a [role] I want to [...] so that [principle is assured].'
- Some of these were then turned into a problem definition in the format: 'How might we [...] so that [principle is assured] for [role]'

The problem of identifying AI supply chain roles and connections between them underlines that they take many forms and complexities. Practically speaking, strict definitions for each role were not needed, as the exercise simply facilitated discussion of which roles are played and how responsibilities arise between them. For example, the building blocks on which an aspirational, fully assured supply chain must be built.

Many of the identified problem statements are actionable and can be tested in the real projects, products or services that Working Group participants are involved in. However, potential interventions are experimental and exploratory by definition, since it remains to be seen how these proposed methods and tools work in practice and provide the satisfactory assurance required.

Spotlight: AI Supply Chain of Assurances

The use of AI and ML (as distinct from traditional IT) causes particular anxiety in companies because it is hard to define and the consequences of its use may be difficult to foresee.

These anxieties are of course felt by individuals inside companies. They have real moral (and economic) incentives to avoid doing harm, even where the law lacks clarity or sufficient accountability. This leads to considerable caution and has a negative effect on the wider-adoption of machine learning techniques.

AI SUPPLY CHAINS CAN BE COMPLEX WEBS OF DEPENDENCIES AND OBLIGATIONS

Complex supply chains only compound the anxiety by multiplying the potential sources of risk. A varied mix of data sources, processing methods and ensembles of ML models, create continuously changing sources - over which no single entity has a complete understanding nor exerts full control. Like any supply chain, AI supply chains can be very complex webs of dependencies and obligations.

So what can be done? The inspiration for an AI Supply Chain of Assurances is the Kimberley Process **(https://www.kimberleyprocess.com)**, which aims to remove conflict diamonds from the global supply chain. While diamonds and AI may not appear to have much in common, they share the requirement for trustworthy systems. Participants must be confident that they are contributing to widespread, sustainable adoption, not inadvertently perpetuating harm.

The Working Group explored the question of what needs assuring and by whom in an AI supply chain [see **methodology**, page 12]. In most cases, a company will not perform all roles in an AI supply chain and can not have full control over the resulting system. The AI Supply Chain of Assurances must be just that, a construction of chains of detailed assurances between supply chain participants, which in aggregate would allow end-users, regulators, auditors and other stakeholders a holistic view of whole system responsibilities, risks and benefits.

For instance, data owners might need to provide assurances that they have appropriate consents for the proposed data-use or that their method of collection has not introduced biases. Data processors might need to provide an assurance that they have provided adequate training and support to human labellers.
There are a number of interesting existing proposals to document machine learning models (such as Google's **Model cards for Model Reporting**, **IBM's AI Factsheets**, and **Partnership on AI's AboutML** project) and a range of tools to help to monitor or adhere to ethical principles (many of which can be found in this typology) when building and deploying machine learning systems.

Although best practices for their use are still in development, these models could form part of the chain of assurances. The whole chain must add up to something that is dynamic, accessible, legible and actionable. In addition, there must be clarity about where responsibility lies and how issues can be remedied.

A tall order?

Conclusions

Widespread adoption of AI requires overcoming the barriers presented by compounding uncertainties in AI supply chains.

Ethical concerns highlighted in the use of AI with one apparent owner, are of greater concern in a supply chain with less visibility. These include poor transparency, propagation of bias, lack of remedy or redress and the impact on safety. It is not surprising that individuals and companies baulk at taking responsibility for systems they do not fully understand or control.

Modelled on the Kimberly process, the AI Supply Chain of Assurances has been proposed as a method to provide clarity on responsibilities, capabilities and controls across supply chains, to offer the transparency and accountability needed for widespread, sustainable adoption.

At this stage, what an AI Supply Chain of Assurances actually looks like is ill-defined. While it is an important academic exercise to provide more definition, our view is that a fully-functional model for an AI Supply Chain of Assurances will not emerge through thinking alone. Many hypotheses for what might work will need to be tested in practice, iterated upon, and communicated, in order to converge upon workable solutions.

By definition, this will need collaboration from all stakeholders in AI supply chains, not only including those represented in the Working Group, but also the subjects of AI models, the regulators and wider society.

Call to action

Over the next year, we would like to collate case studies that demonstrate chains of assurances in specific contexts.

These will be used firstly, to illustrate the concept, and secondly, to form the basis for discussion and improvement towards more general principles that can readily be adapted and reused.

THE KNOWRISK PROJECT

Digital Catapult's contribution will be the KnowRisk project. This project will examine what a chain of assurances could look like in a federated setting, where multiple data-holders train models locally and contribute their model updates to a central server for aggregation. The project will preserve data privacy and benefit from collaboration at the same time.

If you are interested in contributing a case study or joining the Working Group, please contact us at: **appliedAIethics@digicatapult.org.uk**

With thanks to Working Group contributors

In January 2020, Digital Catapult convened the Industry Working Group, with its members assembled from UK-based organisations actively engaged in artificial AI deployment and procurement.

With thanks to our Working Group members including:

Stephen Pattison, Arm Ltd.
Myrna Macgregor, BBC
Catherine Brien, Guardian
Oliver Smith, Koa Health
Lee Glazier, Rolls-Royce Plc
Mark Chattington, Thales

About the Responsible AI Adoption Industry Working group

Digital Catapult's Responsible AI Adoption Industry Working group is part of a portfolio of activities aimed at leveraging the unique capability and appetite of the UK AI ecosystem to grow and drive responsible AI adoption.

Our portfolio of activities is underpinned by our Ethics Committee; an independent group of AI ethics experts. The Committee is chaired by Professor Luciano Floridi of the University of Oxford. Other portfolio activities include supporting AI startups with hands-on consultations and tangible roadmaps to embed ethics into practice, and research that sets theoretical foundations in AI Ethics and Responsible AI adoption. Additional details on our portfolio of Responsible AI Adoption activities to include the full list of the Ethics Committee Members, click **here**.

About Digital Catapult

ABOUT DIGITAL CATAPULT

Digital Catapult is the UK authority on advanced digital technology. Through collaboration and innovation, we accelerate industry adoption to drive growth and opportunity across the economy.

We bring together an expert and enterprising community of researchers, startups, scaleups and industry leaders to discover new ways to solve the big challenges limiting the UK's future potential. Through our specialist programmes and experimental facilities, we make sure that innovation thrives, and the right solutions make it to the real world.

Our goal is to accelerate new possibilities in everything we do and for every business we partner on their journey – breaking down barriers, de-risking innovation, opening up markets and responsibly shaping the products, services and experiences of the future.

Visit **www.digitalcatapult.org.uk** for more information.

If you are interested in contributing a case study or joining the Working Group, please contact us at:
appliedAIethics@digicatapult.org.uk

Digicatapult.org.uk
DigiCatapult
Digital Catapult