Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery

How to Use AI for Good? The Ethical and Societal Implications of Using AI in Scientific Discovery - an AI^3 Science Discovery Network+Workshop at the WebSci'20 Conference

07/07/2020

AI^3 Science Discovery Network+ & WebSci'20

WebSci'20 Online Conference

Michelle Pauli
Michelle Pauli Ltd

21/07/2020

AI3SD-Event-Series:Report-18

How to Use AI for Good? The Ethical and Societal Implications of Using AI in Scientific Discovery - an AI^3 Science Discovery Network+Workshop at the WebSci'20 Conference
AI3SD-Event-Series:Report-18
21/07/2020

DOI: 10.5258/SOTON/P0026
Published by University of Southampton

Network: Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery
This Network+ is EPSRC Funded under Grant No: EP/S000356/1
Principal Investigator: *Professor Jeremy Frey*
Co-Investigator: *Professor Mahesan Niranjan*
Network+ Coordinator: *Dr Samantha Kanza*

Contents

1 Event Details

Title	How to Use AI for Good? The Ethical and Societal Implications of Using AI in Scientific Discovery
Organisers	AI^3 Science Discovery Network+ & ACM Web Science 2020 Conference
Dates	07/07/2020
Programme	Programme
No. Participants	28
Location	WebSci'20 Online Conference
Organisation Committee	Dr Samantha Kanza – AI^3 Science Discovery Network+, Dr Nicola Knight – Physical Sciences Data science Service, Mr Samuel Munday – University of Southampton
Discussion Group Leads	Dr Peter Craigon – University of Nottingham, Dr Nicola Knight – Physical Sciences Data science Service, Dr Cian O'Donnovan – University College London

2 Event Summary and Format

This workshop was run by AI^{3}SD (Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery) as part of the ACM Web Science 2020 Conference [1]. This workshop was run online via zoom as it took place during the COVID-19 pandemic lockdown. The programme was made up of several related presentations that were designed to get the attendees thinking about different ethical considerations for AI in Scientific Discovery. Each presentation was followed by a short discussion session where the attendees could ask the speaker questions and could discuss the main topics of the presentation. The workshop also included an interactive activity using Moral IT Cards [2] which were designed as a tool to help reflect on different ethical issues.

3 Event Background

Artificial and Augmented Intelligence systems have the potential to make a real difference in the scientific discovery domain however this brings a new wealth of ethical and societal implications to consider with regards to this research (e.g. human enhancement, algorithmic biases, risk of detriment). This workshop looks to explore the ethical and societal issues centered around using intelligent technologies (Artificial Intelligence, Augmented Intelligence, Machine Learning, and in general Semantic Web Knowledge Technologies) to further scientific discovery, with a strong consideration of data ethics and algorithmic accountability. Advances in technology and software are rarely inherently bad in themselves, however that unfortunately does not preclude them from being subverted to ill intent by others; furthermore, as demonstrated by the examples above, even an unintentional lack of care towards ethical codes and algorithmic accountability can lead to societal and ethical implications of scientific discovery. It is our responsibility as researchers to consider these issues in our research; are we conducting studies ethically? What ethical codes can we put in place for scientific discovery research to mitigate against ethical and societal issues. These are really important issues, and they require an interdisciplinary focus between scientists, social scientists and technical experts in order to be comprehensively addressed. This workshop is one of a set of ethics workshops that AI^3 Science Discovery Network+ have run to facilitate the necessary conversations about ethics for AI for Scientific Discovery.

4 Introduction

We are living through an AI and data revolution. Artificial and augmented intelligence systems are already being used in the scientific discovery domain and have the potential to make a groundbreaking impact. However, using AI in this way comes with a wealth of ethical and societal considerations, from algorithmic bias to explainability and privacy concerns. Ethical frameworks covering many of these issues abound but are they enough?

Tensions, trade-offs and transparency in AI were the themes running through the AI^3 Science Discovery Network+ 's all-day workshop, part of the ACM Web Science 2020 Conference [1]. In a vibrant example of the interdisciplinary approach that is essential to addressing these big challenges, philosophers, computer and material scientists, sociologists, ethicists and many more gathered to explore the topic. Four presentations set out some of the key issues. Participants then tackled the ethical dimensions of a very current real-life application, using data from a Covid-19 contact tracing app for scientific research, in an interactive session using Moral IT cards [2]. Due to the Covid-19 situation, the entire workshop was conducted via Zoom.

5 Presentations

5.1 Ethical Frameworks. Ethical Judgements – Dr Will McNeill, Lecturer in Philosophy at the University of Southampton

Figure 1: Dr Will McNeill

Dr McNeill opened the workshop with an overview of the tensions and trade-offs inherent in the different requirements for trustworthy and ethical AI systems.

There is no shortage of ethical frameworks for AI and Dr McNeill pointed, in particular, to the EU Commission's 2018 Communication on AI [3], which set out guidelines for trustworthy AI: that it needs human agency and oversight plus technical robustness and safety, it needs to consider privacy and data governance, societal and environmental wellbeing needs to be built in and it should not inherit our own biases. It must be transparent and accountable.

While all these requirements are familiar and reasonable, Dr McNeill was keen to emphasise that the framework itself does not provide answers to the difficult decisions involved in resolving some of the tensions raised by potentially competing requirements, however laudable each appears to be individually.

What might those tensions be? Take technical robustness and transparency, where there may be an inverse correlation between transparency and technical robustness. While designers have a duty to make systems both as transparent and as technically robust as possible, there is necessarily a pay off. Equally, ensuring greater transparency might mean having access to the data training set used to develop the AI, but that could have implications for privacy and data governance. And while privacy and data governance are very important, designers also need as much data as possible to train and test the robustness and safety of artificial systems, which is also critical.

Dr McNeill pointed to the case of NICE [4]. It uses transparent and non-discriminatory algorithms to decide which drugs can be deployed in the NHS and when. The algorithms show how those tensions are resolved, revealing why NICE has chosen to allow a drug to be released or withheld in the NHS, using the same reasoning as in other cases eg cost-benefit analysis, QALYs [5] etc. But, at the same time, it could be argued that this may be at the cost of moral fairness (when cutoffs to drug availability are arbitrary rather than ethical), human agency and choice and, indeed, to a certain extent, robustness. There is a loss of accuracy on these dimensions by opting for dimensions of transparency and non discrimination. It is an ethical trade off, not a resolution of an ethical problem. It is a decision to favour one set of dimensions over another.

We cannot resolve such tensions. They can only be balanced and traded off against each other. No ethical system or higher order algorithm can be used to do this automatically: it will always require human judgement. But we can do this more or less ethically. Ethical agential or institutional transparency is required to recognise the difficult payoffs, understand that they cannot be resolved, only reconciled, and ensure transparency around the discussion that takes place and how the tensions are reconciled. Just as in everyday life when ethical duties conflict (such as when the duty not to kill is balanced against an act of self defence), It is ultimately a matter of human judgement.

As Dr McNeill concluded, ethical frameworks are necessary and desirable but cannot be treated as a replacement for ethical judgement. Instead, they highlight the tensions that only judgement can reconcile. It is our duty to be transparent and provide narratives about which ethical trade offs have been made and why.

Discussion of Dr McNeill's presentation covered:

- Whether algorithms can help (yes, but genuinely ethical systems use judgement so a full algorithmic system may not be an ethical system, even if it might be the right thing to do) and the balance to be struck between transparency and uncertainty, when more information isn't going to tell us what to do or the answer. In those cases you need institutional transparency and honesty about the scope of a system, how explainable it is, how it has been tested and how the decision to use it was reached. In some cases where a system is not transparent and is making judgements, the better it is the less explainable it will be, just like humans. There are different levels of transparency but if we are seeking trust from transparency then it must come through genuine engagement with stakeholders to achieve real commitment and engagement with the decision and an understanding of the difficulty of the decision.

5.2 AI Ethics from the Ground Up: Cultivating Capabilities for Care – Dr Cian O'Donovan, Research Associate, Dept of Science and Technology Studies, UCL

Figure 2: Dr Cian O'Donnovan

The theme of the interplay of transparency and trade-offs was continued by Dr Cian O'Donovan through the lens of the complexities of interdisciplinary research. Interdisciplinary research has led to significant breakthroughs in the fields of AI, robotics and autonomous systems. Yet it has also been the basis of significant controversies, such as the recent example of Cambridge Analytica and Facebook's amalgam of psychometrics, data science and engineering at scale. Interdisciplinary research is often suggested as an answer to societal challenges but what kind is required and how do we assess and evaluate it? How do we go beyond ethics frameworks and into practices and processes?

Dr O'Donovan explained that he 'does research on research' and has been exploring how researchers approach research into AI and robotics - two of the most significant interdisciplines that have emerged over the last few years – and looking at some of the tensions and choices AI researchers make. He has been undertaking situated ethnographic research at the Bristol Robotics Lab on the University of the West of England (UWE) campus. Many of its projects are specifically aimed at the public good, such as the potential of robotics in caring for the elderly and work on turning urine into energy using microbial fuel cell technology.

His concern is that technology is all too often seen as a treatment, with the approach that if we deploy more of the right kind of technology it will solve our problems. However, this risks neglecting the different voices, values and interests of a wide range of people. As a result, innovation policy tends to focus narrowly on inducing acceptance. For Dr O'Donovan that misses the point: technology can be steered in different directions and the benefits distributed to a variety of people and communities, depending on the choices we make and how we make those choices.

Of three possible ways to think about the politics of interdisciplinary research (reach, logics and discourses, and capabilities), Dr O'Donovan has been focusing on the capabilities that are needed and can be cultivated during interdisciplinary research. These include cognitive capabilities but also broader skills such as how to manage a research team, how to collaborate, be reflexive, recognise power and build democratic struggle and recognise the tensions that cannot be resolved through algorithms. He described how he has used bibliometric mapping to locate capabilities at the Bristol Robotics Lab, to look at the different kinds of knowledge in the building and how it comes together. What actual capabilities did researchers individually and collectively have to resolve tensions, highlight choices, deal with underground politics, broaden participation, pull in expertise and accelerate and steer innovation?

Capability mapping has proved to be a useful tool in showing how research and technology can be steered by the people doing it and, crucially, how they deal with ethics principles in practice.

Discussion of Dr O'Donovan's presentation covered:

- The need for more attention to be placed on translation between disciplines, between institutions and between researchers and people who may benefit from the research. The question of who gets to do that translation and on what terms is critical. Mutual facilitation that recognises values and ethical concerns, and also recognises that values and voices might be missing, is important.
- The risk that, in interdisciplinary settings, there is too often a desire to wallpaper over intrinsic tensions in order to produce consensus and 'get everyone to agree and move on', with the result that marginal voices can be silenced. How do we resolve the issues, values, conflicts? The first step is to admit that they are there and that technologies do treat people differently.

5.3 Data Ethics for AI – Ms Jacqui Ayling, PhD candidate in Web Science CDT at the University of Southampton

Figure 3: Ms Jacqui Ayling

Jacqui Ayling also brought together the themes of tensions, transparency and frameworks in a presentation that focused on the fact that, faced with a deluge of high-level ethical principles for AI, we need to recognise that, in reality, moral decision-making tends to be messy, contingent and contextual.

Ms Ayling suggested a number of places we can look for how to address the ethical challenges of emerging technology and its application, including cybersecurity, risk assessment / management, product safety regulation, professional ethics, human rights law and normative social values and ethics. She pointed to an interesting analysis (Jobin et al, 2019) [6] in Nature that ranked the key terms from ethical guidelines. 'Transparency' is the clear winner, referenced in 73 of 84 AI guidelines, followed by 'justice and fairness' (68/84), 'non-maleficence' (60/84) and 'responsibility' (60/84). 'Privacy' fared somewhat worse at 47/84.

High-level ethical principles are important, argued Ms Ayling, as they can condense key points and create shared sets of value to cascade through an institution or team and form the basis for more formal standards. However, they have limits. They do not provide guidance for action in specific situations or how to resolve conflicts. Interpretations of guidelines will be different according to ideology, politics and moral standards. While there appears to be a convergence in the AI community and other stakeholders with finding, creating and using, it is unclear whose voices are missing and what might be missing from the principles as a result.

She also reiterated the tension between transparency and explainability, citing the UK House of Lords AI committee report that states that: "it is not acceptable to deploy any artificial intelligence system which could have a substantial impact on an individual's life, unless it can generate a full and satisfactory explanation for the decisions it will take." However, if, for example, a medical application for diagnosis may deploy an algorithm that cannot meet this criteria – do we not use it, even though we may be satisfied by its accuracy, and it may save lives?

Ms Ayling ended with a call for more more tools that can be used in real situations by real teams, and fewer large, high-level policy documents.

Discussion of Jacqui Ayling's presentation covered:

- The need for pragmatism, reflecting on what other people do in other disciplines, the problem of oversight of complex assemblages and the need to speak to people outside the 'academic bubble'.
- The issue of transparency coming up against the fact that some algorithms are extremely valuable to the companies that own them and requiring them to unveil what they do introduces tricky notions around privacy, corporate secrecy and national border issues. There is a need for tougher regulation.

5.4 Ethics for AI for Scientific Discovery - Dr Samantha Kanza, Enterprise Fellow at the University of Southampton and AI3SD Network+ Coordinator

Figure 4: Dr Samantha Kanza

"There is no doubt that AI and machine learning is flavour of the month again," said Dr Samantha Kanza, as she began a whistlestop tour of some of the key issues and challenges in the field of ethics for AI for scientific discovery.

Firstly, there is the thorny issue of **data sharing**, which comes up in every network meeting. More data is always needed but it needs to be the right data. Garbage in garbage out applies - if the data isn't good quality then the algorithm won't work. In some cases, such as a covid-19 contact tracing app, it is a matter of life and death that the data is accurate. Tensions between data sharing and privacy also come to the fore.

Data bias is a growing problem, especially with examples of biased medical studies if certain groups are (albeit unintentionally) excluded. There is a need to diversify not just the data but the people collecting the data.

Decision-making must be transparent - is it a human or a machine that is making the decision or both? Who should be trusted to take a life or death decision in a medical context?

There are a number of other questions around **transparency and explainable AI**. How can you make a decision about how much to trust an AI if you don't understand it? There is a need to think about it in terms of layers: we need to understand the data but also how a decision is being made, how an algorithm is using the data and disseminate that in a way that people can make decisions while appreciating that we may not be able to explain all of the processes behind it. But who should be making informed decisions? If a doctor presents me with an option that comes from a machine, do they or I or the machine make the decision?

Responsible AI and AI 4 Good raises questions: should we be directing AI for Scientific Discovery towards the greater good? Where should we put our efforts? Who judges if that is the right decision for the greater good?

Finally, hacking and vulnerability and subverting research with ill intent are crucial issues that also produce tensions: how can we protect our research / data / systems from hacking or subversion and how does this conflict with transparency? What about unintended consequences, such as where the legitimate use of drones in agriculture might be subverted for use in military conflict?

Dr Kanza concluded by emphasising the AI^3 Science Discovery Network+'s belief that augmented intelligence - getting the best out of human and machine intelligence is the best way forward.

Discussion of Dr Kanza's presentation covered:

- The question of when to share and not to share data and the difficulty of imagining the uses to which someone with ill intent could put the data was raised as a serious issue, along with the challenge of persuading companies to share datasets (while recognising the good reasons why some might not be able or willing to do so). There is a need to work on better agreements between academia and business in order to have access to more high-quality, well-described data.
- The question of whether it is ethical to collect data 'just in case' it is useful was also raised, with the recognition that there may not always be easy or tidy answers. Recognising the value of storytelling around how the data will be used and the benefits it might bring and to whom – and the issue of who tells those stories – is vital.

6 Interactive Session

Given the virtual nature of this conference, this activity was facilitated by using the TableTop Software TableTopia [7] whereby the attendees were split into groups via breakout rooms and could use the software to explore the cards.

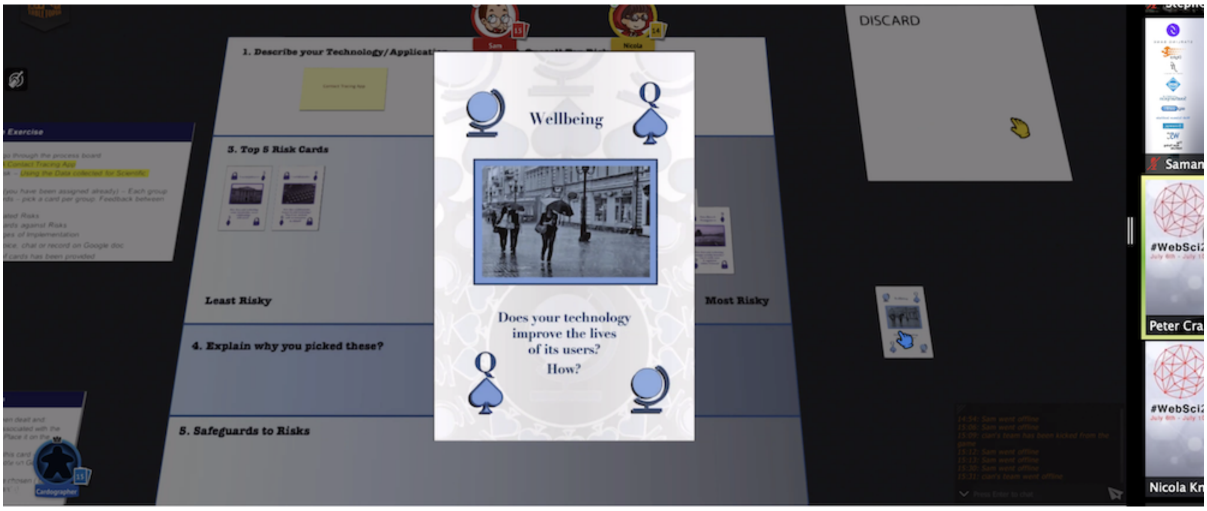
1. Describe your Technology/Applicat
3. Top 5 Risk Cards
Least Risky
Most Risky
4. Explain why you picked these?
5. Safeguards to Risks
DISCARD
Wellbeing
Q
Does your technology improve the lives of its users? How?
Peter Cra
Nicola Kr
Saman
#WebSci

Figure 5: The Interactive Activity using the Moral IT Cards and TableTopia

6.1 The Moral IT and Research Ethics Cards – Dr Peter Craigon, Research Fellow at the University of Nottingham

Figure 6: Dr Peter Craigon

Dr Peter Craigon introduced the Moral IT cards [2] before workshop participants separated into breakout groups to discuss a real-life scenario using the cards.

The playing cards-style deck with four suits (+1): privacy, ethics, law, security, (+ narrative - setting the context) were designed as a tool to encourage reflection and engagement with ethics. The goal is to make ethics discussions less high level and to ensure ethics by design - that ethical considerations are designed in rather than agreed afterwards.

The cards enable ethical engagement in an equal and engaging way with no hierarchy of knowledge. Each card features a principle and an open question, which is broad enough to focus on 'your' technology. While the card deck was originally created with IT-based systems in mind, "it could work just as well with a horse and cart," Dr Craigon remarked. Adaptable and flexible to individual use cases, the step by step approach focuses on the risks associated with the technology and then possible safeguards against the risks, such as issues of trust safeguarded by transparency, and then the challenges and the pragmatic realities eg if it would be expensive or difficult.

Participants were set the challenge of using the cards to discuss the ethical issues around using data from a Covid-19 contact tracing app for scientific research. Dr Craigon introduced Tabletopia as the online platform used to deal and examine the cards in a virtual situation. The three groups selected cards for the risks associated with a covid-19 contact tracing app and also cards for possible safeguards. They then discussed the challenges those safeguards might create.

6.2 Covid-19 contact tracing app risks

Cards selected: Data security, Data breach management, Trustworthiness

- All these cards are linked to trust. Healthcare data is the most sensitive data of all, especially when combined with location data. The public has to trust that the app works well, that their data is going to be used for the specified goal and that it, and they, will be protected. Confidentiality and Resilience cards also feed into trustworthiness.

Card selected: Confidentiality

- Confidentiality is key in contact tracing. Who gets access to the data and how might it be repurposed?

Card selected: Resilience and Low redundancy

- Vulnerability to hacking is a concern along with falsely reporting the data and manipulation on the technical and social side. Two-factor authentication could possibly be used to mitigate against this.

Cards selected: Privacy in public, Usable security

- How is the data used outside of its original intent? Could someone download the data and find out when you are away from your house? How much control do you have over this data? Does permission for the data to be collected now mean permission for it to be used in the future?

6.3 Covid-19 contact tracing app safeguards

Card selected: Transparency

- Explaining to the public how the technology works can help to engender trust.

Card selected: Liability

- If there is a data breach, someone will be held accountable.

Card selected: Consumer protection

- The app could be used to discriminate against people such as barring entry to places or making assumptions about them based on their location cluster. Data could be subverted to know where clusters of people are to target them for certain things such as selling something or, more seriously, a terror attack. How are consumers protected?

Temporality

- How long do we keep the information for?

Obfuscation

- How does the technology protect people's identities? Does it use anonymisation or pseudonymisation techniques?

Accessibility

- There is a need to safeguard against inequalities and the risk of excluding people, such as those without a smartphone.

Limited data collection

- Data should be collected at the most minimal level possible, which also links into temporality.

Data security

- Inadvertent release is not the same as deliberate sharing.

Lawful processing

- If we give consent for the app to track people for Covid data, and then we want to use the data for other types of research, how would we get consent for that?

Wellbeing

- If people are informed about how the data could benefit them and their lives they might be more willing to share data.

Power asymmetry

- Who holds the power to make the safeguarding decisions? What decisions are being made, and what is the thought process behind those decisions? If we can understand that then we can properly define the safeguards that are needed.
- Due to the significant societal need to rush this app out, it is being tested in the real world and not a lab. The key questions are, where does the power reside, what oversight and accountability is there and how can we as the users/the people make sure that these safeguards are implemented?

6.4 Covid-19 contact tracing app safeguarding challenges

The challenges are many and include:

- Obfuscation making the data useless when divorced from the people who provided it.
- False reporting due to wanting to restrict other people's movement or through excessive caution.
- Explaining the benefits, how the app works, why the data is needed, how the data will be used, what is the benefit of the app - but not everything can be explained. Or, there is the potential of revealing an unknown vulnerability with too much detail.
- How long to keep the data - the purpose of the application may change with the passing of time and not all of the data will be valuable. The data could be reused and repurposed for a different use.
- Balance of security v capacity to do data analysis. Can this be done without private data being exposed to anyone?
- NHSX centralised approach v Google / Apple API held on a user's phone. If you were using the app for scientific research you'd have to take the NHSX approach. Citizens might be happier about the Google / Apple approach.
- Does it go far enough? No opportunity for enforcement - is privacy more important than life? What if you have multiple mobile phones? Or do not have a smartphone at all?
- If the app doesn't provide anything past sending a message, would a human-based team be better?

- Uptake issues.
- GDPR and right to be forgotten. How technically possible is it for me to request my data back?
- It is political will that underpins the workings of this app. We don't know how many corners are cut because the app needs to be rolled out quickly. Is it being rolled out for health reasons or political ones (or both)?
- The only way to have properly assessed the ethics is to acknowledge that we've not asked all the questions, we've not answered all the questions and that there will always be new ethical issues to understand.

7 Conclusion

This thought-provoking and timely workshop from the AI^3 Science Discovery Network+set out clearly that ethical frameworks are useful but, alone, are not enough. The ethical decisions that have to be made when AI is used for scientific discovery are complex, involving trade-offs and judgements. We cannot resolve those trade-offs or rely on an algorithm to make the 'correct' ethical decision automatically on our behalf. However, what we can do is to be as transparent as possible about how and why those trade-offs have been made and ensure that as many different voices, values and interests are included in those discussions by engaging with diverse stakeholders and providing clear narratives. Engagement and communication are key along with – once again – recognising the power of augmented intelligence, which brings together the best of human and machine intelligence.

References

[1] 12th ACM Web Science Conference 2020. WebSci'20 [Internet]; 2020. [cited 2020 Jul 21]. Available from: https://websci20.webscience.org/.

[2] Urquhart L, Craigon P. The Moral-IT & Legal-IT Decks [Internet]; 2020. [cited 2020 Jul 21]. Available from: https://lachlansresearch.com/the-moral-it-legal-it-decks/.

[3] Commission E. Communication Artificial Intelligence for Europe [Internet]; 2018. [cited 2020 Jul 21]. Available from: https://ec.europa.eu/knowledge4policy/publication/communication-artificial-intelligence-europe_en?page=1.

[4] for Health NI, (NICE) CE. Improving health and social care through evidence-based guidance [Internet]; 2020. [cited 2020 Jul 21]. Available from: https://www.nice.org.uk/.

[5] Wikipedia. Quality-adjusted life year [Internet]; 2020. [cited 2020 Jul 21]. Available from: https://en.wikipedia.org/wiki/Quality-adjusted_life_year.

[6] Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nature Machine Intelligence. 2019;1(9):389–399.

[7] Tabletopia. Tabletopia [Internet]; 2020. [cited 2020 Jul 21]. Available from: https://tabletopia.com/.