Risk and Trust Perceptions of the Public of Artifical Intelligence Applications

Keeley Crockett[1], Matt Garratt[2], Annabel Latham[1], Edwin Colyer[1], Sean Goltz[3]
1School of Computing and Mathematics,
Manchester Metropolitan University, Manchester, M1 5GD, UK, K.Crockett@mmu.ac.uk
[2]School of Engineering and IT. University of New South Wales, PO Box 7916, Canberra BC 2610, ACT 2902,
Australia, m.garratt@adfa.edu.au
[3]Business & Law School, Edith Cowan University, Perth, Australia, n.goltz@gmail.com

Abstract— **This paper describes a study on the perceived risk and trust of members of the general public regarding artificial intelligence applications. It assesses whether there is a difference in the perceptions of risk and trust in artificial intelligence expressed by the general public compared with those studying computer science in higher education. We define the general public as people having no specific level or specialist knowledge of AI yet with a high stake as potential users of AI systems on a regular basis with or without their knowledge. In the study, participants engaged in an AI debate on topical news articles at a public national science museum event and a University in the UK and completed a questionnaire with two sections: their assessment of trust and risk of an AI application based on a topical news story, and a set of general opinion questions on AI. Results indicate that in specific applications there is a significant difference of opinion between the two groups with regards to risk. Both groups strongly agreed that education in how AI works was significant in building trust.**

Keywords- risk, trust, perception

I. INTRODUCTION

Trust in terms of Artificial Intelligence systems is not easy to define. The Cambridge English dictionary defines trust as "*the belief that you can trust someone or something: Their relationship is based on trust and understanding.*" [1]. Ferrario et al. [2] define AI trust through a multi-layer model designed to analyze human-AI interactions. They define reasons of trust in AI to be either pragmatic in the case of simple trust or epistemic when trust in AI interactions is reflective. This human-centered approach to trustworthiness has also been adopted by the European Commission's (EC) High-Level Expert Group on AI in their published Ethics guidelines for trustworthy AI [3]. The EC adopt Soau and Wang's definition of trust as "*Trust is viewed as: (1) a set of specific beliefs dealing with benevolence, competence, integrity, and predictability (trusting beliefs); (2) the willingness of one party to depend on another in a risky situation (trusting intention); or (3) the combination of these elements.*" [4] yet also acknowledge that the definition of trust is a more universal concept applied to more than just machines. Finally, in [3], the European Commission identifies the need to consider public policy and the effects of AI on the public "*to ensure trustworthy, human-centered AI systems*" are developed and deployed. In this paper we analyze public perceptions of trust through human interactions with AI systems.

Risk associated with AI systems can occur at multiple points in the AI system lifecycle [5], from conceptualization (application in an unethical way), data management (poor quality data or insufficient data governance), generating models from biased and non-representative data, to incorrect implementation and training of the "human in the loop", cybersecurity threats and technological environment errors. The public are influenced by the media reporting of AI; this is coupled with a lack of understanding of how an AI systems work, so people often amplify what they hear about particular systems in a negative way. A survey conducted in February 2019 on media trust, suggested that 32% of respondents from the US stated that they trusted news content most of the time, compared with 40% from the UK and 59% in Finland [6]. Organizations such as The British Heart Foundation recognize that fake news surrounding AI contributes towards the spread of misinformation and in 2019 stated that "*it is vital that as well as the NHS, other sectors including industry, charities and academia, must make the effort to improve public understanding about the developments of AI in healthcare to dispel any mistrust they feel towards it.*" [7]. The user's perception of trust and risk is multi-faceted in that trust increases the acceptance of an AI system, while increased perceived risk contributes to its rejection [8].

Grass roots education in understanding AI is therefore essential to building trust with the general public. In 2018, the Finnish Government aimed to educate 1% of European citizens in the basics of AI through a free online course entitled Elements of AI [9]. The course is designed for a wide range of people including business professionals, unemployed persons, dental assistants and pensioners – aged between 20 and over 75. The course takes between 5 and 10 hours to complete with only basic math's and no programming required [9]. The course however is described as "*a university-level online course, free and open for everyone*" [10] which may be a barrier to members of the public who feel they have not met the standard of education required to take a University course. For example, the EU reports [11] that in 2018, 35.2 % of EU citizens aged 25–54

held a university degree and just 21.7 % of those aged 55–74. Also 10.6 % of young people (aged 18–24) had only completed lower secondary school education, effectively leaving all opportunities for further academic education and training [11]. Therefore, fundamental and easily accessible courses about AI should be developed for those people who have non-academic skills and talents. It is therefore essential to find out current opinions that the public have on AI applications, and obtain more insight into what they understand and how much they trust AI systems.

The research in this paper attempts to answer the following research question:

Is there a difference in the risk and trust perceptions towards artificial intelligence expressed by the general public compared with those expressed by students studying computer science in higher education?

We assume that the general public have no specified knowledge level with regards to AI so the study therefore used a series of news articles to facilitate debate on different AI applications covering fake news, cybersecurity breeches in smart energy grids, medical diagnostics, healthcare, driverless cars and robotics. During the debate participants completed a two-part paper-based questionnaire which included free text answers about their opinions. For comparison purposes, a second independent group of people studying computer science in higher education also took part. The debates took place in a public science museum and in a University during October and December 2019. The results of the questionnaire analysis are presented in this paper.

This paper is organized as follows: Section II examines related work on risk and trust perceptions in AI and highlights the results of existing surveys that have attempted to capture public perception. Section III describes the experimental methodology and questionnaire design, while Section IV analyses and discusses the results on risk, trust and general perceptions of AI. Finally, Section V concludes and makes recommendations on how basic courses on AI could be designed for different education attainment levels and skills of the population.

II. RELATED WORK

A) Risk and Trust

Perry and Uuk [12] state that risks associated with AI systems fall into two categories: AI technical safety and AI governance (which includes political, military, economic, governance and ethical issues). The Future of Life Institute has identified two scenarios where they believe AI will pose a risk to society [13]. The first concerns autonomous AI systems that are programmed to kill and destroy human life, a risk highlighted by many international organizations such as the IEEE [14], the EU [3], the UK Office for Artificial Intelligence [15] and smaller companies such as morse.ai [16]. The second scenario is based on an AI system which benefits society but does so through developing "a destructive method for achieving its goal", generally when human and machine goals are misaligned.

Cheatham et al. [5] identify five areas that can lead to AI risks. 'Data difficulties' refers to the correct usage of data including compliance with the GDPR and other regulatory bodies. 'Technology troubles' refers to when an AI system fails to do the job as expected, for example missing a fundamental outlier. 'Security snags' is concerned with risks associated with AI-driven cybersecurity and the implications to the data and hence the data model. 'Misbehaving models' refers to models developed from biased or unrepresentative data. Finally, the nature of human-machine AI system interactions causes risks, for example due to the lack of understanding of the results by a human interpreter due to inappropriate training.

Bias is one of the biggest risks in using AI systems and this includes bias that is embedded into organizational or industrial cultures, personal and unconscious bias and data bias. Data bias must be considered and addressed in the selection of training data for AI systems. Data which has been labelled by humans for training may be subjective. Where training, validation and testing is dynamic and models continually evolve and learn, it should be monitored to ensure that there is no bias creep. It is also important to recognize that applications such as human profiling, and a one size solution to fit all humans may not be appropriate. Different models may need to be developed for different genders, cultures etc. as it may not be possible for the models to generalize on the human population. Arnold et. al. [17] propose the use of an AI factsheet which provides information on statement of purpose, basic performance, safety, security, and lineage which are aligned with AI trust principles. The aim is that suppliers of AI systems and services voluntarily populate these factsheets. Customers who purchase from these suppliers can therefore review the factsheet and decide if that product meets their ethical and data governance standards.

Explainable AI (XAI) is necessary in building trust so that users and subjects can understand how the AI made a decision. However, the big question is to whom? Solutions such as Google Cloud's AI Explanations product [18] include end users as stakeholders "*who want to understand a model prediction to incorporate it into their decision-making process*" [18]. This is not the same, for example, as providing an explanation to a user who applies for health insurance and is rejected based upon an AI system's automatic profile of them from facial micro-expressions [19]. Crockett et. al. [20] propose a new Hierarchy of Explainability and Empowerment that allows information and decision-making complexity to be explained at different levels depending on a person's own perception of their knowledge level.

Accountability of the AI system also affects trust and is difficult to determine due to the limited legislation and the lack of substantive case law [21]. In order to be accountable, decisions need be explainable so that errors can be identified. In Lord Sales's (Justice of the UK Supreme Court) 2019 lecture, the clear need for direction within the legal system was noted, "*we need to build a structure of legal obligations on those who design and operate algorithmic and AI systems*

which requires them to have regard to and protect the interests of those who are subject to those systems." [22].

Similar to all software products, usability and reliability of the AI system will also factor in how much people trust the system. Amershi *et, al.* [23] propose 18 human-AI interaction design guidelines to produce more usable, AI-centric systems. Data governance and data privacy also play a significant role in perceived trust, especially following the well published Facebook Cambridge Analytica scandal which continues to reveal leaked documents in 2020 [24].

B) Public Surveys on AI Perception

There has been an intensification of polls and surveys designed to capture the opinion of the public and businesses on AI in recent years. This section provides a summary of those that have had major influence. The UK Government poll on Artificial Intelligence: public awareness survey, surveyed 2,467 people online to understand public awareness of AI, and its benefits. The survey ran from 9 April to 15 April 2019 [25]. Two interesting findings were that 75% of men said they knew something about AI compared with 53% of women and that 74% of people aged under 45 knew something about AI compared with 54% over [25]. The survey report stated that "*A lack of knowledge about AI is preventing people from knowing the impact it could have on the economy and jobs, with half not knowing how many jobs would be created and a similar proportion unsure how much money would be added to the economy through AI.*" [25]. The Global Artificial Intelligence Survey by ARM [26] interviewed 3,938 consumers across eight countries using an online survey; participants were pre-screened to ensure they had basic knowledge of AI. Whilst the results regarding knowledge and understanding agree with the UK Government poll, answers given by males and females tended to agree. 61% of people stated they thought AI would change society for the better. The analysis found that "*..as AI technology is applied in ways that help preserve human health and enhance people's quality of life, we can expect to see public opinion swinging even more forcefully in favor. However, this is predicated on AI developers ensuring the technology is safe and secure*" [26]. This survey also found that 85% of participants were concerned about the security of AI systems [26]. In July 2019, IPSO conducted a global poll for the World Economic Forum that surveyed 20,107 adults from 27 countries [27]. The survey found that regardless of educational level, people (41%) were just as likely to worry about the use of AI in general in society, and approximately 40% wanted to restrict government use of AI more with almost 49% wanting regulation of business usage. In 2019, Edelman also conducted an AI survey which compared answers of the U.S. general public's perceptions of AI with senior technology executives [28]. 100 members of the public were surveyed compared with 300 executives. One major conclusion was that "*respondents see benefits but think AI will benefit the wealthy, harm the poor and cause societal disruption.*" AI regulation was also seen as important, with 60% of the general population and 54% technology executives seeing this as essential. A study published by the Center for the Governance of AI (GovAI) [29] involving 2,000 American adults found that after reading a short definition of AI, 41% supported its development, compared to 22% who opposed it. The study found that the most trusted organizations to develop AI were universities and the U.S. military; Facebook was least trusted [29]. However, the study concludes that due to the sample size, results should be interpreted with caution and that a more substantial study was required.

All these surveys, however, lack a robust description on what constitutes the general population. Moreover, only one of the studies takes the educational level of respondents into consideration (participants were asked if they considered their education level to be low, medium or high). It is not only educational level that might shape a person's perception of AI, but also interactions in their daily life, an experience at work (e.g. being made redundant due to automation [30]), or reading a media article that may not contain factual information. Going forward, education on AI systems is required for all members of society regardless of age range, educational and social economic background.

III. PUBLIC PERCEPTION EXPERIMENTAL METHODLGY

A) Overview and Ethics

This section describes a public perception study called "*You, me and "AI": What's the risk in giving AI more control?*" that took place as part of a large-scale public engagement event - Cybersecurity and AI Playground at the Manchester Museum of Science and Industry, UK, in October 2019. The museum has free entry and attracts individuals and families from a wide variety of social and economic backgrounds. The event was part of a series of science public engagement activities organized by the museum to attract families during a school holiday.

In order to conduct this study, a full ethical application was submitted and approved by Manchester Metropolitan University. As participants were asked to engage in a debate and complete and return a questionnaire, a covering letter, similar to a participant information sheet was produced for participants to take away. No consent form was needed as consent was implied by returning the questionnaire. The covering letter explained that the researchers would ask participants a series of questions on artificial intelligence with images from recent topical stories. A question/answer sheet was provided for participants to write their answers or choose opinion to each question. No personal identifiable information was collected and thus answers were anonymous. To enable a comparison of opinions, students in the Faculty of Science and Engineering, Manchester Metropolitan University also had the opportunity to take part in the same debate (December 2019) to enable their perceptions to be captured and compared with the general public who attended the Cybersecurity and AI Playground.

This study was designed to test the following hypothesis:

H₀: There is a statistical significant difference in the risk and trust perceptions towards artificial intelligence expressed by the general public compared with those studying computer science in higher education.
H₁: There is no difference in the risk and trust perceptions towards artificial intelligence expressed by the general public compared with those studying computer science in higher education.

B) Questionnaire design

The questionnaire was designed to facilitate debate and discussion to explore how people perceive trust and risk in relation to AI systems. The questionnaire was divided into two parts. The first section presented a series of questions which first showed a recent relevant news story and then asked the participant how much they would trust the type of AI system featured in the story (Answers were Yes; No; Abstain). Participants were also asked to rate the risk of the system on a scale of [0-10] were these systems to be used. On this scale, a 0 represented no perceived risk and 10 represented high perceived risk. For example, questions 1 and 2 were related to the use of AI to detect deep fakes [31.33]:

Q1. Do you trust AI to detect fake videos?
Q2. On a scale of 0 to 10, how much are you willing to risk trusting the content of videos that are posted online?

Deep fakes create digital impersonations of people from audio and videos through the use of deep neural networks by creating and inserting synthesized faces [34]. Their impact is significant in that they can be used to create interviews and events that actually never occurred. In December 2019, the Cyberspace Administration of China announced a new law criminalizing the publication of fake news content that uses artificial intelligence or virtual reality [33]. Research continues to develop automated deep fake detection algorithms but is continually playing a catch up exercise [34]. For each question pair, first there was a discussion led by the researcher about the topic, using a media article (Figure 1) to stimulate the debate. The participants were asked to record their answers to the two related questions on a question sheet.

Figure 1. Deep fake article from IEEE Spectrum [31]

The second section of the questionnaire comprised a set of statements with a response using 5 point Likert scale [35] where 1 represented *strongly disagree* and 5 represented *strongly agree*. These questions were designed to assess a participant's opinion on trust, bias, explainability, ethics and whether they supported the development of such systems. A full list of questions can be found in Tables I, II and III in Section IV Results and Discussion.

C) Participants

Two groups of participants took part. Group 1 were members of the public who attended the museum event (*N=54*) and Group 2 were those who were studying in a course in the Faculty of Science and Engineering (*M = 25*) and hence had some knowledge of computer science. These groups were chosen to see if there was a difference in trust and risk perceptions between the general public and those studying computer science.

IV.   RESULTS AND DISCUSSION

Table I shows the responses to questions regarding Trust showing whether participants answered (Y)es, (N)o or (A)bstained. The results are also presented visually in Figure 2. The correlation of Yes responses between groups was 0.91 and for no responses 0.93. The percentage of abstain responses from the general public across all questions was 20% compared with 36% for University students. It is possible that students from a science and engineering discipline had more scientific knowledge about certain applications and therefore had a more informed opinion than the general public.

Table I: Trust Questions

Q'No	Question	Group 1 %			Group 2 %		
		Y	N	A	Y	N	A
1	Do you **trust** AI to detect fake videos?	63	35	2	52	29	19
3	Do you **trust** an AI system to make a diagnosis from medical images?	65	26	9	57	33	10
5	You have a tumour, but would you **trust** a diagnosis from an AI system?	52	28	17	48	43	10
7	Would you **trust** an AI cybersecurity system to find and fix vulnerabilities on your electronic devices?	76	20	4	71	29	0
9	Do you **trust** AI to filter your spam email if it thinks it is malicious?	74	19	7	90	5	5
11	In 2017, the NHS was crippled by an international cyber-attack. Would you **trust** an AI system to predict attacks in advance and provide early warning?	70	22	7	71	19	10
13	Fake voice recordings can impersonate powerful people.	0	96	4	10	90	0

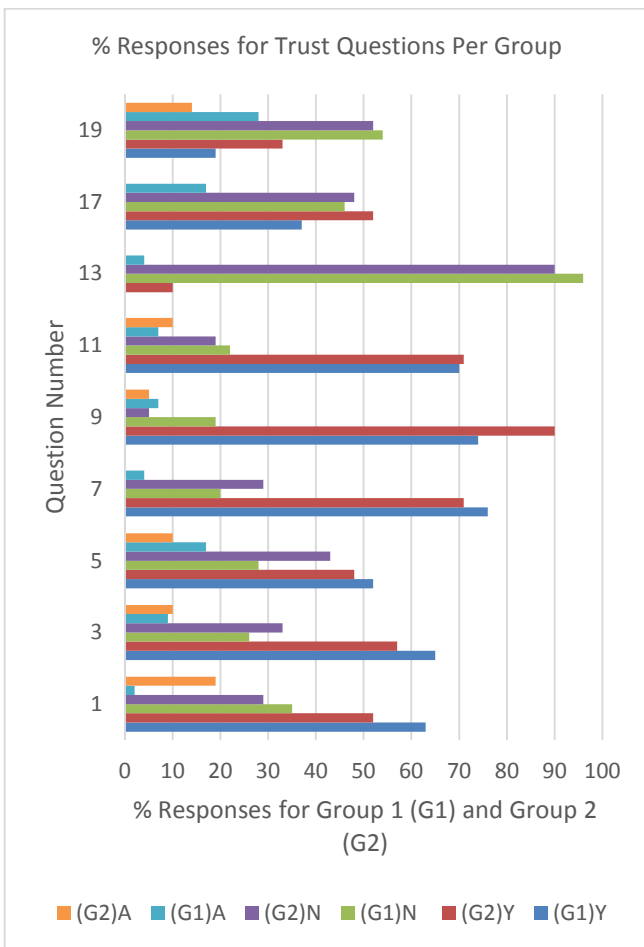| Q'No	Question						
	Would you **trust** an automated message from your boss asking you to deposit money in an account?						
17	Driverless cars will need a 'digital MOT' to check they can't be hacked. - Would you **trust** your safety to a car that had passed this test?	37	46	17	52	48	0
19	Security robot Pepper is crammed with cameras and sensors. Now it is hurtling towards you at top speed… - Would you **trust** that Pepper was protecting you and hadn't been hacked to cause you harm	19	54	28	33	52	14

Figure 2: % of Yes / No Responses for Trust Questions

The results show that the issue of trusting content of automated voice recordings (Q13) generated the most number of No responses across both groups (96% and 90% respectively). Both groups may have been influenced by the heavy media attention given to deep fakes at the time of the study. Both groups agreed that they could trust AI to filter potentially malicious email (Q9) and to find and fix

vulnerabilities in personal electronic devices (Q19). The groups differed on Q17, which looked at a news article that discussed how driverless cars would be required to have a digital MOT which would ensure they could not be hacked. 52% of University students (Group 2) would trust a car's safety based on this MOT compared with only 37% of the general public (Group 2). Generally, the results show there is some differences in opinion between the two groups, but it was dependent on the application they were asked whether they trusted or not.

Table II shows the responses to questions regarding Risk, which were scored on a scale of 0 to 10 with 0 indicating no risk and 10 indicating high risk. Median values are shown for the two groups along with the *p-value* derived from the Mann-Whitney statistical test.

Table II: Risk Questions

Q' No	Question	Group 1	Group 2	p-value
2	On a scale of 0 to 10, what is the risk to you of trusting of the content of videos that are posted online?	5	7	0.05155
4	On a scale of 0 to 10, what is the risk to you of trusting an automated diagnosis of your condition from your medical image?	5	7	*0.02382*
6	On a scale of 0 to 10, what is the risk to you of trusting an overall diagnosis of your medical condition by only an AI system?	6	8	0.02088
8	On a scale of 0 to 10, what is the risk to you of trusting an AI program to keep your devices safe from hackers?	5	5	0.29834
10	On a scale of 0 to 10, what is the risk to you of giving full control to an AI system to delete what it thinks are malicious emails?	5	5	0.267
12	On a scale of 0 to 10, what is the risk to you of trusting an AI system falsely identifying you as a hacker and stop you booking an appointment online?	5	7	*0.00672*
14	On a scale of 0 to 10, what is the risk to you of following instructions from a voice you trust?	6.5	9	*0.0164*
16	On a scale of 0 to 10, what is the risk to you of trusting AI systems in making power grids more vulnerable to hackers?	5	5	0.5485
18	- On a scale of 0 to 10, what is the risk to you that someone hacks into a driverless cars and controls it remotely?	8	9	0.28914
20	On a scale of 0 to 10, what is the risk to you of leaving your personal security to AI systems?	7	8	0.29834

The results in Table II show that for four of the risk questions there was a significant difference of opinion between the two groups *(p-value < 0.05)*. These questions related to risks associated with medical imaging and diagnosis (Q4 and Q6), an AI system falsely identifying them as a hacker (Q12) and the use of deep fakes for voice recordings (Q14). Both groups associated medium risk with AI taking greater control over cybersecurity (Q8, Q10 and Q15). Figure 3 shows a visualization of the median risk scores for each question. Green (Group 1) represents the general public and blue represents the higher education students (Group 2). Visually, figure 3 shows that in general, the perceived risk of AI applications is higher in Group 2.
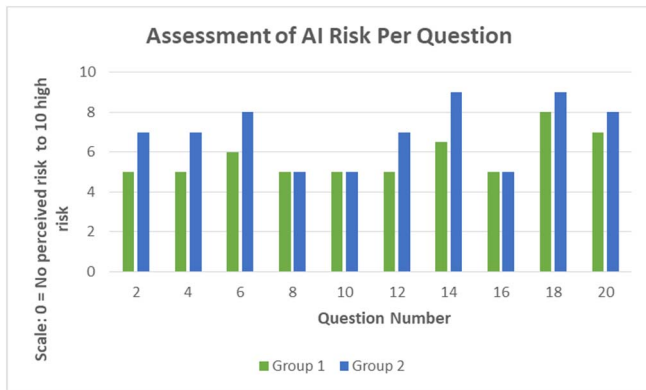

Figure 3: AI Risk Perception per Question

Table III shows the median responses to a series of statements scored on a 5 point Likert scale, where 5 - strongly agree, 4 - somewhat agree, 3 - neither agree nor disagree, 2 - somewhat disagree, and 1 - strongly disagree.

Table III: Median Summary of AI perceptions Statements

Q' No	Statements	Group 1 Median	Group 2 Median
1	"I understand how artificial intelligence works"	4	4
2	"I support the development of AI systems"	3	3
3	"The use of AI systems is ethical"	3	3
4	AI is a technology that requires careful management"	5	5
5	"I believe that the minority of AI systems are biased"	3	3
6	"Explaining how an AI system makes a decision is important to me"	5	4
7	"The majority of AI systems are fair"	3	3
8	"I believe the benefits of using AI systems outweigh the risks"	4	4
9	"I would like to be involved in developing an ethical code of AI"	4	4
10	"Educating people in how AI works would build trust"	5	5

Application of the *t-test* to examine how significant are the differences between the two groups gave a p-value of 0.11 implying that there is only an 11% chance that the participant could come from the same group, however this is not significant. On analysis of the results in Table III, the median value indicated agreement between groups. Question 6, *"Explaining how an AI system makes a decision is important to me"* was the only question where there was a difference in opinion. The general public strongly agreed with the statement, whereas students somewhat agreed. Free text comments received from the general public included:

- *"We should not let development accelerate faster than ethics"*
- *"we must .. and should be able to understand how deep learning algorithms work"*
- *"AI still requires human input – we should never become a black-box' society which doesn't check and understand the outputs of AI"*
- *"the management, documentation and transparency of AI systems is critical to its success and for people to trust its implementation"*

Both groups strongly agreed that education in how AI works was significant in building trust, one participant wrote *"If I knew how AI worked and where it was used it would allow me to feel more confident in asking questions"*. A majority of participants felt they would like to be involved in developing an ethical code of AI that represented the opinions of the general public.

V.   CONCLUSIONS AND FURTHER WORK

This results of the risk and trust perceptions study reported in the paper have a clear message that the general public is uncomfortable about *"being left behind"* in research and development of AI systems. Perception of risk is greater when the outcome of an error is more personal or serious (e.g. life and death) and therefore it is clear that there is a need to address people's concerns, especially in specific areas of AI application.

A majority of participants felt they would like to be involved in developing an ethical code of AI that represented the opinions of the general public. Despite there being a number of international initiatives [14-16], they all take a top-down approach to guidelines and regulation and do not take into consideration the general public (end user) voice of concern. In Greater Manchester, UK, a group of academics and businesses are focusing on the creation of an AI Charter for Ethical AI that gives more transparency around compliance with key AI principles. The approach taken in developing this charter is more "bottom-up" and is more suitable to integrating the public voice into debate and discussion. Through focused events, the general public will have the chance to engage in AI fundamental educational activities and speak to academics to help their understanding, and businesses about how their AI systems actually work and make decisions that may affect the public. By facilitating knowledge exchange the aim is to empower all members of the general public. Knowledge and education is essential to enable informed debate and discussion around ethical AI.

There is much further work to undertake. Firstly, we will evaluate and update questionnaires used in this study to include more detail about educational attainment level (both actual and perceived) and employment background. Secondly, we will develop, trial and evaluate a short workshop on AI fundamentals and trial with different subsets of the general public. We will examine public trust and risk perceptions before and after the workshop and look to compare results against subsets who undertake the Finnish Elements of AI online course. Thirdly, we will review the school's curriculum (keys stage 2 and 3 in the UK) to support younger generations to be aware of ethical AI topics such as bias, trust and risk of using AI applications.

REFERENCES

[1] Cambridge English Dictionary, [online], Available: https://dictionary.cambridge.org/dictionary/english/trust, [Accessed 6 Dec. 2019]

[2] Ferrario, A. Loi, M. Viganò, E. (2019), In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions, Philosophy & Technology, https://doi.org/10.1007/s13347-019-00378-3

[3] Ethics guidelines for trustworthy AI (2019), [online], Available: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, [Accessed 6 Dec. 2019]

[4] Siau, K., Wang, W. (2018), Building Trust in Artificial Intelligence, Machine Learning, and Robotics, Cutter Business Technology Journal (31), S. 47–53.

[5] Cheatham, B. Javanmardian, K. Samandari, H. (2019), Confronting the risks of artificial intelligence, McKinsey, [online], Available: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence# [Accessed 6 Dec. 2019]

[6] Share of adults who trust news media most of the time in selected countries worldwide as of February 2019 (2019), [online], Available: https://www.statista.com/statistics/308468/importance-brand-journalist-creating-trust-news/

[7] Mcdonald, A, (2019) Improve public understanding of AI, or risk spread of fake news report warns, [online], Available: https://www.bhf.org.uk/what-we-do/news-from-the-bhf/news-archive/2019/may/improve-public-understanding-of-ai-or-risk-spread-of-fake-news-report-warns

[8] Yanushkevich, S. Howells, G. Crockett, K. O'Shea, J. Oliveira, H.C.R, Guest, R. Shmerko, V. (2019) Cognitive Identity Management: Risks, Trust and Decisions using Heterogeneous Sources. In: Proceedings of the First IEEE International Conference on Cognitive Machine Intelligence. Proceedings of the First IEEE International Conference on Cognitive Machine Intelligence, *in press*, Winner Best Paper Award.

[9] Elements of AI, (2018), [online], Available: https://www.elementsofai.com/eu2019fi, [Accessed 7 Jan. 2020].

[10] Elements of AI Questions and Answers Press Report (2019), [online], Available: https://eu2019.fi/en/presidency/elements-of-ai [Accessed 7 Jan. 2020]

[11] Educational attainment statistics, (2019), EU Commission, [online], Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Educational_attainment_statistics#Level_of_educational_attainment_by_age, [Accessed 7 January 2020].

[12] Perry, B. Uuk, R. (2019), AI Governance and the Policymaking Process: Key Considerations for Reducing AI Risk, Big Data and Cognitive Computing, Vol3(2):26, Available: https://doi.org/10.3390/bdcc3020026

[13] Future of Life (2020), [online], Available: https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1, [Accessed 5th Jan. 2020].

[14] Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2018) [online], Available: https://ethicsinaction.ieee.org/, [Accessed 10th Jan. 2020].

[15] Office for Artificial Intelligence - UK Government, (2020), [online], Available: https://www.gov.uk/government/organisations/office-for-artificial-intelligence, [Accessed 10th Jan. 2020].

[16] Morse.ai (2020), [online], Available: http://www.morse.ai/ [Accessed 10th Jan. 2020].

[17] Arnold, M., Bellamy, R.K., Hind, M., Houde, S., Mehta, S., Mojsilović, A., Nair, R., Ramamurthy, K.N., Olteanu, A., Piorkowski, D. and Reimer, D., 2019. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. IBM Journal of Research and Development, 63(4/5), pp.6-1.

[18] AI Explainability Whitepaper, (2019), Google, [online], Available: https://storage.googleapis.com/cloud-ai-whitepapers/AI%20Explainability%20Whitepaper.pdf, [Accessed 10th Jan. 2020].

[19] Neil, C. (2019), China Knows How to Take Away Your Health Insurance, [online], Available: https://www.bloomberg.com/opinion/articles/2019-06-14/china-knows-how-to-take-away-your-health-insurance, [Accessed 10th Jan. 2020].

[20] Crockett, K. Stoklas, J. O'Shea, J. Krügel, T. Khan, W. Reconciling Adapted Psychological Profiling with the New European Data Protection Legislation, Computational Intelligence, Eds: Sabourin, C. Mereio, J. Barranco, N. Madani, K. Warwick, K. Springer, *in-press*

[21] Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Schieber, S., Waldo, J., Weinberger, D. and Wood, A., 2017. Accountability of AI under the law: The role of explanation. arXiv preprint arXiv:1711.01134.

[22] Algorithms, Artificial Intelligence and the Law, The Sir Henry Brooke Lecture for BAILII, Freshfields Bruckhaus Deringer, London, Lord Sales, Justice of the UK Supreme Court, 12 November 2019, [online], Available: https://www.supremecourt.uk/docs/speech-191112.pdf, [Accessed 10th Jan. 2020].

[23] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K. and Teevan, J., (2019), April. Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (p. 3). ACM.

[24] Cadwalladr, C. (2020), Fresh Cambridge Analytica leak 'shows global manipulation is out of control, The Guardian, [online], Available: https://www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation, [Accessed 9th January 2020]

[25] Artificial Intelligence: public awareness survey, (2019), [online], Available https://www.gov.uk/government/publications/artificial-intelligence-public-awareness-survey, [Accessed 17 Dec.2019].

[26] Global Artificial Intelligence Survey (2019), [Online], Available: https://www.arm.com/solutions/artificial-intelligence/survey. [Accessed 17 Dec.2019].

[27] Ipsos Global Poll for the World Economic Forum, (2019), [online], Available: https://www.ipsos.com/sites/default/files/ct/news/documents/2019-07/wef-ai-ipsos-press-release-jul-2019.pdf, [Accessed 17 Dec.2019].

[28] Edelman's 2019 Artificial Intelligence (AI) Survey, (2019). [online], Available: https://www.edelman.com/research/2019-artificial-intelligence-survey, [Accessed 17 Dec.2019].

[29] Zhang, B. Dafoe, A. (2019), Artificial Intelligence: American Attitudes and Trends, Center for the Governance of AI, Future of Humanity Institute, University of Oxford, January 2019 available: https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/

[30] Crockett, K. Goltz, S. Garratt, M. (2018), GDPR Impact on Computational Intelligence Research, IEEE International Joint conference on Artificial Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2018.8489614, ISSN: 2161-4407

[31] Hsu, J. (2019), Can AI Detect Deepfakes To Help Ensure Integrity of U.S. 2020 Elections?, [online] Available: Source: https://spectrum.ieee.org/tech-talk/robotics/artificial-intelligence/will-deepfakes-detection-be-ready-for-2020 [Accessed 3 Dec. 2019]

[32] Chesney, R. Keats, D. (2019), Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," 107 California Law Review (2019, Forthcoming); U of Texas Law, Public Law Research Paper No. 692; U of Maryland Legal Studies Research Paper No. 2018-21.

[33] Solovieva , D. (2019), China Steps Up Deepfake Policing, Banning Fake News As Problem Spreads, [online], Available: https://karmaimpact.com/china-steps-up-deepfake-policing-banning-fake-news-as-problem-spreads/ [Accessed 4 Dec. 2019]

[34] Yang, X., Li, Y. and Lyu, S., 2019, May. Exposing deep fakes using inconsistent head poses. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8261-8265). IEEE.

[35] 1. de Winter, J. C., & Dodou, D. (2010). Five-point Likert items: t test versus Mann-Whitney-Wilcoxon. Practical Assessment, Research & Evaluation, 15(11), 1–12