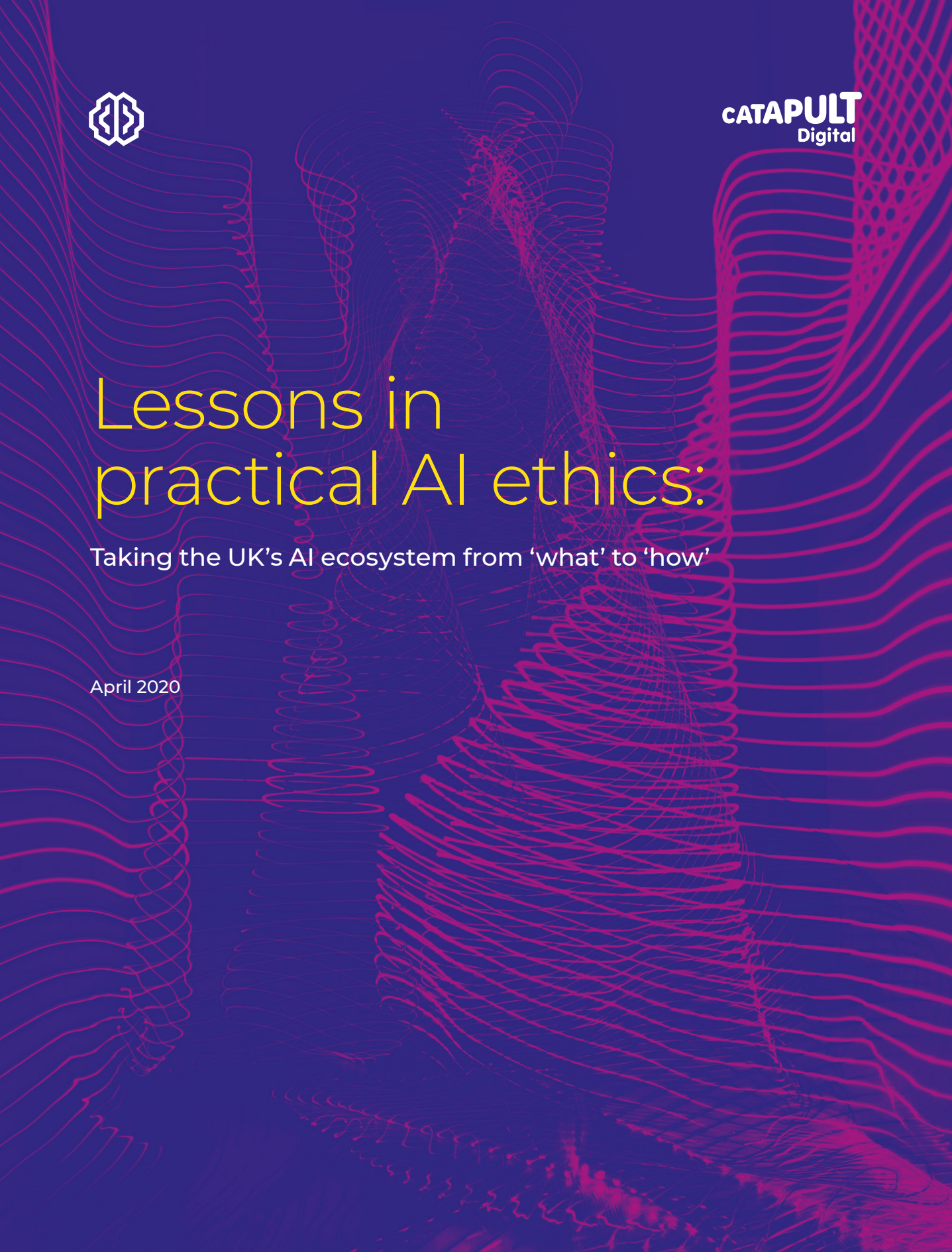

Lessons in practical AI ethics:

Taking the UK's AI ecosystem from 'what' to 'how'

April 2020

Contents

From Theory to Practice: Applied ethics for artificial intelligence (AI) system

The widespread development and deployment of artificial intelligence (AI) technologies is deeply impacting individual lives, society, and the environment. Now more than ever, at a time when our reliance on digital technologies is increasing due to the COVID-19 pandemic, it is crucial to ensure that AI systems are designed, developed, and deployed in ways that are socially beneficial and environmentally sustainable.

This is an urgent challenge for our times. Unfortunately, good intentions alone do not guarantee success; usable, interpretable and efficacious mechanisms for bridging the gap between ethical aspirations and reality are necessary. It would be unrealistic however, to expect these to emerge fully-realised in theory without some sort of testing in reality. Testing is exactly what this report is about. A handful of practical interventions that were designed carefully, but also with the humble awareness of their inevitably tentative and fallible nature, have been tested with a group of companies which have helped to critique and refine them over time. This report highlights the lessons learned and the many questions still to be answered.

It is, of course, important to share what works and what does not, yet it is equally obvious that this is easier said than done. It is therefore remarkable that each of the companies that has participated in this work has committed time and resource to the effort and exposed itself to constructive criticism, and potential failures as well as successes. They are pioneers and we are grateful and indebted to them for their collaboration and trust, without which this report would not have been possible.

The practical interventions described in the report were trialled with startups, in the UK. It is a particular typology and a specific country. There may be other, exclusive, challenges for larger organisations, for public and third sector organisations or challenges related to operating in different geographies and jurisdictions. No generalisation should be presupposed, but understanding what commonalities, and what differences there are, will help us collectively to address the overall challenge of developing AI systems that are socially beneficial and environmentally sustainable. We offer this as a concrete and constructive contribution to a much larger dialogue.

We hope that reporting on our findings will encourage others to do the same, and we welcome feedback to appliedaiethics@digicatapult.org.uk

Luciano Floridi
Chairman of Machine Intelligence Garage
Ethics Committee

Introduction

This report provides an overview of the work being undertaken by Digital Catapult and its collaborators to introduce and embed artifical intelligence (AI) ethics into the culture, products and practises of machine learning innovators.

It covers the establishment of an independent Ethics Committee, and the resulting development of an innovative Ethics Framework, and the introduction and application of practical, hands-on initiatives to action the principles of AI ethics: consultation, workshops, deep dives and the Applied AI Ethics Hub.

Ethical issues arise within AI and machine learning on an everyday basis, and so this report details the challenges encountered to date, the learnings that have been applied, and feedback and reflections from the people and organisations involved in developing and adopting this model approach.

Lessons in practical AI ethics has been produced in consultations with over 35 participating companies, and concludes by considering the approaches to ethics already being undertaken with other technologies.

The gap in the landscape

Contents

The gap in the landscape

THE DISCONNECT BETWEEN THEORY AND PRACTICE IN ETHICAL AI

Artificial intelligence (AI) and its application has profound effects on our lives: from its use in diverse sectors such as healthcare to manufacturing, and from retail to recruitment.

> It is reasonable to say that 'AI' and 'ethics' have become buzzwords. Given the technology's pervasiveness and the publication of guidelines for ethical AI use by more than 70 institutions, as of 2019.[1]

Establishing ethical ideas and values has been important in laying the groundwork for understanding the opportunities and risks in responsible machine learning development.

In 2019, the Leverhulme Centre for the Future of Intelligence and the Nuffield Foundation conducted a literature review of the AI ethics landscape, and found that a shared set of concepts were starting to emerge, with a number of repeatedly used ideas appearing within AI Ethics Frameworks and guidelines (including 'fairness', 'transparency', 'privacy', 'accountability' and 'security').

This finding indicates that there is a growing set of shared parameters for AI ethics.[2]

While common language and concepts are starting to appear in the AI and ethics narrative, what is less clear is how to define and consistently implement these abstract concepts in a practical manner for machine learning development. For example, there are more than 21 different mathematical approaches to 'fairness', each with different outcomes.[3] Interpretation by philosophers or lawyers might differ entirely.[4]

Alongside the wide-ranging interpretations of 'fairness' in relation to data collection, selection and analysis, there has also been an absence of practical ethics advice relating to the use of AI. There is no established guidance on which ethical principles are applicable to the development and use of different machine learning solutions, and which tools or methodologies best support each application.

ADDRESSING THE GAP THROUGH DIGITAL CATAPULT'S AI ETHICS INITIATIVES

Digital Catapult has unique access to a thriving ecosystem of early-stage machine learning startups in the UK. Through Machine Intelligence Garage - our flagship AI acceleration programme - we provide high growth potential innovators with access to computational power, alongside business and investment support. While supporting startups in growing and developing their solutions, we ensure that they have the resources, advice and guidance to be able to do so responsibly and ethically.

> In response to the evident need for guidance, Digital Catapult has created an applied and practical methodology for AI ethics, designed for businesses and individuals wanting to adopt an ethical and responsible approach to their machine learning development.

Machine Intelligence Garage supports participating startups with ethics consultations, deep dives and workshops to help them develop responsible AI solutions.

This practical work enables these companies to consider the long-term implications of their solutions, and take an ethical and sustainable approach to impactful and thoughtful decision-making within their business.

Digital Catapult also undertakes wider industry engagement through the Applied AI Ethics Hub. The Hub gives industry partners using AI access to the necessary tools to support the responsible development and deployment of machine learning (ML). It also provides a forum for industry to engage with policymakers, regulatory bodies and key government departments about their challenges, ensuring that the right problems are being addressed in the early stages. This unified approach across our complex ML ecosystem is essential to the UK's effective progress within AI and ethics.

> These four initiatives form the foundation of Digital Catapult's approach to practical AI ethics, addressing the gap between the conceptualisation and application of ethics for machine learning.

The four initiatives are:

1. Ethics consultations
2. Ethics deep dives
3. Ethics workshops
4. The Applied AI Ethics Hub

The gap in the landscape

KEY LEARNINGS

AI ethics receives a lot of attention and interest from startup communities and industry in general. Yet, despite this interest and a general upward trend in companies wanting to be more conscientious and transparent, it is clear that there has been a gap between intent and practical action.

It is also clear from this early engagement that the technology landscape has lacked a co-ordinated initiative for the practical application of AI ethics, and so Digital Catapult works to provide companies with the tools and advice they need to start actioning responsible and sustainable decisions within machine learning development.

To date, Digital Catapult's ethics initiatives have been enthusiastically received and positive impacts are already being seen. More than 35 machine learning companies have already benefited from an AI ethics consultation; two deep dives have been completed, and six are currently being undertaken. An industry working group has been engaged and the Applied AI Ethics Hub supports three feasibility studies currently looking at privacy and trust tools.[5]

The Applied AI Ethics Hub also helps to address the potential time and resource frictions that can arise when building a responsible business. It allows ML innovators to find relevant tools, techniques and processes to support the ethical development and deployment of their products and services.

Context and core initiatives

Contents

Context and core initiatives

Digital Catapult's independent AI Ethics Committee helped us to develop the Machine Intelligence Garage Ethics Framework. This lays the foundation for our AI Ethics initiatives.

THE AI ETHICS COMMITTEE

Digital Catapult convened an independent Ethics Board of experts and thought leaders from academia and industry, chaired by Professor Luciano Floridi of the University of Oxford. The Committee consists of two groups: the Steering Board, and the Advisory Group. Both groups advise our ethics strategy, and the Advisory Group also works directly with startups on the Machine Intelligence Garage.

Steering Board members

Luciano Floridi - Chair of the Ethics Board
Luciano Floridi is the Oxford Internet Institute's Professor of Philosophy and Ethics of Information at the University of Oxford, where he is also the Director of the Digital Ethics Lab of the Oxford Internet Institute, and Professorial Fellow of Exeter College. He is a Turing Fellow of the Alan Turing Institute and Chair of its Data Ethics Group. His research primarily concerns Digital Ethics (Computer Ethics), the Philosophy of Information, and the Philosophy of Technology.

Sir William Blair
Sir William is a High Court Judge in England and Wales. He became President of the Board of Appeal of European Supervisory Authorities in 2012. He is a Judge in Charge of the Commercial Court in London since 2016. He is also an active member of London's Financial Markets Law Committee, and chairs the Committee on International Monetary Law of the International Law Association (MOCOMILA), which brings together leading people in the financial law field.

Professor Dame Wendy Hall
Dame Wendy Hall is Regius Professor of Computer Science, Pro Vice-Chancellor (International Engagement) and Executive Director of the Web Science Institute at the University of Southampton. She was Co-Chair of the UK government's AI Review, published in October 2017, and was recently announced by the UK government as the first Skills Champion for AI in the UK. With Sir Tim Berners-Lee and Sir Nigel Shadbolt, she co-founded the Web Science Research Initiative in 2006.

Hetan Shah

Hetan Shah is Executive Director of the Royal Statistical Society and chair of the Friends Provident Foundation, a grant-making trust that seeks to create a fairer economy. Hetan is a visiting senior research fellow at the Policy Institute, King's College London, and Honorary Vice President of the Geographical Association. He is also a member of the Social Metrics Commission, which is designing new poverty indicators for the UK, and sits on the Big Lottery Fund's Impact Advisory committee.

Jo Twist

Jo Twist is CEO of Ukie, the trade body for UK games and interactive entertainment. She is also Deputy Chair of the British Screen Advisory Council, London Tech Ambassador, Chair of the BAFTA Games Committee, an Ambassador on the Mayor of London's Cultural Leadership Board, and a Creative Industries Council member. In 2016 she was awarded an OBE for services to the creative industries, and won the MCV 30 Women in Games award for Outstanding Contribution.

Jeni Tennison

Jeni Tennison is the CEO of the Open Data Institute, which has a mission to work with companies and governments to build an open, trustworthy data ecosystem. She gained a PhD in AI from the University of Nottingham, then worked as an independent consultant, specialising in open data publishing and consumption, before joining the ODI as Technical Director in 2012 and becoming CEO in 2016.

Philippa Westbury

Dr Philippa Westbury is a Senior Policy Advisor at the Royal Academy of Engineering, which brings together the UK's leading engineers and technologists for a shared purpose: to promote engineering excellence for the benefit of society. She leads the Academy's digital and data policy work, and works on other aspects of engineering policy, such as infrastructure and built environment, manufacturing and decarbonisation in which digital and data play increasingly important roles. Prior to this, she worked in the built environment sector in policy, consultancy and research roles. She has engineering degrees from the University of Cambridge and Imperial College London.

Context and core initiatives

Advisory Group members

Burkhard Schafer

Burkhard Schafer is a Professor of Computational Legal Theory at the University of Edinburgh. Burkhard Schafer studied Theory of Science, Logic, Theoretical Linguistics, Philosophy and Law at the Universities of Mainz, Munich, Florence and Lancaster. His main field of interest is the interaction between law, science and computer technology, especially computer linguistics. As co-Founder and co-Director of the Joseph Bell Centre for Legal Reasoning and Forensic Statistics, he helps to develop new approaches to assist lawyers in evaluating scientific evidence and developing computer models for these techniques.

Shahar Avin

Shahar Avin is a postdoctoral researcher at the Centre for the Study of Existential Risk (CSER). He works with CSER researchers and others in the global catastrophic risk community to identify and design risk prevention strategies, through organising workshops, building agent-based models, and by frequently asking naive questions. Prior to CSER, Shahar worked at Google for a year as a mobile/web software engineer. His PhD was in Philosophy of Science, on the allocation of public funds to research projects. His undergrad was in Physics and Philosophy of Science, which followed a mandatory service in the IDF. He has worked at and with several startups over the years.

Laura James

Laura James is Entrepreneur in Residence at Cambridge Computer Lab. She works with emerging technologies in new and growing organisations across sectors, and has been active in the tech responsibility space since 2016, with a focus on practical ways to improve industry practice. Working with businesses and learning about their technologies, challenges and opportunities has always been fascinating to her, and she enjoys supporting early stage and growing organisations. Laura is looking forward to helping the Machine Intelligence Garage startups act responsibly as regards their users, society more broadly, and other stakeholders, as well as exploring the tradeoffs and choices they face.

Josh Cowls

Josh Cowls researches the ethics and politics of data science and AI at the Alan Turing Institute in London. Josh is also a Research Associate at the Oxford Internet Institute, University of Oxford. Josh has helped launch the Turing's Ethics Advisory Group, and has contributed to the creation of the Ada Lovelace Institute, in partnership with the Nuffield Foundation. Josh's research interests lie at the intersection of ethics, politics and communication, and his forthcoming PhD will explore the legitimacy of algorithmic decision-making in society.

Christine Henry

Dr Christine Henry is a Product Manager at Amnesty International, working on Amnesty Decoders, an online volunteering platform. She previously worked at IQVIA as a Product Manager on a healthcare platform to explore and analyse patient data. Christine has many years of experience in healthcare data analysis, forecasting, and market access, as well as knowledge of machine learning and data science. She holds a PhD in physical chemistry from the Australian National University, and a law degree. Christine is passionate about investigating the ethical and social impacts of new technologies and data. She is a volunteer at DataKind UK, where she works with teams of pro bono data scientists to help charities and nonprofits to use data science techniques for greater impact.

We are in the process of expanding our Ethics Committee, due to the high demand we are receiving for our AI Ethics initiatives. New members will be announced in the upcoming months.

THE ETHICS FRAMEWORK

Digital Catapult's initiatives in practical AI ethics are all based on the Machine Intelligence Garage AI Ethics Framework, the first iterative and practical guide to be designed and developed to meet the needs of early-stage startups developing machine learning solutions across the UK.

The AI Ethics Framework was developed in collaboration with the Ethics Committee, and translates high level principles into practical questions that illuminate how they are relevant to business, people and technology decisions. This is the approach that was later taken by Europe's High Level Expert Group for Trustworthy AI. This translation process is a critical first step in putting principles into practice, and needs to take place over multiple layers and in all contexts.

The Framework has been designed to ensure that companies are encouraged to have dynamic and iterative conversations about ethics, with a view to implementing any processes or changes necessary. Each of its core principles has an associated set of questions to facilitate a reflective, consultative and deeply practical approach.

Rather than being an audit or 'tick box' exercise, the AI Ethics Framework is used to support conscientious decision-making and promote positive and ethical startup cultures. The objective is not to brand companies as 'fair' or 'ethical', but to instil a reflective, iterative process that allows startups to continuously improve and become fairer and more ethical.

Context and core initiatives

The AI Ethical Framework operates in the realm of 'post-compliance ethics'[6]. Digital Catapult does not offer a legal framework; the baseline assumption is that the startups will already be complying with relevant legislation, such as the Universal Declaration of Human Rights and GDPR. Instead, the AI Ethics Framework provides a set of considerations that enable startups to go beyond regulatory requirements and legal rules. The intent is for startups to adopt a 'constructive' ideal of responsibility - doing good where possible - rather than a 'constrained' sense of responsibility, which solely focuses on following legal rules.

- The Framework is ruthlessly practical, and devoid of any abstract and high level concepts.
- The Framework does not impose value judgments, assume any objective truth, or dictate any universal concepts to be applied. It accepts there will be a level of relativism contingent on an individual's own values; shift in public perception over time; and for changes in priorities as technologies evolve.
- The Framework asks questions that will provoke thoughtful discussion and considerate development and deployment of machine learning technologies. It provides companies with a process that will facilitate conscientious, practical decision-making.

Example principles and questions within the framework include:

Know and manage your risks

- Who or what might be at risk from the intended and non-intended applications of your product/service? Consider all potential groups at risk, whether individual users, groups, society as a whole, or the environment.

Use data responsibly

- Are the data uses proportionate to the problem being addressed?
- Can individuals remove themselves from the dataset? Can they also remove themselves from any resulting models?

Consider your business model

- What happens if the company is acquired? For example, what happens to data and software?

All machine learning teams using the AI Ethics Framework (whether through Machine Intelligence Garage or not) will be in a position to consider the issues key to their technology, activities and users, and make relevant and impactful changes within their organisations wherever necessary.

Digital Catapult encourages the widespread implementation and use of ethical AI, and our framework and activities are designed to be a valuable blueprint for replicable success across organisations of all types and sizes.

We are already seeing its adoption by others for their projects and initiatives, including Nesta (for their AI challenge[7]). Our AI Ethics Framework has enabled teams to think about and address issues that had previously not been considered.

ETHICS CONSULTATIONS

Digital Catapult works to ensure that all supported organisations are provided with a robust process to enable them to think ahead and consider the risks within their solutions.

– Each startup joining the Machine Intelligence Garage[8] benefits from a kick-off ethics consultation (based on the Ethics Framework) as part of their onboarding process. This consultation is led by two members of the Ethics Committee Advisory Group. The objective of each consultation is to have a practical conversation about the ethics of the innovator's product or service, and initiate a tangible change to their business.

Ethics consultations are carried out at the beginning of the Machine Intelligence Garage programme to encourage thoughtful discussion about day-to-day development decisions, taking into account the implications for products, processes and policies right from the start.

It's made clear to participants that the purpose of these consultations is not to judge or remove companies from the programme; the initial consultation is advisory and collaborative in nature. More than thirty-five startups have now undergone and benefited from these ethical consultations.

ETHICS DEEP DIVES

Following the success of ethics consultations and the positive reception from participants, Digital Catapult has added the ethics deep dive to the activities available to Machine Intelligence Garage startups. The year-long deep dive begins with intensive conversations held between two Ethics Committee members and the business team, resulting in a defined ethics roadmap for their organisation.

Recognising that effective change cannot happen overnight,

the roadmap is divided into stages, with each stage indicated by one of a series of milestones for progressing ethical development. Each stage is supported by consultation with the Board, with tangible, advisory and actionable points that will positively impact on their products and services, company culture, and ultimate outcomes.

Running the deep dive over a year ensures that the participants are conducting a meaningful and sustained review of every aspect of their business, including benefits, risks, data use, and processes.

ETHICS WORKSHOPS

Digital Catapult provides a one-day ethics workshop for Machine Intelligence Garage cohort members. The workshop facilitates hands-on peer-to-peer learning between startup co-founders, and is designed to spark interest in AI Ethics. Real-life case studies are reviewed and discussed, and resulting ethical questions and tensions are explored together. This helps the innovators to reflect on their own internal processes and policies as they take steps towards making their businesses more responsible.

The workshops are closely steered and facilitated by the Ethics Committee and Digital Catapult's Machine Intelligence Garage team, and participating innovators and team members benefit hugely from interacting with others in their sphere. Attendees often ask for further individual sessions with the Ethics Committee, demonstrating the value and importance they place on these activities.

THE APPLIED AI ETHICS HUB

Digital Catapult's research and interaction with companies of all sizes has underlined that well-intentioned practitioners are often searching in vain for tools (methodologies, frameworks and software) that can help them to build values-aligned AI. Even when such tools are found, there is still considerable uncertainty about how and when to use them.

We worked with the University of Oxford to discover and map what tools are currently available[9], and the results of this research were published in the Journal of Science and Engineering Ethics, winning 'Best Paper' at the 'AI for Social Good' workshop[10], NeurIPS 2019. In response to the findings, we are co-ordinating the specification, co-development, testing and documentation of a suite of tools: as part of the Applied AI Ethics Hub. This idea was developed through collaborative R&D funding bids and via conversation with industry working groups.

This foundational work is paving the way to the launch of the full scale "AI Ethics Hub" in 2020/21[11].

The Applied AI Ethics Hub will address and enable the following key areas:

- **Industry engagement:** to establish the requirements of companies that are building, integrating, or buying AI-driven products and services, and support their engagement with wider stakeholders.
- **Experimentation to enhance trust and security:** the creation of sandbox environments for testing (existing and novel) tools in a de-risked manner enables parties to explore and experiment with a variety of solutions in advance of deployment, including privacy and security enhancing tools. This work is also well placed to support open-source tools and community efforts.
- **Translation of research** into robust, documented, usable and accessible solutions, while fostering the coordination between academia, industry and government.

Context and core initiatives

- **Development of novel solutions:** these will serve to answer real-life industry challenges, which may currently be underserved in the market.
- **Building and maintaining a directory,** including resources, tools, methods and best-practices. This directory will provide information on the state of these tools': state of maturity, scope of application, and limitations. This will enable all players - whether adopters or developers - to understand the rapidly evolving landscape of AI tools.
- **Development and dissemination of best practice,** including the creation of an evidence base for responsible AI ROI, so that practitioners can better make the case for the required investment of time and resources.
- **Providing physical space** to showcase resources and facilitate knowledge exchange between all parties.

This work will provide three core benefits:

- An internationally accessible resource for the state of the art in applied AI ethics, including currently available solutions and highlighting priority areas to address.
- A physical facility to bring together developers, practitioners and leaders in the field to co-develop new solutions.
- Further establishing the UK as a leader in the global context of Applied AI Ethics.

Partners & Delivery Mechanisms

Machine Intelligence Garage companies are key to the success of the AI Ethics Hub. Their diversity of application and technology approaches make them an ideal (and enthusiastic) testbed for applied AI ethics tools.

The challenges presented by applied AI ethics require expertise from many disciplines and areas, including ethicists, domain experts, lawyers, civil society and user groups, researchers, technologists, government, and private sector companies. An advisory group representing these stakeholders will be an essential component of the hub.

Key findings: Frequently recurring ethical considerations

Contents

Key findings: Frequently recurring ethical considerations

These are the core themes relating to machine learning and ethics that have surfaced through Digital Catapult's engagement with more than 35 companies. The outputs of each consultation were documented, anonymised and then clustered according to challenge. Here they are presented in alignment with the seven principles of the AI Ethics Framework.

BE CLEAR ABOUT THE BENEFITS OF YOUR PRODUCT OR SERVICE

Benefits take centre stage when developing new technologies. Companies were able to speak confidently on this topic, and explain the benefits of using their specialised machine learning technologies. Machine Intelligence Garage's startup companies are diverse and cover a host of use cases and industry domains, including materials engineering, financial services, retail, healthcare and cyber security.

KNOW AND MANAGE YOUR RISKS

Risks may be inevitable, but being able to anticipate, recognise and manage them will lead to responsible decisions for the long term. The most recurring themes within this subset of the framework included use of the solution:

- outside its intended original use case, industry or domain;
- for surveillance purposes;
- for consumer 'nudging';
- and its impact on the environment and sustainability.

Use of product or service outside the intended industry or domain

> Startups were conscious of the risks of their product or service being used for different purposes than initially intended, for example, use by violent or malicious non-state actors, or use for (unintentionally) supporting any other harms or crimes.

Key findings: Frequently recurring ethical considerations

A number of startups explicitly expressed their opposition to working specific industry sectors. For these companies, the ethics advisors recommended the open publication of their values, such as on their website, and including this condition as part of their company policy.

Surveillance

The importance of understanding surveillance has been raised, especially where AI technology is deployed in a workplace environment, where it is important to understand how it is being used by managers and organisations. Some businesses had put safeguards in place to make sure that appropriate transparency and permissions are required for use of their product. For example, in one manufacturing site, when managers add their team members to applications, those employees are notified of the addition and asked for their consent.

However, such measures assume that an employee will always be in a position to choose whether to provide consent or not. In reality, not all workers may be in a position where they feel they are able to refuse consent, either because they are not comfortable enough to do so, or because they believe refusal would disadvantage them in the workplace. This leads to larger questions about the usage of machine learning for surveillance, and the power that such use may give to organisations if safeguards are not in place.

Consumer 'nudging'

Startups working in B2C sectors raised the concern of how their product or service might be used to target users and influence consumer behaviour. Digital marketing already uses machine learning systems and feedback loops to target individuals with advertisements based on data collected about their activity and preferences. There were discussions about whether this targeting can inherently change user behaviour, and consequently, what the responsibility of the companies building machine learning tools might be in terms of the products being promoted.

This is especially important when audiences and users may include children. When one startup had their app approved by the app store as being suitable for a teenage audience, this gave them the impetus to ensure that they were not promoting drinking, smoking, gambling or sexualised depictions of any kind. This is an essential first step of ethical consideration, and exploring the legitimacy of promoting products - even those which seem benign or innocuous - to a teenage audience is discussed during ethics consultations.[12]

Environment and sustainability

The Ethics Committee is keen to ensure that startups are thinking carefully about choices that might impact the environment and sustainability of their product or service. Questions surrounding the environment have ranged from the sustainability of the materials used to whether a product or service might promote unsustainable behaviours.

Key findings: Frequently recurring ethical considerations

USE DATA RESPONSIBLY

> Responsible data use is the cornerstone of ethical machine learning development.

Three main topics emerged frequently during discussions: regulatory centrism, data aggregation, collection and sharing, and training data bias and data quality.

Regulatory centrism

During consultations, it was commonplace for responsible data use to be conflated with legislation such as GDPR. Companies would either assert that no personal data was being captured, so they need not engage with this aspect of the framework, or they would state their compliance to GDPR. It was therefore necessary for the Advisory Group and Digital Catapult to consistently raise the need to engage with this question, irrespective of whether or not their company's activities fell within or outside of GDPR.

> While GDPR compliance is essential, it's important that companies take a responsible stance to data management beyond the parameters of this type of legislation - especially as GDPR does not cover all data questions, still leaving a lot of 'grey areas'.[13]

Due to the evolutionary nature of AI and ML, policies and regulations relating to this technology are often more reactive than proactive, as legislators and policy makers cannot always respond to developments quickly enough to avoid harm to individuals, society or the environment.

While startups are usually diligent about observing current legislative requirements, there can sometimes be a lack of analysis of options and responsibilities beyond the minimums set by law.

> Digital Catapult is ideally placed to pioneer the "What more could we do?" or "What could we do better?" conversations that lie outside the parameters of current legislation.

Data aggregation, collection and sharing

Machine learning companies operate across different domains and industry sectors. Startups have discussed how they could legally and ethically share data or learnings across industries or sectors to benefit companies and individuals. For example, participants talked about how they might be able to share relevant learnings from working in one sector with a completely different sector. There would be no apparent conflicts of interest in sharing declassified data, and there might be benefits to both sectors and their respective ecosystems. This highlights the interesting synergies that AI can foster between sectors, given its cross-cutting nature and use cases in different industries and disciplines.

Bias in training data and data quality

Biased and poor quality data has been widely expressed as a potential concern. Startups were aware of this potential problem and hoped to avoid the entrenchment of bias, or errors resulting from poor quality data. Conversations on this subject tended to centre around data sources. For example, publicly available data might be incredibly rich, but may have limitations - such as erroneous or missing data, or selection bias during collection - that will skew insights drawn from the data set.

Key findings: Frequently recurring ethical considerations

BE WORTHY OF TRUST

> Being worthy of trust requires an understanding of where a solution's blind spots might be.

The two most common themes in this area were black boxes and explainability, and what the implications of fairness might be.

Black boxes and explainability

The need to mitigate against black boxes, and enhance explainability, is very sought after amongst startups and their clients. The ability to interpret ML models and better understand its prediction pathways allows for better transparency and credibility into the system.

It is important for companies to build trust by highlighting where there are black boxes, and consequently, what impacts this might have for the resulting recommendations and predictions from the model..

Fairness

There were multiple conversations about fairness and how it relates to products and services. As already mentioned, 'fairness' is poorly defined and can be used in different contexts or to refer to different values or sets of outcomes. Fairness was often used in the context of biased data: while data quality is important, fairness is a different issue to bias mitigation. Startups need to be aware of the different concepts of fairness and think carefully about which is most suitable for them to adopt.[14]

Some of the startups had thought carefully about the impact their product or service would have on different individuals and decided not to proceed with projects in the interest of fairness. It was encouraging to hear startups exhibiting an elevated level of awareness on the limitations of their innovations, and deciding to not engage in projects that may lead to unfair outcomes.[15]

PROMOTE DIVERSITY, EQUALITY AND INCLUSION

Most participating companies had engaged in discussion about this principle and had already sought to apply it, although it was evident that some would have had more opportunity to apply it in practice. The two main themes that emerged were ensuring diversity and educational diversity.

Ensuring diversity of gender, race and other protected characteristics

Data that is diverse in gender, race and other protected characteristics is important to ensure high levels of accuracy and precision across use cases. The Ethics Committee has frequently highlighted the importance of ensuring data accuracy across racial groups: for example, it has been well-documented that some face recognition technologies for connected autonomous vehicles work best with white faces and bodies.[16] Consequently, the use of expansive and diverse datasets will enable better and more consistent accuracy across groups.

One approach to ensuring data diversity is to ensure diversity of technology teams, based on the view that different perspectives can be useful in reducing bias. There is a body of research which indicates that diverse teams also perform better, not only effective in reduction of bias, but in terms of development and long-term product success. Digital Catapult aims to support and accept diverse startups onto its programmes, and consultations often raised the question of ensuring expansion plans were forward-thinking around diversity and inclusion.

Educational diversity

For some early stage startups, with teams of fewer than five people, recruiting from a range of educational backgrounds had proved difficult. For example, advertising opportunities on university boards or at university fairs can result in multiple applications from very similar candidates. As startup companies begin to scale, it is important for them to consider the value of educational diversity when recruiting.

It was highlighted that hiring teams from similar universities may lead to similar approaches in problem solving. There was also a wider point that hiring should be skills-focused, and less about which avenues (for example, universities) are taken to attain those skills. This was tied in to the idea that prestigious universities are predominantly attended by more privileged socioeconomic groups - with one of the advisors highlighting that 'class' diversity is paramount within the advanced tech spaces - both for better product outcomes, and in promoting equality and inclusion.

BE OPEN AND UNDERSTANDABLE IN COMMUNICATION

Effective communication, when coupled with a principled approach to ethical considerations, is a competitive advantage, even when hard moral issues are on the line. The two most frequently discussed aspects of communication were the clarity of language used and clarity in the terms and conditions.

Clarity of language employed

Many startups had already considered the importance of using appropriate language and tone to describe their product or service. Making content straightforward and conversational could be especially important if the ML tool was to be used in non-technical areas or by non-technical communities. It was common for startups to be eager to make sure that all content relating to their product or service would be accessible to anyone.

Clarity in the terms and conditions

Some founders were keen for their terms and conditions to be as user friendly as possible: concise and written in plain English. The difficulty in achieving this was discussed: benchmarking against the policies of large and established technology companies was considered inadvisable, and to write new policies that improved on those already in use within the market would require a strong legal team with deep legislative knowledge. However, this does represent a positive opportunity, and if undertaken correctly, could lead to competitive advantage, especially in the face of user mistrust or apprehension.

CONSIDER YOUR BUSINESS MODEL

Business model discussions largely covered pricing and accessibility, and the issue of controlling how a product or service might be used, and in which industries or sectors.

Pricing and accessibility

Machine learning startups frequently need to consider different pricing models. Discussions revolved around aspects such as tiered pricing based on access to different levels of features, and ways of handling accessibility and inclusivity for different-sized client companies. For example, some startups had considered offering their services at a reduced rate for early stage companies or SMEs. It is clear that a number of startups had thought carefully about how to make their product or service more accessible, especially if their offerings were for general use and benefit of wider society.

Controlling how a product or service can be used

Many founders expressed concerns about retaining ethical values while attracting new investment, ensuring control over algorithms and data, and limiting uses of products, especially for APIs. Many organisations were explicit in their desire not to work with, or have their product or service used by, players in particular industry sectors, and were seeking ways to exercise control over their downstream supply chain.

Ethics advisors questioned what would happen to an organisation's data, services and tools if the business was later sold to another company with an entirely different set of values.

Key findings: Observations from ethics initiatives

Contents

Key findings: Observations from ethics initiatives

This section explores observations and feedback from our expert advisors on the Ethics Committee and within Digital Catapult, companies that have engaged in our ethics initiatives, as well as suggestions made for potential future activities.

INTERNAL AND ETHICS COMMITTEE OBSERVATIONS

After conducting ethics consultations with a number of Machine Intelligence Garage startups, Digital Catapult and the Ethics Committee have considered the key learnings.

> Even when applications to the Machine Intelligence Garage seemed to be broadly unconcerning, there was always something of relevance and importance to discuss. Engagement with companies at this stage was validated by the universally high levels of interest and willingness of startups to engage and be receptive to constructive feedback.

The Advisory Group observed that the AI Ethics Framework was instrumental in framing issues succinctly and facilitating dynamic conversations during the consultations, as companies knew the angle and direction discussions would take. Each consultation was future-focused, highly practical and typically ended with concrete anticipatory steps that each company could take to increase their responsible innovation going forward. Framing ethics as risk prevention (as opposed to harm mitigation) is a most beneficial and productive approach, and has enabled a focus on the positive impact that projects will hopefully have. This has enabled advisors to directly address the trade-offs (that is, maximising the good while minimising the bad) that each startup faces.

Finally, Digital Catapult and the Ethics Committee have been able to reaffirm the decision to make the Advisory Group purely advisory, rather than make ethical judgements on each project application.

> Whether or not a company is ethical is not (usually) a binary question. Not having to make a binary ruling enables the Advisory Group to conduct open and candid discussions that are more likely to result in concrete positive change.

Key findings: Observations from ethics initiatives

FEEDBACK FROM PARTICIPATING COMPANIES

Ethics Framework and ethics consultations

It is clear that the framework and initial ethics consultations were well received, and the vast majority of the companies who used them displayed the will and enthusiasm to engage with the issues.

- The Ethics Framework was seen as a valuable asset that provided impetus for internal discussions around ethics.
- The Framework enabled participants to start developing actionable changes for their business.
- From data usage to internal company processes, the feedback indicated that the Framework is comprehensive and explores many different issues that had not been previously considered.
- There was value in having a formalised set of questions that enabled co-founders to unpack their own personal moral values and see how it could be appropriate to extrapolate these into a business setting.

By way of feedback, it was suggested that to be able to engage with ethics on a deeper level, a greater understanding of how to delegate responsibility for each item would be needed. For example, it might be expected that two different people are to account for the two separate ethical issues of "Is my data biased? Is the data suitable for this context?" versus questions such as "Are we hiring a diverse set of people? Is our work culture inclusive?" Clarity on responsibility and accountability for differing issues might catalyse the institutionalisation of ethics within their business.

While there have been frequent requests for some kind of checklist for startups to use, this is precisely the kind of approach that the AI Ethics Framework is intended to prevent. Digital Catapult and Ethics Committee advisors understand the difficulties in knowing where to start, but operating in an ethical manner is not about doing the minimum possible to obtain a stamp of approval. Instead, it is important to foster the right kind of thinking, mindset and culture to encompass a proactive and practical approach to ethical AI on an ongoing basis. It is the need for this kind of shift in thinking that validates the initiatives that Digital Catapult and the Ethics Committee are pioneering.

It is interesting to consider startups' underlying motivation for incorporating ethical considerations and practices into their product or service. The expected reasons of best practice and benefiting society have been raised, but it has also been commented that implementing ethical practices can result in competitive advantage, especially in an era of skepticism and nervousness about AI and ML tools. If this becomes an upward trend and competitive advantage becomes evident, there may be an acceleration in ethical activities within startups, and the establishment of new industry incentives and recognised standards. However, any obvious cynical pursuit of advantage through ethical AI (real or exaggerated) could lead to loss of public trust.

Some companies would have also benefited from a deeper level of awareness on how to prioritise each of the issues. The framework is complex and more direction on how to develop these ideas might be needed. While this need is actually addressed through the deep dives, Digital Catapult is now looking at ways to provide answers to these questions outside the deep dive process.

It has been suggested that the Ethics Committee provides drop-in clinics (as one-to-one sessions), as new issues requiring specific attention may arise over time. This would leave the door wide open for frank conversations around ethics when they are needed.

These suggestions highlight the importance and value of the work that Digital Catapult is doing in this space, and shows that startups are interested in access to more ethics-focused activities.

Conclusion and next steps

Conclusion and next steps

Applying a strong ethical framework to the development of machine learning technologies is essential. The groundwork of what responsible machine learning development looks like has been laid; it is now important to address how this can be achieved through co-ordinated cross-industry initiatives. Digital Catapult works continuously to bridge this gap between theory and practice by engaging all the relevant parties within the AI and ML ecosystem and offering practical applied ethics solutions. The impact of these interventions is already being seen and continues to grow.

In 2020/21 we hope to launch the Applied AI Ethics hub as a cross industry coordinated effort to promote the responsible development of AI technology for societal benefit.

CONSIDERING ETHICS ACROSS THE ADVANCED DIGITAL TECHNOLOGY STACK

Alongside machine learning ethics work, Digital Catapult is exploring how to develop practical guidance on ethics for other technology areas. This work has already begun, with consideration of the possible ethical challenges relating to augmented, virtual and mixed reality content, and the resulting publication of our report on immersive and ethics.[17]

In future, this may extend to considering the ethical implications of combining advanced digital technologies, for example through the scale and ubiquity of data being collected, processed and analysed using next generation connectivity, such as 5G.

HOW TO JOIN THE INITIATIVES

We aim to develop the AI Ethics Hub to serve the multidisciplinary community needed to responsibly develop AI technology for societal benefit. We are interested in engaging with different stakeholders, including but not limited to: startups, technologists, industry, lawyers, regulators, standards bodies, academics, civil society groups and any other interested parties - to get involved, we can be reached at: appliedAIethics@digicatapult.org.uk

Footnotes

[1] See these two repositories: Algorithm Watch. The AI Ethics Guidelines Global Inventory (9 April 2019): https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/ and Winfield, A. (18 April 2019): An Updated Round Up of Ethical Principles of Robotics and AI. http://alanwinfield.blogspot.com/2019/04/an-updated-round-up-of-ethical.html

[2] https://www.nuffieldfoundation.org/wp-content/uploads/2019/12/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf

[3] See this online tutorial from Arvind Narayanan, Associate Professor at the University of Princeton for more information on the 21 mathematical approaches to fairness

[4] https://www.nuffieldfoundation.org/wp-content/uploads/2019/12/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf

[5] These feasibility studies are already up and running, focusing on a host of areas: federated learning, chain of custody and deep fakes.

[6] Term coined by Luciano Floridi, https://link.springer.com/article/10.1007/s13347-018-0303-9

[7] https://www.nesta.org.uk/case-study/longitude-explorer-prize/

[8] As of the sixth cohort, in 2018, when this initiative began.

[9] https://link.springer.com/article/10.1007/s11948-019-00165-5

[10] https://aiforsocialgood.github.io/neurips2019/accepted track2/posters/26_aisg_neurips2019.pdf

[11] At the time of writing, it is pending funding.

[12] This has come to fruition as Instagram recently banned promotional posts for dieting supplements for users under 18.

[13] This has been recently documented by Sandra Wachter and Brent Mittelstadt, who highlight that GDPR does not sufficiently protect individuals, given the inferences that machine learning models can make from data, even if it is devoid of personal information. See: Wachter, Sandra, and Brent Mittelstadt. "A right to reasonable inferences: re-thinking data protection law in the age of big data and AI." Columbia Business Law Review (2019).

[14] The difficulty in the concept of fairness for machine learning can be further drawn out in this tutorial which highlights 21 mathematical definitions of fairness.

[15] In this case, the company referred to 'fairness' in terms of prioritising equality and equal opportunity.

[16] https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/

[17] https://www.frontiersin.org/articles/10.3389/frvir.2020.00001/full

Digital Catapult
101 Euston Road
London NW1 2RA

0300 1233 101

www.digicatapult.org.uk

migarage.ai