DOI: https://doi.org/10.1109/SSCI50451.2021.9660153

Publisher: IEEE

Version: Accepted Version

The Ethical Landscape of Data and Artificial Intelligence: Citizen Perspectives

Keeley Crockett Senior Member IEEE[1], Edwin Colyer[2], Annabel Latham Senior Member IEEE[1]
[1]Department of Computing and Mathematics,
Manchester Metropolitan University, Manchester, M1 5GD, UK, K.Crockett@mmu.ac.uk
[2]Directorate for Research and Knowledge Exchange, Manchester Metropolitan University, Manchester, M1 5GD, UK

Abstract— **Globally, there is growing acknowledgement that those involved in the development and deployment of AI products and services should act responsibly and conduct their work within robust ethical frameworks. Many of the ethical guidelines now published highlight a requirement for citizens to have greater voice and involvement in this process and to hold actors to account regarding compliance and the impacts of their AI innovations. For citizens to participate in co-creation activities they need to be representative of the diverse communities of society and have an appropriate level of understanding of basic AI concepts. This paper presents the preliminary results of a longitudinal survey designed to capture citizen perspectives of the ethical landscape of data and AI. Forty participants were asked to participate in a survey and results were analyzed based on gender, age range and educational attainment. Results have shown that participant perception of AI, trust, bias and fairness is different but related to specific AI applications, and the context in which is applied. Citizens also are also very receptive to undertaking free courses/workshops on a wide range of AI concepts, ranging from family workshops to work-based training.**

Keywords- ethics, citizen, data, artificial intelligence

I. INTRODUCTION

Following the European Union's publication of the Proposed Regulation Framework on Artificial Intelligence [1] (April 2021) and the US Government's "Guidance for Regulation of Artificial Intelligence Applications" [2] (June 2021), there is a new emphasis on citizen involvement in the development of AI solutions. There are other publications that place a strong requirement on academics, public and private sector businesses to involve citizens in the conceptualization and development of AI products and services that have a direct impact on them [3,4]. The need for new technology to have a public "license to operate" is growing as public and private organizations and businesses face media backlash [6,7] and severe consequences (e.g. loss in revenue) if they do not develop ethical AI solutions. As AI is further incorporated into our everyday lives, citizens may not be aware of the extent to which automated decisions affect them; they are effectively disempowered from scrutinizing the "how" and "why" of AI-powered systems. Policy makers and advocacy groups, aware of this growing divide between developers, decision subjects and wider society, realize that greater co-operation between all parties is vital for public trust in AI technologies. In 2020, the UK AI Council made a recommendation in its AI Roadmap to "Ensure public trust through public scrutiny" [8] and the World Economic Forum emphasized the role of civil society in bringing the public voice to the table in the development of responsible AI [4]. From a Nordic perspective, Robinson [9] concludes that without citizen involvement and education in AI there is a risk of alienating those who are supposed to benefit from this technology.

If AI is to benefit everyone in society, then everyone, irrespective of their educational level, needs to comprehend at some basic level what AI is, along with concepts such as bias and fairness, and how decisions made by an AI system are driven by a model derived from data. Decision subjects should be aware of their right to an explanation of automated decisions. In the absence of these "concepts of ethical AI" in any current national curriculum, it is important to establish current levels of awareness, knowledge, and prevailing attitudes among citizens. Many surveys and studies [10..12] have captured societal viewpoints of AI over previous years, but they typically reach people who are relatively digitally literate, such as university graduates, rather than a more representative sample of populations or incorporating analyses of specific, seldom heard subgroups i.e. those aged 60 and over. In addition, as the citizen viewpoint is highly dynamic, even recent studies may fail to capture current views that may have shifted following public discourse around AI ethics, for example as a consequence of OfQual's failed A-level grade prediction algorithm [6] and the use of the NHS track and trace algorithm beyond the pandemic for surveillance [7]. A variety of free online AI and ethics courses are currently available [13..18] that seek to educate the public in AI. However, they are often targeted at those with a certain educational levels, assume a level of digital literacy and/or access to digital resources, or require an expert facilitator.

This paper presents the results of a novel survey which seeks to investigate how a citizen's age, gender and educational attainment affects their perception of the ethics of data and AI in general and their trust of the use of AI in different settings and investigates the education and training needs of the general population in AI. The research presented in this paper attempts for the first time to investigate and analyze citizen perspectives of AI concepts in terms of their age, gender and educational attainment through a survey which allows deeper exploration of understanding through two cases studies: the use of AI in a loan classification system, and the accuracy verses ethical usage of a deception detection system. The research questions to be addressed are:

RQ1: How does a person's age, gender and educational attainment affect their perception of the following: AI in general, ethics, AI applications, understanding of bias, who is responsible for decisions made by an AI system, explainability of a decision and the role of a human in an AI system?

RQ2: How does a person's age, gender and educational attainment level affect their perception of bias, fairness, and trust of the use of AI in different applications?

RQ3: What are the needs and requirements of the general population for education and training in AI?

The survey is part of a longitudinal study to look at how opinions of citizens with regards to their awareness of trust, fairness, bias and explainability change over time as AI applications become more abundant. In this paper we present the first quarter results from 40 participants.

This paper is organized as follows. Section II presents a summary of related work in the development of trustworthy and responsible AI, the citizen viewpoint and the current state of AI ethics education. Section III and IV present the survey methodology, results, and discussion. Finally, Section V, describes the responsibilities of the academic community in the role of citizen education and presents further work.

II. RELATED WORK

Governments, organizations, and businesses continue to publish ethical guidelines, principles and polices for the use of AI [1,2]. For example, Schiff *et al.* [19] reviewed 112 documents on principles, frameworks, and guidelines on ethics from 25 countries across private and public industry sectors which highlighted differences in key ethical topics such as bias and accountability. The depth to which an organization engages in responsible and ethical AI depends on which guidelines/principles they adopt (or develop) and the resources they have available to implement them. This section examines the challenges of citizen engagement in developing responsible AI and reviews current free courses on offer for those citizens who would like to learn more.

A) Developing responsible AI

There are many definitions of responsible AI. Responsible AI as defined by Buhmann *et al.* [20] has three dimensions: the responsibility to do no harm, the responsibility to do good and a responsibility for good governance. For an AI system to be trusted by citizens, all aspects of responsibility and accountability to stakeholders and decision subjects need to be defined and understood. When transparency is poor, decision explainability is confused or non-existent, and the accountability chain either breaks or is obscured [21]. It becomes difficult for citizens to engage in dialogue or scrutinize the work of system developers. It is firmly believed that to develop responsible AI, it is essential to breakdown the barrier of explainability in deployed AI systems so all stakeholders, regardless of role, understand how a model, induced from data, has made a decision, along with the caveats of that decision. Arrieta *et al.* [21] undertook a systematic review of eXplainable Artificial Intelligence (XAI) and produced a global taxonomy of different XAI techniques and highlighted many XAI challenges such as model interpretability, especially within deep learning, constraints of data privacy and the confidence level of a model. Such techniques are mainly targeted as those designing and creating AI solutions and not suitable for explaining to general citizens. To succeed in developing responsible AI, employees need organizational structures, policies, and procedures in place to support them [23].

A further challenge is in understanding where citizens are in terms of not only their perception of how their data is used and what AI is, but also if they feel digitally excluded or suffer from digital or data poverty. Lack of access to the right equipment or broadband speeds should not exclude a person from participating in responsible AI juries, panels, and debates. Organizations that seek to co-create AI with the public need to appreciate existing levels of knowledge, skills, and awareness among their citizen collaborators so that interactions are meaningful and based on common understanding.

B) Ethics, AI and education

It is challenging to design a course on AI and ethics which can be accessible to everyone, regardless of digital literacy and educational attainment. Improved citizen involvement in developing responsible technology comes from education [5]. In this section, we review a recent sample of free courses to highlight some of the limitations (e.g., educational barriers) associated with such courses.

Data and/or AI ethics is being introduced more significantly into both undergraduate and postgraduate courses within universities globally. For example, Practical Data Ethics is a series of six free online lessons that was originally delivered as an evening course at the University of San Francisco and is now be offered by Ai.Fast [16]. The syllabus (not exhaustive) is sourced from an analysis of 100 technical ethics syllabus [17] and covers misinformation, bias and fairness and algorithmic colonialism. [18] maintains a list of over 292 courses currently running tech ethics curricula at university level. Whilst this is excellent for training future graduates, the courses are designed for those who meet the educational attainment required to attend university. For those already working as AI developers and data scientists, there are numerous online, self-taught courses available; Kaggle offers a series of five "Intro to AI Ethics" tutorials which cover human-centered design, bias, fairness and how to use model cards [24].

The Ethics of AI is a relatively new course run by the University of Helsinki [13] and was designed for anyone (public administration, businesses, and the general public [13]) with an interest in the ethical aspects of AI. The course builds on the free Elements of AI course [14] which was restricted to those with university entry qualifications. The course covers fairness, human rights, accountability, transparency and non-maleficence and registration is required to compete the exercises. Although the course is open to anyone, it is recommended that participants have "*familiarity with basic AI concepts*" [13] which could be a limiting factor for those with lower educational levels. "We are AI" is a five-week course to introduce citizens to AI, launched in May 2021 by the Centre for Responsible AI at New York University's Tandon School of Engineering [15]. This course features five engaging modules including "What is AI?", "Learning from Data", and "All about Bias". Each module is designed to be delivered as a learning circle and run by a facilitator (virtual or in person) in 90-minute sessions. The creative learning resources are diverse and inclusive and include videos, comic books, and activities, making this course very accessible. The key motivation of the course was that citizens are largely unaware of the influence that AI has on their day-to-day lives. Alexandre *et al.* [5] built a hybrid AI MOOC for citizens aged 15 and over, which features short videos, quizzes, online activities, and unplugged activities for families. Analysis of the first three months revealed more than 13,000 persons had engaged in the MOOC, and 600 had completed it. Evaluation of the MOOC (1140 respondents at the start of the course and 217 at the end-of-course) had positive feedback and several areas for improvement, but the authors acknowledge that the analysis is biased, due to people who have experience in MOOCs tending to respond. This brief review shows that free courses are emerging that are appropriate for citizens to learn, but they are fragmented in approach, often require facilitation by a person with sufficient knowledge in AI and ethics, and prerequisite levels of education or digital literacy for enrollment. Current offerings also do not address those suffering from digital poverty.

III. SURVEY METHODLOGY

A) Study design

This novel study aims to capture information from a diverse group of citizens about their current knowledge, existing skill sets, and moral behaviors toward AI applications and services that impact. It analyzes their perspectives of AI concepts against their age, gender, and educational attainment. Asking specific questions about two cases studies: the use of AI in a loan classification system, and the accuracy verses ethical usage of a deception detection system will allow citizens to contextualize their responses and allow analysis on the depth of understanding. This will be achieved through collecting data thorough a longitudinal survey which opened on 15th May 2021. Results in this paper are presented on data collected up to 1st August 2021. The survey comprises 35 questions and is available online (https://mmu.onlinesurveys.ac.uk/public-opinions-on-artificial-intelligence). The questions have been formulated based on a) reviews of recent national and international public surveys [19,25]; b) discussion with community groups in the UK; c) public engagement activities carried out at the Museum of Science and Industry, Manchester UK in October 2018 and October 2019 [26]; and d) a review of educational courses for members of the public to learn about AI. Table I lists all demographic questions in the survey, whilst Table II lists the subset of questions from the survey relevant to the specific aim of this paper. *RQ1* is analyzed using the following survey questions (2,3,4,6,7,14,15,17,18), *RQ2* using the following survey questions (2,3,4,6,7,14,15,17,18,22 - 28) and finally *RQ3* was analyzed using the following survey questions (32,33).

TABLE I. DEMOGRAPHIC QUESTIONS

Q-no	Question
1	I confirm that I have read the participant information above. Please provide or decline your consent
2	What gender do you identify as? If you selected Other, please specify:
3	What is your age?
4	What is your ethnic group? If you selected Other, please specify:
5	5a. Which country do you currently reside? 5b. Which Town or City do you live in?
6	What is your current education? If you selected Other, please specify:
7	What is your current employment status?

TABLE II. SUBSET OF SURVEY QUESTIONS

Q-no	Question
8	In one or two sentences, please write what you understand by the term "artificial intelligence"
9	In this survey, we use the following definition of Artificial Intelligence. - Machines which learn use artificial intelligence (AI) to enable computer applications to make decisions by learning from experiences without the help of a human. They learn through looking for patterns in data that has been previously used by humans to make decisions about how to do a task. If the data is good quality and representative of the human population then the AI could learn to make better consistent decisions. Do you think this definition of artificial intelligence is clear?
10	What does "being ethical" mean to you in your everyday life?
14	On a scale of 0 to 5 how much would you trust the use of artificial intelligence (where 0 indicates no trust and 5 indicates full trust) in the following areas {12 areas}
15	On a scale of 0 to 5 how much would you trust the use of artificial intelligence ***if being used on you, your family, or friends*** (where 0 indicates no trust and 5 indicates full trust), in the following areas: {same list as Q15}
18	An AI application only makes decisions based on what it has learnt from the data it has been given. So, who should be responsible when things go wrong? E.G., a person is mis-diagnosed, or a person is rejected for a loan? For each statement indicate your opinion on the scale from always responsible to not at all responsible. {8 statements}
22.25	Case study 1 questions
27.28	Case study 2 questions
32	If you were offered free courses about AI, what would you like to learn about? Tick all that apply.
33	How would you like to learn about AI? Please tick all that apply.

Questions were answered using free text, and both 5-point and 6-point Likert scales [27,28] were used. A 6-point Likert scale was used to allow participants to mark their agreement or disagreement with a particular statement on a symmetrical scale thus removing the neutral option. This allows negative and positive responses to be emphasized and according to [28] gives discrimination and reliability values which are higher than a 5-point Likert scale.

To explore a deeper understanding with participants of practical AI applications, a series of questions focused on two mini case studies. Case study 1 required participants to read the following example and answer a series of questions. The aim of the case study was to introduce citizens to a typical online automated decision-making system, describing a model which had been generated by an underrepresented data set in terms of gender and investigate their responses to questions related to bias, fairness and trust in relation to age, gender and educational level. Case study 1 and the associated questions are as follows:

An online company offers quick loans from £500 to £5000. To apply you must fill in the quick loan eligibility checker with your data. The AI application will make a decision on whether you are eligible by using your data to predict if you will pay back your loan on time. You are asked to provide personal data such as your gender and age. The AI application has been trained on data that contains 20% females and 70% males and 10% other. 85% of people in the training data are also aged between 18 and 25.

- Q22. Do you think the decision made by the AI system on wether to give you a loan is biased? Please explain your answer
- Q23. Do you think the decision made by the AI system on whether to give you a loan is fair? Please explain your answer
- Q24. To what extent would you trust a decision by the system for you? Please explain your answer
- Q25. What would you do if your application was rejected by the system?

The second mini case study asked participants to first read the following scenario: *You are travelling for a holiday and reach a border crossing point where a border guard asks you some questions about your trip. While you are answering the questions, you are being filmed and an automated detection system is looking at your face and seeing if you are telling the truth or not. The border guard can see on their screen the results in front of them and can use the information to help them decide whether to let you pass through or ask you more questions.*

Participants were asked to then answer two questions:

- Q27. On a percent scale of 0 to 100, how accurate do you think the deception detection system should be before it is used to provide information to the border guard?
- Q28.When do you think it is ethical to use the deception detection to provide information to the border guard? Tick all that apply

The aim of the second case study was to first examine what citizens thought would be a suitable accuracy level for a high-risk AI system to achieve before it should be used to provide information to a human – in this case the border guard. Secondly, to find out if they thought this system was ethical. To conduct this study, a full ethical application was submitted and approved by Manchester Metropolitan University (ETHOS REF: 27706).

B) Participants

Given the nature of COVID and associated restrictions, the survey was advertised mainly online through professional and social media channels, and through an email to one community leader. To ensure inclusivity, a paper-based format of the survey was made available on request and a person could also request to participate via a phone conversation where one member of the research team verbally asked the questions and recorded answers directly into the online survey. Invitations to participate were also emailed to potential candidate organizations and individuals. All participants were provided with a participant information sheet prior to taking part and had the opportunity to ask questions. An e-consent form was then signed prior to starting the survey. A participant could withdraw at any time without giving reason. The inclusion criteria for participants were that they gave explicit consent to participate in the survey after reading the participant information provided at the start of the online survey and were consenting adults over the age of 18.

IV. Results and Discussion

Forty responses were collected between 18 May 2021 and 1 August 2021. 40% were female, 55% were male and 5% preferred not to say. 72% of participants were from the UK and the remainder from India, Germany, Spain, Brazil, Jordan and Canada. Figures 1 and 2 show the distribution of age and education attainment of participants.

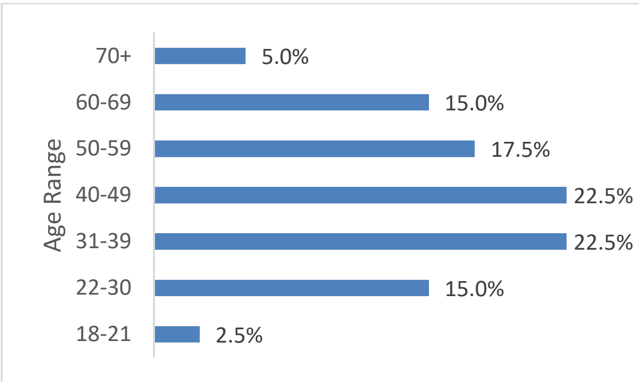

Fig. 1. Age ranges of participants

55% of participants were employed full time, 17% part-time, 20% were retired and 8% were seeking opportunities. Questions were analyzed according to three themes: a) citizen understanding of AI and ethics; b) trust and confidence; and c) case study analysis.

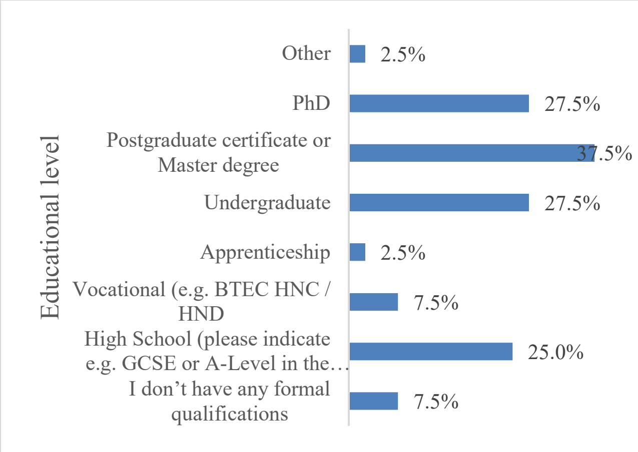

Fig.3. Educational Attainment of participants

A) Citizen understanding of AI and ethics

At the start of the survey, participants were asked to write what they understood by the term "artificial intelligence" (Q8) to assess their initial understanding. Answers were varied including "*I think it's to do with your brain and what you know*" (age range 60 – 69 and no formal qualifications), "*to do with the brain how intelligent you are*" (age range retired, high school qualifications), "*A set of techniques that make computers do things better than humans*" (age range 40 – 49, PhD qualification). Participants were then provided with a definition of AI - adapted from a school curriculum suitable for 11–16-year-olds – and asked if it was clear and whether they understood it (Table II). 18% of participants said the definition was not clear, of these 16% were over 60 and had no formal or only high school qualifications and female. All participant responses to Q10 (*What does "being ethical" mean to you in your everyday life?)* revealed a good understanding of ethics, with common definitions focusing on the moral differences between right and wrong, upholding human rights, acting with integrity and being honest. One participant wrote "*Having traits such as integrity and honesty. It means trying to do the right thing, living one's values, and showing concern for others and for society*."

B) Trust and Confidence

Q14 captured the participants' feelings about trust in the use of general AI applications, whilst Q15 focused on the perception of trust if a specific AI application was being used on them, their family, and friends. The aim of this question was to investigate if citizens have increased or decreased trust in AI applications when they are being used or applied to themselves, family or friends compared with the general population. Responses were also analyzed for differences between gender, age and educational level. To analyze the statistical significance of the results from the Likert scale questions (Q14-15), the nonparametric Man Whitney [27] test was used as the normality assumption does not hold and the sample size is small. Table III shows the median values (M_{Q14} and M_{Q15}) for the two groups along with the *p-value* derived from the Mann-Whitney statistical test, across the whole dataset for each statement {*a..m*} in Table III. The column *No' diff* reports the number of participants who changed their opinion on a given statement i.e., Q14.a and Q15.a had a different response.

TABLE III. MEDIANS AND P-VALUES FOR Q14 AND Q15

Area of Use	M_{Q14}	M_{Q15}	p-value	No' diff
a. Using AI to help support education / training in schools	3	3	0.410	5
b. Using AI to help predict the future spread of viruses such as COVID-19	4	4	0.916	12
c. Use of Health Care robots to support nurses	2	2.5	0.709	14
d. The use of face recognition cameras in public spaces to identify potential suspects	3	3	0.707	14
e. The use of automated deception detection system to support police officer interviews	2	2	0.439	11
f. Improvements in home energy efficiency to save you money	4	4	0.764	12
g. Identification of plastics in the sea to prioritize an area for clean-up activities.	4	4	0.975	5
h. Using a self-driving car to go to work	2.5	2	0.811	10
i. The use of drones to identify crop disease early on so it can j. be treated.	4	4	0.975	9
j. Automated robots in manufacturing to do repetitive jobs	4	4.5	0.381	6
l. The use of AI to help border guards make decisions on whether to ask a person more questions at borders.	2	2.5	0.584	17
m. The use of AI systems to diagnose the type of tumors in the human body by looking at images.	4	4	0.720	13

The median values of results (M_{Q14} and M_{Q15}) were identical across all applications in all except four cases. *The p-value > 0.05* in all cases indicated that the different between the medians was not statistically significant. Trust in a given AI application by a participant is very similar regardless of whether it is applied to the general population or to a participant's own family and friends. However, examination of the data on a statement-by-statement basis showed that each participant had several personal preference differences. For example, given the statement "*The use of AI systems to diagnose the type of tumors in the human body by looking at images*", nine participants increased their trust score when the system was to be used on themselves, their family or friends and four reduced their trust levels thus implying they were happy with the use of AI in general in this scenario but not when applied to their own family. All statements with had some differences of opinion. Comparing the male and female responses for each question statement, the *p-value* was always greater than 0.05 indicating that the median responses between males and females was different (in 71% of cases) but not statistically significant. Those with a PhD qualification varied their answers the most between the same statement on Q14 and Q15. For the statement "*Use of health care robots to support nurses*", PhD students gave a median answer of 3 for Q14 and 2 for Q15, indicating that they had greater general trust for this application (Q15), but if being used on themselves, their family, and friends the level of trust was reduced. A similar pattern emerged across all educational levels and age ranges where there was a difference in participant answers in terms of trust, when the application was applied to the participant (and family) themselves trust was lower (Q15) than in response to trust in general (Q14). The age range 60-69 showed the greatest number of differences between answers given to Q14 and Q15 on the same statements.

In Q18 participants were first told that an AI application only makes decisions based on what it has learnt from the data it has been given. Participants were then asked about who they thought should be held responsible when things go wrong. Participants were given 8 statements and asked their opinion if they thought that person or entity was always, somewhat, or not responsible (Table IV). This question was designed to capture opinions about who was responsible. Examples were given to contextualize the question: e.g., a person being mis-diagnosed, or a person is rejected for a loan.

TABLE IV Q18 – WHO IS RESPONSIBLE WHEN THINGS GO WRONG?

No	Statement	Always	Somewhat	Not
18.1	The company who purchased and uses the AI application is responsible when things go wrong	50%	42.5%	7.5%
18.2	The human operator who is using the results of the AI to support their decision making	20%	72.5%`	7.5%
18.3	The people who collected the data as it was poor quality	35%	55%	10%
18.4	The software developers who coded the AI algorithms to learn from the data	30%	65%	5%
18.5	The people who tested the AI application	27.5%	65%	7.5%
18.6	Members of the public for not voicing their concern	10%	32.5%	57.5%
18.7	The Law	35%	42.5%	22.5%
18.8	The user of the system as they did not understand the decision correctly	20%	55%	25%

For all statements except 18.6 and 18.7, 10% or fewer respondents thought that the person or entity in question had no level of responsibility for the error. For statement 18.1 (Table IV), those that answered that the company was not responsible had either a UG or master's qualification and were less than 49 years old. 57.5% of participants thought the public were not responsible when things go wrong, and they didn't voice concern (18.6) while 80% thought the law was either somewhat or not at all responsible. The results indicate that citizens recognize that there are multiple actors and entities that may be held responsible when things go wrong, but the degree of responsibility may depend on understanding and personal experiences. The results overall show that opinion greatly differs. There are no clear right/wrong answers, indicating the complexity of ethics and apportioning moral or legal accountability. There is strong support for the system development companies to be held accountable for the outcomes of their systems' decisions. In addition, individual employees shouldn't be held fully and personally responsible for the faults and biases within the system as they are following procedures and making the best decisions on the basis of what is before them, but they are partially responsible.

C) Case Study 1 analysis

Analysis was conducted on Q22 – 25 to examine if there was a difference of opinion based on age range, educational level and gender (*RQ2*). Figures 4 to 6 show comparisons of participant answers for questions 22-24 analyzed by gender, educational levels and age. Participants were first asked to think about whether the decision made by the AI system about a loan application was biased (Q22). 70% of participants thought the decision would be biased and of these 50% explained it was because of the training data. 23% were unsure and 1% thought the decision would not be biased. Two respondents said that the decisions were not biased: "*Not at all cause the system uses simple rules as generated by the company and only the company review or change cause of no human factor involved decision can never be biased*" and "*No. Every applicant is treated equally in the eyes of the AI*".

Figure 4 presents a bar chart where a 'yes' or 'no' response indicated that the participant thought the system was or was not either biased or fair. A "?" response indicated that they were unsure, where PNS is Preferred Not to say. The results indicate that females are less likely to think AI is fair and do not trust AI. Males were more concerned about bias than females and more likely to be unsure about trust. Due to the small sample size, results are not statistically significant.

The bar chart in Figure 5 shows three possible responses (yes, no, ?-unsure) for (Q22, Q23, Q24) in relation to the participant's educational level. What is evident is that all those with a PhD qualification, regardless of subject, agreed that the loan system was biased, not fair and they did not trust it. 31% of participants with a master's qualification, thought the system was not biased, compared with 25% who did, however 31% said it was not fair. For those with no qualifications or high school qualifications, 42% thought the system was biased and 32% said they were uncertain, whereas for fairness, 51% thought the system fair and 25% were uncertain. Across this group 13% trusted the system, 31% did not and 19% were unsure. Interestingly, 38% of master's students were also uncertain about trusting the system.

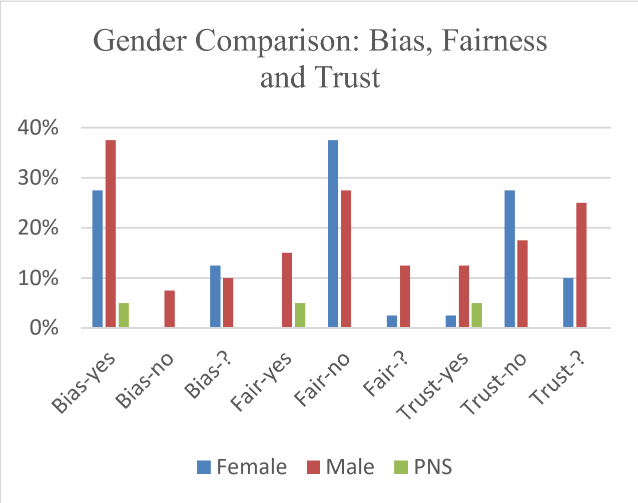

Fig.4. Case Study 1 Gender Comparison (Q22, Q23, Q24)

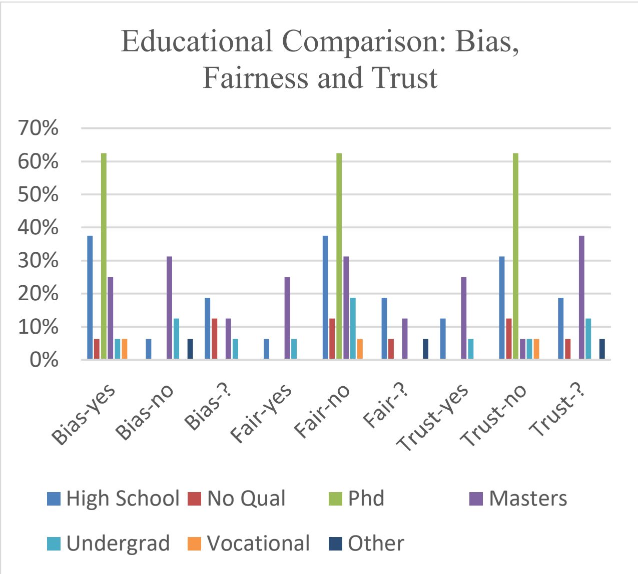

Fig.5. Educational Levels for Bias (Q22), Fairness (Q23) and Trust (Q24)

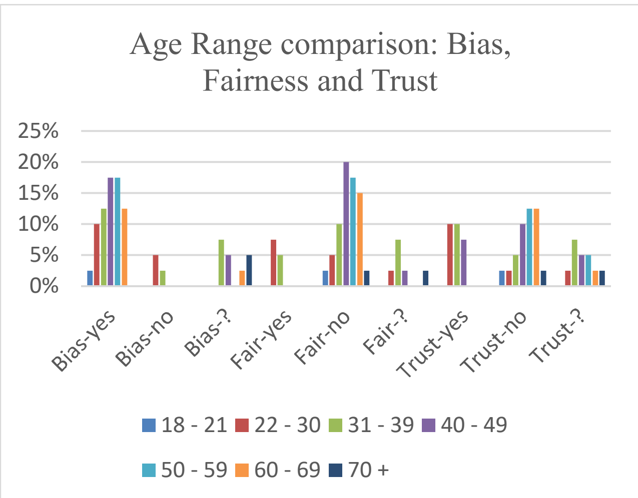

Fig.6. Case Study 1 Age Comparison (Q22, Q23, Q24)

Figure 6 presents a bar chart showing the three possible responses for (Q22, Q23, Q24) in relation to a participant's age. Across all age ranges the majority thought the system was biased. 8% aged over 60, and 13% between 31 and 49 were unsure. Analyzing age independently, those aged between 22 and 39 tended to trust the system more compared with those aged 40 and above who did not trust the system or were unsure.

The initial responses to the survey have demonstrated that overall that many participants are critically thinking about bias (most answering yes), fairness (most answering no), and are a little trusting (fewer participants answering yes). A much larger proportion of people were unsure about trust, suggesting that trust could be won.

Q25 asked participants to explain what they would do if their loan application was rejected by the system. The answers were coded in to five key themes:

1. 23% of participants said they would *seek contact with a human*. Answers included "Contact customer services", "Appeal to a human".
2. *Seek an explanation of why the decision was made.* One participant noted "I would ask a detailed written description of how the rejection was decided." and another said "have (the decision) qualified by a human".
3. *Feel angry and upset.* Two participants gave examples of when a similar situation happen to them "I would be very upset and want to know why. The problem is there is never anyone to call, and companies hide behind email" and "When this happened to me, I tried to email the company through the contact us page, but they never got back to me. In the case of the loan, I need to know why I was rejected - for example what happens if it is an error in my credit rating?"
4. *Complain to a regulatory body*
5. *Nothing / Walk Away.* One participant clarified their response by saying "I wouldn't use it in the first place (at least not if I knew the facts above)" – referring to the data description in the case study

D) Case Study 2 Analysis

Q27 asked "*On a percent scale of 0 to 100, how accurate do you think the deception detection system should be before it is used to provide information to the border guard*?" There was no difference between participant responses in terms of age range. 28% gave an answer of 100% accuracy, 60% said 90% or above and 13% stated 50% and below. In terms of gender, 33% of females, 23% of males, and 1% others, gave an accuracy rating of 90% or above. Analysis in terms of educational levels revealed that 80% of participants with a high school qualification and 80% with a Ph.D. gave an Accuracy of 90% and above. In contrast, for those with a masters this was only 33% and could indicate outliers. Whilst an interesting finding, more data would be required to determine if this was significant. Figure 7 shows the participant responses to Q28 and their viewpoints about when they consider it acceptable to use the deception detection system to provide information to a border guard. Participants could tick more than one answer. 67.5% considered that system accuracy of 100% would make the system ethical to use, with 75% agreeing that the border guard must first be trained on how to use the information provided by the system. 40% also highlighted that they would find it acceptable to be used if it was known that the border guard had unconscious bias.

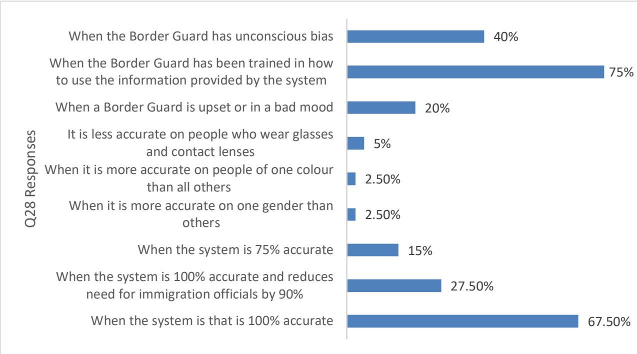

Fig.7. When do you think it is ethical to use the deception detection to provide information to the border guard?

E) Education and training

Q32 asked participants to say what they would like to learn about if they were offered free courses about AI. Figure 8 shows a summary of responses.

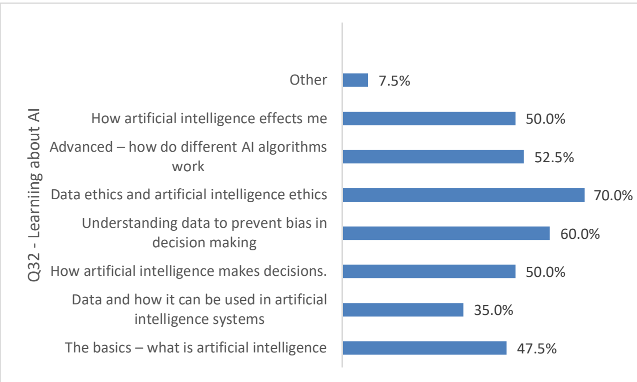

Fig.8. AI Content Preference

100% of those aged between 50-59 wanted to learn about the basics, with 71% of this group wanting to know how AI made a decision. Those over 60, also wanted to learn about data and AI ethics and how AI affected them personally. 46% aged 49 or under were interested in understanding how the AI algorithms worked. There was no notable difference between gender and the selection of answers. 6% of those with a high school or no qualification wanted to learn about the basics of AI, and so did 50% of those with a Ph.D. 80% of Ph.D. students also wanted to understand data to prevent bias in decision making. People with high education attainment have more interest in learning, whereas the challenge will be that people with lower educational attainment may have less interest in learning and this is a serious barrier to overcome. Given the small sample size, the responses to this question, are likely reliant on the subject(s) or employment that participants felt most competent in. Under "Other", one participant wrote "*Using data and AI to deliver equity*". Participants were then asked how they would like to learn about AI (Q33). 65% were in favor of introductory courses that were accessible for everyone covering topics: what is AI, what is ethics and what is data governance which could be delivered both face-to-face and online. Work-based introductory continuous professional development (CPD) modules were suggested (47.5%). 17.5% stated that they would like to learn about AI through a Board game and 25% through playing a mobile app. 25% thought family workshops would be a good idea and 27.5% thought booklets would also support learning. 75% of females selected some form of introductory course, compared with 81% of males.

V. Conclusions and Further Work

It is noted that the preliminary results reported in this paper are from a small sample, which is why the survey has been established as a longitude study, with quarterly analysis points. Another caveat is that the education curriculum has evolved along with the teaching of computer science and AI over the years and those in high age ranges may have had little exposure before completing this questionnaire. Nevertheless, this also provides a justification, that if we expect our growing elderly population to engage with such systems online (especially in the increased automation of services during COVID and reduction in face-to-face services, e.g. the banking sector) then we need to provide a way to educate and train this population group to be more confident and digitally literate and empower them to ask questions and know their rights. To answer *RQ1* and *RQ2*, results have shown that perception of AI, ethics and trust, biasness and fairness is affected by gender, age and educational level and related to the specific AI application and the context in which is applied. For example, participants aged 22 and 39 tended to trust the specific applications of AI more, compared with those aged 40 and above. Males were more concerned about bias than females and more likely to be unsure about trust. When an AI system fails to give a correct decision response, 57.5% of participants thought the public were not responsible when things go wrong. Results indicate that participants thought responsibility should be shared amongst a range of stakeholders. Participants with a PhD qualification agreed that the loan system was biased, not fair and they did not trust it. For those with no qualifications or high school qualifications, 42% thought the system was biased and 32% said they were uncertain, whereas for fairness, 51% thought the system fair and 25% were uncertain.

Findings for RQ3 revealed that those aged 50 and over were interested in introductory courses on AI, and how AI affected them personally. 85% participants across all age levels selected at least 2 courses. The basics of AI was also selected by 50% of those with a PhD. This supports the need for such a course to be developed, which has consistent learning outcomes, but with multiple modes of delivery i.e. family workshops versus a CPD workshop to ensure a universal understanding of AI concepts of bias, fairness etc. This could be achieved through a toolkit of alternative resources, mapped to each learning outcome, which can be adapted to different groups, who have different concerns about the use of AI. Future work will seek to evaluate and extract best practice from existing free AI pubic courses and co-produce/evaluate with diverse citizens inclusive courses to empower them to be more confident in their use and scrutiny of AI. If you would like to take part in the longitude survey, and contribute your opinion, please go here https://mmu.onlinesurveys.ac.uk/public-opinions-on-artificial-intelligence for more information and to take part.

REFERENCES

[1] European Union, "Regulation Of The European Parliament And Of The Council, Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206, 2021

[2] Guidance for Regulation of Artificial Intelligence Applications, [Online], Available: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf, [Accessed 11 Aug.2021], 2021

[3] BEIS, The use of public engagement for technological innovation Literature review and case studies, BEIS Research Paper Number 2021/003, [Online], Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/955880/use-of-public-engagement-for-technological-innovation.pdf [Accessed 11 Aug.2021], 2021

[4] World Economic Forum, Civil society can help ensure AI benefits us all. Here's how, [Online], Available: https://www.weforum.org/agenda/2021/07/civil-society-help-ai-benefits/, [Accessed 11 Aug.2021], 2021

[5] Alexandre, F., Becker, J., Comte, M.H., Lagarrigue, A., Liblau, R., Romero, M. and Viéville, T., Why, What and How to help each Citizen to Understand Artificial Intelligence? KI-Künstliche Intelligenz, pp.1-9, 2021

[6] Jones, E. Safak, C., Can algorithms ever make the grade? Ada Lovelace Institute [Online], Available: https://www.adalovelaceinstitute.org/blog/can-algorithms-ever-make-the-grade/, [Accessed 11 Aug.2021]., 2020

[7] Guinchard, A., Our digital footprint under Covid-19: should we fear the UK digital contact tracing app? International Review of Law, Computers & Technology, 35(1), pp.84-97, 2021

[8] UK AI Council Roadmap, [Online], Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949539/AI_Council_AI_Roadmap.pdf, [Accessed 11 Aug.2021], 2021

[9] Robinson, S.C., Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). Technology in Society, 63, p.101421, 2020

[10] Artificial Intelligence: public awareness survey, [Online], Available https://www.gov.uk/government/publications/artificial-intelligence-public-awareness-survey, [Accessed 11 Aug.2021], 2019

[11] Neudert, L.M., Knuutila, A. and Howard, P., [Online] Global Attitudes towards AI, Machine Learning & Automated Decision Making, Available: https://oxcaigg.oii.ox.ac.uk/wp-content/uploads/sites/124/2020/10/GlobalAttitudesTowardsAIMachineLearning2020.pdf, [Accessed 11 Aug.2021], 2020

[12] Edelman's 2019 Artificial Intelligence (AI) Survey, [Online], Available: https://www.edelman.com/research/2019-artificial-intelligence-survey, [Accessed 11 Aug.2021], 2019

[13] University of Helsinki, Ethics of AI, [Online], Available: https://ethics-of-ai.mooc.fi/start, [Accessed 11 Aug.2021], 2021

[14] University of Helsinki, Elements of AI, [Online], Available: https://www.elementsofai.com/, [Accessed 11 Aug.2021], 2018

[15] We Are AI, [Online], Available: https://dataresponsibly.github.io/we-are-ai/, [Accessed 11 Aug.2021], 2021

[16] Fast.AI, Practical Data Ethics, [online], Available: https://ethics.fast.ai/, [Accessed 11 Aug.2021], 2020

[17] Fiesler, C., Garrett, N. and Beard, N., What do we teach when we teach tech ethics? a syllabi analysis. In Proceedings of the 51st ACM Technical Symposium on Computer Science Education (pp. 289-295), 2020

[18] Fiesler, C. Tech Ethics Curriculum, [online], Available: https://docs.google.com/spreadsheets/d/1jWIrA8jHz5fYAW4h9CkUD8gKS5V98PDJDymRf8d9vKI/edit#gid=0, [Accessed 11 Aug.2021], 2021

[19] Schiff, J. Borenstein, J. Biddle and K. Laas, AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection, IEEE Transactions on Technology and Society, vol. 2, no. 1, pp. 31-42, 2021

[20] Buhmann, A. and Fieseler, C., Towards a deliberative framework for responsible innovation in artificial intelligence. Technology in Society, 64, p.101475, 2021

[21] Buhmann, A. Pamann, J. Fieseler, C. Managing algorithmic accountability: balancing reputational concerns, engagement strategies, and the potential of rational discourse J. Bus. Ethics (163), pp. 265-280, 2020

[22] Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. Chatila, R., Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, pp.82-115, 2020

[23] Akova, B Yang, J. Cramer, H. Chowdhury. R. Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. Proc. ACM Human Computer Interaction 5, CSCW1, Art. 7 DOI:https://doi.org/10.1145/3449081, 2021

[24] Mitchell,M. Wu, S. Zaldivar, A. Barnes, P. Vasserman, L. Hutchinson, B. Spitzer, E. Raji, I.D, Gebru, T, Model cards for model reporting. In Proceedings of the conference on fairness, accountability, and transparency, pp. 220-229, 2019

[25] Crockett, K. Adaptive Psychological Profiling from Non-Verbal Behaviour – "Why are Ethics Just Not Enough to Build Trust?", Women in Computational Intelligence, Eds. Smith, A. Springer, *in press, 2021*

[26] Crockett, K., Garratt, M., Latham, A., Colyer, E. and Goltz, S. July. Risk and Trust Perceptions of the Public of Artificial Intelligence Applications. In 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 1-8, 2020

[27] Likert R. A technique for the measurement of attitudes. Arch Psychology. 1932;22 (140):55.

[28] Chomeya, R., Quality of psychology test between Likert scale 5 and 6 points. Journal of Social Sciences, 6(3), pp.399-403, 2010

[29] de Winter, J. F.C. and Dodou, D.Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon, Practical Assessment, Research, and Evaluation: Vol.15, Art.11, 2010