Towards Trustworthy AI: Raising awareness in marginalized communities

Annabel Latham
Department of Computing and Mathematics
Manchester Metropolitan University
Manchester, UK
ORCID: 0000-0002-8410-7950

Keeley Crockett
Department of Computing and Mathematics
Manchester Metropolitan University
Manchester, UK
ORCID: 0000-0003-1941-6201

***Abstract*— Citizen trust is a central part of an ethical AI ecosystem, and a key part of upcoming AI legislation across the world. However, pockets of society have little awareness of what is meant by AI or its current use by organizations. Therefore, a first step towards building citizen trust is to raise general awareness of what is meant by AI technologies, how data is captured and used in decision making, and existing citizens' rights to question organizations about the use of their data in automated decision making. This paper describes a mechanism to reach different groups of publics through a Community AI Roadshow. The motivation was to develop a way to reach and engage with traditionally marginalized communities and develop a common language and understanding around AI. An evaluation showed that understanding of AI increased by 33% following the roadshow. A resulting set of recommendations for AI researchers engaging with marginalized communities is given. The methodology presented in this paper has since been adapted for other groups of publics, such as local governmental organizations in the UK.**

Keywords*— *Trustworthy AI, Citizen engagement, Marginalized communities

I. INTRODUCTION

Advancing technology for the benefit of humanity is the IEEE's core purpose, and the motivation for the work described in this paper. Balancing the power of AI to improve people's lives against the risks of unethical practices is a challenge facing legislators the world over. Public trust in AI as a benefit will determine whether it is possible to achieve a healthy AI ecosystem where all citizens know that they can question the use of data and AI, and know how to do this. Successful legislation and regulation of AI technologies relies on a general understanding of citizen rights around challenging the application of AI and data-driven technologies. Raising awareness of data and AI is the focus of this research, and in particular reaching out to communities who are often missed or excluded, such as traditionally marginalized communities.

There is no agreed definition of marginality [1], but in this paper marginalized communities refers to those communities who are minoritized and underserved by larger society. Rather than solving societal injustices, new technologies such as AI can actually perpetuate inequalities and marginalization [2]. Marginalized communities are often overlooked by organisations and regulators in the AI space, and traditionally public engagement has focused on one-way information sharing. Even researchers taking steps to avoid unethical practices, such as the adoption of participatory or human-centred design, should be aware that this does not guarantee equity [3]. True engagement with marginalized communities requires researchers to reflect on several tensions and challenges faced by the groups with whom they wish to engage. Digital exclusion is a particular challenge that affects publicity, recruitment, and engagement with citizens. Narrative and storytelling approaches are an effective way to explore and debate technology, however different stories will speak to different participants depending on their lived experience. There are currently no mechanisms to support organisations and researchers who want to engage with marginalized communities as part of their AI R&D process.

This paper describes a pilot study based in Greater Manchester in the UK, where there are over 1 million digitally excluded residents. Two different communities who are in the lowest 1-3% nationally (UK) for indices of multiple deprivation were selected for the pilot study. A narrative, game-based adaptive Community AI Roadshow was designed to demystify AI, data and bias, with activities exploring the ethical issues around the use of automated decision making. Three AI Roadshows were delivered in central community centres and evaluated in terms of reach, effectiveness and impact on understanding of AI. During each roadshow, stories of AI applications highlighted in the media were selected dynamically according to the interests and lived experience of the participants, observed during discussion and debate. The design of the roadshow evolved based on interactions with different marginalized groups, as a basis for engaging with new communities to raise awareness and understanding of AI.

The research questions (RQ) for this study are:

- RQ1 Is it possible to design a short interactive workshop to explain and raise awareness of AI and data-driven technologies, related ethical issues and citizen rights, for groups with a mixed level of background education and experience?
- RQ2: Does raising awareness of AI and data-driven technologies in a community setting increase citizen understanding of AI and confidence to challenge organisations about automated decisions?
- RQ3: Can raising awareness of ethical issues and citizen rights around AI in a community setting empower citizens to learn more and engage with AI companies?

The main contributions in this paper are:

This work was supported by The Alan Turing Institute Public Engagement Grant 2022.

- A methodology for raising awareness of AI, related ethical issues and citizen rights in traditionally marginalized communities.
- Development and trial of an adaptable community AI workshop that uses storytelling to engage marginalized communities and raise awareness and understanding of AI, ethical issues and citizen rights.
- Evaluation of a pilot study in hard-to-reach communities within Greater Manchester, UK, demonstrating increased citizen confidence and understanding of AI.
- A set of recommendations for researchers designing public engagement activities within marginalized communities.

The rest of this paper is organised as follows: section II gives background on citizen trust in AI, the particular considerations for researchers when working with marginalized groups and the People's Panel for AI; section III describes the Methodology for developing Community AI Roadshows; section IV presents the results and discussion, including recommendations and section V concludes and suggests future work.

II. Background

A. Citizen trust in AI

Democratisation of AI is fundamental in building trust with people to ensure inclusivity and equality in its application. For people who are marginalized, it may require a leap of faith [4] to use AI technology, in that they either lack the knowledge and skills to make an informed choice, or they are presented with no alternative – use the tech or be denied the service. There have been several recent studies and surveys that have tried to capture public perceptions about AI [5..8]. A survey of more than 30,000 individuals across the EU in 2019 [5] showed that men were found to be more trusting than women and the richest people had the least trust in AI. The survey found that how AI was represented also played an important part in establishing a trust relationship with robotic pets facilitating initial trust formation. In [8] a US centric study (n=525) was conducted to investigate what influence ethical requirements of AI have on trust. The study investigated the seven ethical principles (Human agency and oversight; Technical robustness and safety; Privacy and data governance; Transparency; Diversity, non-discrimination, and fairness; Societal and environmental well-being and Accountability) proposed by the European Commission High-Level Expert Group. The survey found a positive correlation between user experience of smart home products, consumer tech and chatbots in relation to trust, however acknowledges that the sample is not truly representative, especially as all participants had internet access. A 2023, MITRE-Harris Poll (n=2050) [9] found a lack of trust in American citizens, especially in high stakes applications such as autonomous vehicles and healthcare. In this study, only 48% believe the use of AI was safe, 82% of Americans and 91% of working in tech industry supported regulation. The poll was accessible to those with internet.

In 2023, Ada undertook a UK survey of 4000 adults to examine public attitudes to artificial intelligence [7]. Whilst the majority of participants took the survey online, 252 participants were interviewed by telephone either because they had no internet access or because this was their preference. The survey found that participants felt that those developing AI technologies needed to be more transparent about their role and approach in order to be trusted. The findings supported previous work [10] that concluded that raising awareness of AI technology alone will not necessarily increase public trust in specific applications.

It is clear from this very high-level analysis of surveys designed to question the public about their opinions, perceptions and trust in AI that marginalized unrepresented people are often excluded from taking part. Access to surveys is often limited to those which have privilege of internet access, sample sizes are typically small and not representative, and often there is a lack of information about reach across different demographic groups. This raises the pertinent question; how do we really know how to build citizen trust if we do not first meet people where they are in terms of understanding and lived experiences in using AI?

B. Considerations when engaging with traditionally marginalized communities

"At a high level marginalisation refers to how a person experiences the world around them based on their identity and how others perceive them" [1]. There are different contexts in which people may experience marginalization or privilege, and these experiences change over time. Intersectionality refers to how people face additional dimensions of oppression due to experiencing marginalization in multiple contexts [11]. It is important for researchers to keep in mind that there is no one-size-fits-all approach for engaging with citizens, whether they experience marginalization or not. The careful formation and development of relationships with participants and their support networks over the long-term enables researchers to understand and adapt to the complexity of citizens lives [12].

In [1] four tensions are described that must be considered by researchers when planning to engage with marginalized communities in order to better understand the situational needs of the project, participants and other stakeholders. Whilst the four tensions (exploitation, membership, disclosure, and allyship) explore the complexity and difficulty when engaging with marginalized communities, the authors stress that researchers should recognise their power that can be used for the promotion of others.

One challenge for public engagement around new technologies such as AI is digital exclusion, which refers to a lack of access to or use of the internet and is understood to be closely linked to poverty [13]. Digital exclusion may be due to digital literacy (a lack of skills to use technologies via the internet) or digital poverty (being fully offline due to a lack of access to devices that connect to the internet). Although in the UK the proportion of adults who are digitally excluded has fallen, 10% of the UK population were classed as digitally excluded in 2018 [14] and in Greater Manchester 1.2 million people are digitally excluded, with 23% of residents not using digital services due to lack of money [15]. The prevalence of digital exclusion in some marginalized communities excludes those citizens from access to online support, learning and public engagement activities as well as many services.

Storytelling and narratives play an important role in public engagement as they are effective in communicating complex ideas and providing a safe space for meaningful discussion with a wide range of audiences [16]. Research has shown that

narratives are easy for people to process as stories are introduced early in life and enable rich mental representations of the substance of a topic [17]. When working with groups of people, a narrative can be successful in establishing common ground around a shared story that enables discussion around an issue, such as the use of automated decisions [18].

C. The People's Panel for AI

The People's Panel for AI (PPfAI) is a framework that aims to engage citizens in the research and development of AI products and services, thus building public trust in AI [19]. The framework consists of four phases, as follows:

1) Community AI Roadshows: The subject of this paper. The aim of the Community AI Roadshows is to raise awareness in the community of how AI works and the ethical concerns around AI and automated decision making. The roadshows are also used as a device to motivate citizens to sign up for further training and to be part of the People's Panel for AI.

2) People's Panel Training Days: Citizens who volunteer to be part of the PPfAI attend two interactive training days that aim to build confidence, group cohesion and prepare panelists for the PPfAI panel sessions. The training takes a deeper dive into the use of data and AI, introduces tools for evaluating harms and consequences of AI, and prepares citizens for the panel sessions by holding a mock panel.

3) PPfAI Panel Sessions: PPfAI panels are facilitated formal sessions that aim to bring together citizens and researchers or businesses who are developing AI technologies. The presenters (researchers or businesses) explain their AI product or service to the PPfAI panelists via a short non-technical presentation and by answering pannellists questions. The panel then convenes privately and conducts Consequence Scanning [20], followed by a feedback and discussion session with the presenters. The Facilitator documents the process and sends written feedback to the researchers.

4) PPfAI Evaluation and Feedback: All phases of the PPfAI framework are evaluated, and the results feed into the co-produced PPfAI Terms of Reference document [21]. Reflection on the process is collected from all stakeholders to inform the evolution of the PPfAI. The PPfAI panel facilitator closes the loop by sharing further questions and comments from the panelists with the researcher/business, and sharing feedback and actions from the researcher/business with panelists. Events are held in the communities to allow the panelists to share their experience of the PPfAI.

III. Raising Awareness through AI Community Roadshows

The purpose of this research was to engage with people within traditionally marginalized communities on their thoughts, opinions and understanding of AI and why it might matter to them. The challenge is that there is no shared understanding or language around AI and related ethical issues. This is due to multiple factors, such as lack of awareness in communities who are marginalized or digitally excluded, negative views and fear as a result of media hype and the often-biased presentation of AI technologies in the media. The development of AI community roadshows was the mechanism to introduce people to how AI works and the ethical challenges it poses using storytelling and debate.

The aim of the roadshow is to introduce people to how AI works, the ethics, risks and benefits of AI systems and UK citizens' rights to question organisations about their use of AI.

A. Methodology

To reduce possible barriers and attract a larger number of community members, it was decided to develop a free, short, interactive "eat and learn" AI roadshow that would be held in a central community venue. Each roadshow lasted two hours and included food and drinks.

1) Identify target communities: The aim of this project was to reach underserved communities in Greater Manchester, which has several areas of deprivation. The two target communities, Ordsall and Levenshulme in Greater Manchester, UK, are in the lowest 1-3% nationally for indices of multiple deprivation (i.e. the most deprived communities). Levenshulme is approximately four miles south of Manchester city centre and is predominantly residential, with many public houses, fast food shops and antique shops. Levenshulme has a multi-cultural and multi-ethnic population of 19,647 at the 2021 census [22]. Ordsall is south-east of Salford city centre and the southern part is almost entirely residential, with a high proportion of social rented housing [23]. Ordsall has a population of 10,481, an increase of 53% between 2011 and 2021 [23], and despite a recent urban regeneration project remains one of the most deprived parts of Greater Manchester with higher than average crime rate. Neither community had previously engaged with the R&D sector.

2) Arrange community venue and catering: Each target community had a central community social space with good reach into the community: The Tatton Café and Community Hub, Ordsall [24] and The Levenshulme Inspire Centre Community Hub [25]. The team sought advice from contacts working in the community centres, who advised on the most suitable day and time to attract people from the community and also the type of meal that would be most appreciated. This was important to maximise the number of citizens reached and also to remove potential barriers by selecting a time and space and food that community members were familiar with. The arrangements were different for each community group – at The Tatton an evening AI roadshow and hot meal, held in the café, was best for the community. At Inspire, two AI roadshows were organised – the first was open to all and took place in a function room during the afternoon and included afternoon tea and cake. A second AI roadshow at Inspire was organised for members of an existing community group for older people. As the community group took place in a function room two afternoons each week, a hot lunch and an afternoon AI roadshow was integrated into one of the sessions.

3) Advertise AI roadshows in the community: Following the advice of contacts working in the community centres, publicity for the AI roadshows was based on offline methods for better inclusivity and reach into the community. Colour posters and flyers were printed and displayed in the community centre cafes, local libraries and community

newsletters. Interested people were encouraged to sign up for the Meal and AI Roadshows either digitally by scanning the QR code on the posters on their mobile phones to regsiter via Eventbrite, or by writing their name on a list held at the community centre reception. Regular communication with community contacts allowed the team to combine online and offline registration details for catering purposes.

4) Design accessible, interactive AI roadshows: A narrative, game-based adaptive Community AI Roadshow was designed to introduce and explore AI, data and bias, with activities exploring the ethical issues around use of automated decision making. It was important to keep the duration of the roadshow short, a maximum of 45 minutes, to allow plenty of time for questions and discussions over the shared meal. Designing for inclusivity meant being aware that no knowledge could be assumed, so concepts were introduced and discussed in simple terms. The background and interest of the attendees was difficult to predict as the roadshows were open to anyone in the community and there was no prior knowledge of people who had registered. Therefore a number of games and scenarios were designed (more than there was time for) so that stories could be selected dynamically during the session depending on the response of the attendees. For accessibility the material used, such as presentation and evaluation forms, was printed, with copies also in large font. Activites did not require access to the internet and could be done using pen and paper. Simple, coloured voting cards with one side red with a tomato and the other green with a pepper (as used in a popular TV show Ready Steady Cook) were designed, printed and laminated for participation in debate. At the Tatton there was no access to digital equipment in the café so the roadshow used poster-sized printed slides. At Inspire a mixture of projected electronic presentation and posters was used depending on the need of participants. The design of the roadshow evolved based on interactions with different marginalized groups, as a basis for engaging with other communities to raise awareness and understanding of AI. The structure and content of the AI Roadshow is given in III.B.

5) Deliver interactive community roadshow, selecting content dynamically according to the participant need: A key concern for the researchers was that any percieved barriers that might prevent citizens from attending and engaging in the AI Roadshow must be minimized. The aim was that the AI Roadshow should be like a social event rather than work or school. Therefore much consideration was given to appearance such as the style of dress (informal), the use of titles (Dr. and Prof. were dropped), the method of delivery (informal, discursive) and the layout of the room (café style, not lecture or classroom style). In this context, informality is critical for creating a shared and safe space where equals can engage in discussion. During each roadshow, stories of AI applications that had been highlighted in the media were selected dynamically according to the interests and lived experience of the participants, by observing scenarios with best engagement during discussion and debate.

6) Evaluate AI roadshows: Evaluation is important to measure change, even when aspiring to evoke a social mood for the AI Roadshow. Therefore attendees were asked to complete a short anonymous survey (pre-event) at the start to capture what people understood about AI and whether they trusted it, in a given context. People then completed a short anonymous survey (post-event) that asked similar questions to the pre-event survey. The aim was to determine if people learned anything during the roadshow, if their opinions changed with regards to trust and also whether they enjoyed the roadshow. People also had the option of completing a diversity and inclusion monitoring survey. All surveys were anonymous.

7) Recruit citizens for further training and involvement in People's Panel: The Community AI Roadshow was used as a vehicle for recruiting interested citizens to sign up for two days of further free AI training and to take part in the People's Panel for AI (see section II.C). An introduction to the PPfAI was given at the end of the AI Roadshow, and a paper information sheet supplied. Interested attendees added their names and contact information to the paper sign-up sheet. The perceived value of taking part in the PPfAI had to be very clear to community members.

B. Structure and content

AI is frequently portrayed in the media, with a significant proportion of news stories portraying AI in negative late as "Bad news sells" [26]. Media comes in many different formats where typically the same story or theme can reach different audiences through different media (print, social media, television, radio, etc). Storytelling is an effective way to communicate complex ideas to a wide range of audiences, so we adopted a storytelling approach to raise awareness of AI in our roadshows. The roadshow structure is shown in Table I.

TABLE I. WORKSHOP STRUCTURE

Item	Description
Welcome and Icebreaker	Overview of the roadshow and obtaining participant informed consent.
Pre-roadshow questionnaire, available in hard copy format, standard and large font size	Designed to evaluate community members' understanding of AI, ethics and confidence before the roadshow. Anonymous.
What is Artificial Intelligence and why does it matter?	Discussion and simple English definitions of AI, data and ethics. Explanation of how machines learn using real-world examples of how humans make decisions (e.g. choosing to go to the supermarket or not based on weather factors)
Artificial Intelligence in our everyday lives	Utilising storytelling and discussion. Followed by picture card voting on whether community members would trust specific applications.
Deeper dive case studies	Case Study 1: Social Care and AI – adapted from [27] Case Study 2: AI and Education
Community Discussion - What technology don't you like and why?	Open free discussion for community members to reflect on the use of technology in their everyday lives.
Getting involved in the People's Panel for AI and being a community voice	Short presentation on how to get involved in more upskilling in AI and be part of the People's Panel
Post-roadshow questionnaire	Designed to evaluate community members understanding of AI, ethics and confidence after the roadshow (anonymous) and also evaluate what they liked and didn't like about the roadshow
Hot meal and open Q and A	Fundamental for fostering trust with researchers' team. Answering questions and obtaining signups for the People's Panel.

C. Evaluation Methodology

The main challenge in designing an evaluation for the Community AI Roadshows was to ensure that the method was appropriate for the target audience – members of traditionally marginalized communities who may be disengaged from education and employment, and whose motivation for attending may be around the food provided. Evaluation tasks had to be quick and easy; it was important that the evaluation did not become another barrier to participation. Therefore, it was decided to design two short, tick box surveys, one to be completed at the start of the roadshow and one at the end. The surveys had to be short (maximum 5 questions), carefully worded and easy to complete on paper. The pre-event survey contains four questions to score from 1 to 10 that evaluate participants' understanding, trust and confidence in AI (Questions 1-4 in Table II). The post-event survey was identical to the pre-event survey with an additional question to evaluate the roadshow itself (Question 5 in Table II).

To capture background information on participants an optional, anonymous Background Information Sheet was designed. The form design follows best-practice advice [28] and includes ten questions: 1) age, 2) gender, 3) ethnicity, 4) health condition or impairment, 5) caring responsibilities, 6) employment type, 7) eligibility for free school meals[1], 8) would you describe yourself as from lower socio-economic background, 9) highest level of qualification, 10) postcode[2]. However, the form unavoidably appeared intimidating, as the resulting paper questionnaire had a lot of text (as many option boxes were included for ease) and was two pages in font 12 point, four pages in large font. This led to an important research question for future work: What is the most appropriate method to capture EDI data from marginalized communities?

The motivation of participants to learn more and engage with AI companies was evaluated using the number of citizens who signed up for further training.

The quantitative evaluation was supplemented using qualitative methods: by observation and recording of the comments made by attendees, and the questions asked, the design of the AI Roadshow evolved.

IV. Results and Discussion

Three "meal and AI roadshows" were delivered in the two target community venues (The Tatton in Salford, Inspire in Levenshulme) with 29 participants who completed both evaluation surveys (around 50% conversion from registration). It was observed that there were a few attendees who left early and some who chose not to complete the evaluation (around 10 additional attendees).

A. Description of Participants

24 participants chose to complete the optional Background Information Sheet (82% of participants completing the evaluation). The surveys were anonymous and could not be linked to evaluation answers but were used to get a picture of the citizens reached. Participants in each roadshow had different backgrounds with one notable difference being age: as mentioned in section III.A.2, the second AI Roadshow at Inspire was integrated into a community activity for older people, so all 10 participants who completed the Background information Sheet were aged 65+, whereas there was a spread of age groups in the other two roadshows. The overall gender split was 63% female and 37% male. 48% of respondents reported having a health condition or impairment. Fig. 1 shows the range of educational backgrounds of respondents (relates to UK qualifications). It can be seen that there was a range of qualifications, from none to high school to degree level.

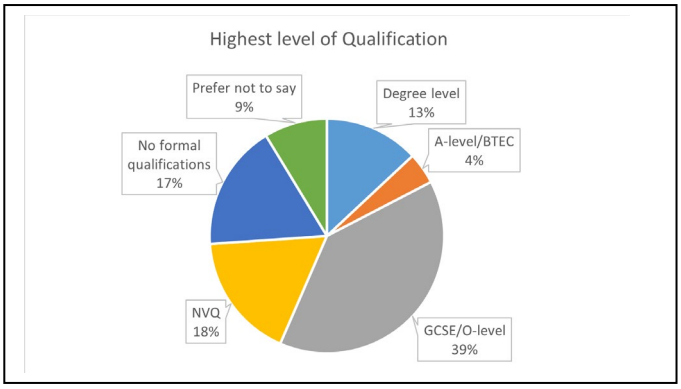

Fig. 1. Highest level of qualification.

B. Results of Evaluation Surveys

The overall response to the AI roadshows was positive and those participants who attended were very engaged, had strong opinions and asked lots of questions. Table II shows the results of the Pre- and Post-Roadshow surveys, where participants rated questions from 1 to 10, with 1 being low and 10 high. In Table II the mean rating for each AI Roadshow (Tatton, Inspire-1 and Inspire-2) is shown separately along with the total mean rating. The difference in community groups is clear to see in the different ratings and is especially notable in the Inspire-2 roadshow which targeted older people.

In Table II the percentage change from the pre-roadshow to the post-roadshow rating indicates an improvement in awareness and confidence over all four questions. Fig. 2 illustrates for each question how the improvements differ across each roadshow along with the mean change. It can be seen from Fig. 2 that increases in ratings are consistently lower for the Inspire-1 roadshow, which suggests that participants had more awareness and understanding at the star of the roadshow. The highest improvement in ratings can be seen in the Inspire-2 roadshow, which targeted older people.

Question 1, 'I understand what AI means' increased by 33% on average (29% median), but the increase in

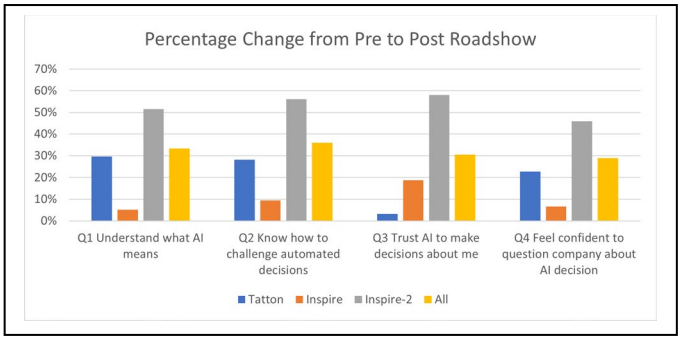

Fig. 2. Percentage Change in Ratings by Location.

[1] In the UK, family eligibility for free school meals is often used as an indicator of deprivation.

[2] In the UK, a postcode designates an area with several addresses, and can be used as an indicator of deprivation by consulting the English indices of deprivation [29].

TABLE II. PRE- AND POST- ROADSHOW SURVEY RESULTS

Question	Pre-Roadshow Mean ratings (1 to 10)				Post-Roadshow Mean ratings (1 to 10)				Total % Change
	T[a] (n=10)	*I-1[a] (n=6)*	*I-2[a] (n=13)*	*Total (n=29)*	*T[a] (n=10)*	*I-1[a] (n=6)*	*I-2[a] (n=13)*	*Total (n=29)*	
1) I understand what Artificial Intelligence means	4.50	7.30	3.00	4.40	6.40	7.70	6.20	6.60	33%
2) I know that computers make automated decisions that affect my everyday life and I know how to challenge these decisions	4.10	4.80	1.80	3.20	5.70	5.30	4.10	5.00	36%
3) I trust the use of Artificial Intelligence to make decisions about me	3.00	4.30	1.30	2.50	3.10	5.30	3.10	3.60	31%
4) I feel confident about how to question a company that used my data to make an automated decision about me	4.10	4.20	2.00	3.20	5.30	4.50	3.70	4.50	29%
5) I enjoyed taking part in this roadshow	--	--	--	--	8.10	9.70	8.50	8.70	--

[a.] T=Tatton; I-1=Inspire Roadshow 1; I-2=Inspire Roadshow 2.

understanding was notably higher at the Tatton (30%) and Inspire-2 (52%) compared to Inspire-1 (5%). This indicates that participants at the first Inspire roadshow were more aware of AI than at the other two roadshows.

In question 2, 'I know that computers make automated decisions that affect my everyday life and I know how to challenge these decisions', ratings increased by 36% on average (40% median). Again, there is a similar difference between roadshows in the change, with Tatton increasing by 28%, Inspire-2 increasing by 56% and Inspire-1 increasing by 9%.

In Fig. 2, the results for question 3, 'I trust the use of Artificial Intelligence to make decisions about me' are interesting, showing that while there was an average improvement in ratings of 31% (75% median), the Tatton participants ratings only increased by 3% whereas the increase was 19% for Inspire-1 and 58% for Inspire-2 participants. This breaks the general pattern across roadshows seen in Fig. 2, suggesting that maybe participants at the Tatton roadshow are less trusting of AI, or the roadshow did not give them more reason to trust AI decisions. The reasons for this are likely complex and a more in-depth study with the community is required to try to uncover and understand the result.

For question 4, 'I feel confident about how to question a company that used my data to make an automated decision about me', ratings increased by an average of 29% (40% median) with the biggest change again noted in the Inspire-2 roadshow participants (46%) and Tatton (23%), with the Inspire-1 increase being 7%.

In Table II, question 5 'I enjoyed taking part in this roadshow' was scored 8.7/10 on average, which suggests that the aim of creating an enjoyable experience was met.

Across all three roadshows, there were 12 citizens who signed up for further training and to be part of the People's Panel for AI, which is 31% of attendees and 41% of those participants who stayed for the whole roadshow and completed the evaluation. This suggests that the AI Roadshows were successful in sparking an interest in some citizens to learn more about the ethical use of AI and engage with AI companies.

Overall, the results of the evaluation show that the pilot study was successful in designing a method to raise awareness of AI, ethical issues and citizen rights in traditionally marginalized communities. In particular:

- the results show that an interactive workshop has been successful in explaining and raising awareness of AI for groups with a mixed level of background education and experience, thus answering RQ1;
- the pre- and post- roadshow surveys demonstrate that raising awareness of AI through an interactive workshop has increased citizen understanding of AI and confidence to question organisations about automated decisions, thus answering RQ2;
- the 12 sign-ups for further training and to be part of the People's Panel for AI indicates that raising awareness of citizen rights around AI in a community setting can empower citizens to learn more and engage with AI companies, thus answering RQ3.

C. *Recommendations for AI researchers engaging with marginalized communities*

The experience of designing a methodology for raising awareness of AI, ethical issues and citizen rights in hard-to-reach communities has led to a number of learnings and recommendations that may be of use to other researchers wishing to engage with traditionally marginalized communities. The following recommendations are drawn directly from the experience during this research and are in line with more general strategies for reaching underrepresented communities [30].

- *Be community-led* – the principle of 'meet you where you are' is vital for engaging with marginalized and disengaged communities so it is critical to deliver the AI Roadshows in familiar community spaces. It is also important to understand the time and place (e.g., type of room within the community space) that best suits community members, and the choice of meal that will be most appreciated.
- *Create a reassuring and safe space* for people to speak up, ask questions and give their opinions. It is important to reinforce that all in the space are equal, and to be very open and supportive when people speak.

- *Remove barriers* – for example consider the way you present yourself and adjust if needed. For example, dress casually, use your first names, don't use titles such as Dr and Prof., be very casual to encourage participation.
- *Minimise bureaucracy* – Keep any evaluation forms to an absolute minimum, with a maximum of five simple questions rated on a scale or LIKERT scale and print out in large font sizes.
- *Carefully consider how to capture EDI data* – common forms are intimidating, appear onerous and many attendees will not want to complete them – this is particularly true of those in minority groups who may view this as a box-ticking exercise. Always provide a 'prefer not to say' option to minimize embarrassment. More research is needed about the most appropriate way to capture this background information.
- *Work in partnership with existing community gatekeepers* – for example, volunteers and community groups who are trusted and who understand the needs and interests of the community.
- *Enable inclusive access* – consider the needs of the group, e.g., use large font sizes, provide printouts on paper, use Post-Its and voting cards rather than electronic versions such as menti often used in universities.
- *Be careful to create a shared language* – avoid the use of jargon or words that you have not explained, e.g., algorithm.
- *Be led by attendees* – be prepared to change the case examples and stories based on what is connecting and working and what is not, be prepared to drop part of the content if interest wanes.

V. Conclusion and Future Work

To truly capitalize on the benefits of AI technologies to humanity, it is vital that different voices and perspectives of citizens are understood. The work presented in this paper has explored a method of promoting public trust in AI and data-driven technologies by raising awareness in typically underserved marginalized communities. Engaging with traditionally marginalized communities comes with a number of challenges for researchers. This paper has described a methodology for reaching such communities by developing and trialling an interactive, storytelling Community AI Roadshow to raise awareness of the use of AI technologies, associated ethical considerations and citizen rights around AI.

The pilot study was based in Greater Manchester, UK, where there are several areas of deprivation and over 1 million digitally excluded residents. Two traditionally marginalized communities were selected, each with a central community café or hub serving the residents. An accessible, interactive AI Roadshow was designed adopting a group-based storytelling approach to provide collective empowerment, a space to share personal experiences and connect meaningfully in discussions around AI, ethical issues and citizen rights. The AI roadshows evolved after engaging with each community, and case stories were selected dynamically according to their ability to connect with the interest and lived experience of attendees. The AI roadshows had to be inclusive, accessible and suitable for groups with a mixed level of background and education. Partnering with community groups enabled the roadshow to be designed with the interests and needs of the community in mind. An informal, social approach was taken, and each 2-hour roadshow included a meal and plenty of discussion and debate. The aim was to foster a safe space where everyone was equal and open to discussing different viewpoints.

The evaluation results showed that the Community AI Roadshows were successful in raising citizen awareness and understanding of AI and data-driven technologies, with a 33% average increase in understanding of AI. Awareness of automated decision systems and how to challenge such decisions increased by an average of 36%, and there was an average improvement of 29% in confidence to question a company about automated decisions. Trust in AI to make decisions about citizens rose by an average 31%. There were notable differences in the change across the different community groups, with the biggest change being seen in the roadshow that targeted older people.

The experience of the research and the challenges of engaging with traditionally marginalized communities led to a set of recommendations for researchers wishing to deliver similar projects.

The key contributions in this paper are: a methodology for raising awareness of AI and related ethical issues and citizen rights in traditionally marginalized communities; the development and trial of an interactive, story-based workshop to engage marginalized groups around AI; the evaluation of a pilot study demonstrating increased citizen understanding and confidence in AI; a set of recommendations for AI researchers engaging with marginalized communities.

The methodology presented in this paper has since been adapted for other community groups, and the roadshow has been adapted for other publics such as local government organisations. Future work involves trialling the methodology in different communities across the UK.

An important gap was discovered in the literature that raised a research question for future work: What is the most appropriate method to capture EDI data from marginalized communities?

Acknowledgment

This work was supported by The Alan Turing Institute Public Engagement Grant 2022. Our thanks go to the communities in Greater Manchester who took part: The Tatton Community Café in Ordsall and The Inspire Centre in Levenshulme.

References

[1] Liang, C.A., Munson, S.A. and Kientz, J.A., 2021. Embracing four tensions in human-computer interaction research with marginalized people. ACM Transactions on Computer-Human Interaction (TOCHI), 28(2), pp.1-47.

[2] R. Benjamin, "Race after technology". In Social Theory Re-Wired (pp. 405-415). Routledge, 2023.

[3] C. N. Harrington, S. Erete, and A. M. Piper. 2019. Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. Proceedings of the ACMon Human-Computer Interaction 3, CSCW, Article 216 (2019), 25 pages. DOI:https://doi.org/10.1145/3359318

[4] Hoff, K.A. and Bashir, M., 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. Human factors, 57(3), pp.407-434

[5] Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F. and Agreda, S.G., 2022. To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts. Technological Forecasting and Social Change, 181, p.121763

[6] Schepman, A. Rodway, P (2023) The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust, International Journal of Human–Computer Interaction, 39:13, 2724-2741, DOI: 10.1080/10447318.2022.2085400

[7] Ada Lovelace Institute and The Alan Turing Institute, 'How do people feel about AI? A nationally representative survey of public attitudes to artificial intelligence in Britain (2023). Available at: https://adalovelaceinstitute.org/report/public-attitudes-ai

[8] Choung, H., David, P. & Ross, A. Trust and ethics in AI. AI & Soc 38, 733–745 (2023). https://doi.org/10.1007/s00146-022-01473-4

[9] Mitre, (2023), [online], MITRE-Harris Poll Finds Lack of Trust Among Americans in AI Technology, Available at: https://www.mitre.org/news-insights/news-release/mitre-harris-poll-finds-lack-trust-among-americans-ai-technology Date Accessed 28-01-24

[10] Ada Lovelace Institute. (2022). The rule of trust: Findings from citizens' juries on the good governance of data in pandemics. Available at: https://www.adalovelaceinstitute.org/report/trust-data-governance-pandemics/

[11] K. Crenshaw. 1991. Mapping the margins: Intersectionality, identity politics, and violence against women of color. Stanford Law Review 43, 6 (1991), 1241‑1299. DOI:https://doi.org/10.2307/1229039

[12] Vines J, McNaney R, Lindsay S, Wallace J, McCarthy J. Special topic: Designing for and with vulnerable people. Interactions. 2014 Jan 1;21(1):44-6.

[13] Holmes, H. and Burgess, G., 2022. Digital exclusion and poverty in the UK: How structural inequality shapes experiences of getting online. Digital Geography and Society, 3, p.100041.

[14] Serafino, P., 2019. Exploring the UK's digital divide. Office for National Statistics.

[15] Greater Manchester Combined Authority (2022). Social impact report 2022: Fixing the digital divide. [online] Accessed 2/12/23, Available at: https://www.greatermanchester-ca.gov.uk/what-we-do/digital/digital-inclusion-agenda/social-impact-report-2022-fixing-the-digital-divide/

[16] Prior A, Leston-Bandeira C. Parliamentary storytelling: a new concept in public engagement with parliaments. J Legislative Stud. 2020. https:// doi. org/ 10. 1080/ 13572 334. 2020. 18480 81.

[17] Mar D. Evaluating whether stories can promote social cognition: introducing the Social Processes and Content Entrained by Narrative (SPaCEN) framework. Discourse Process. 2018;55(5–6):454–79.

[18] Marmolejo-Ramos F, Workman T, Walker C, Lenihan D, Moulds S, Correa JC, Hanea AM, Sonna B. AI-powered narrative building for facilitating public participation and engagement. Discover Artificial Intelligence. 2022 Mar 30;2(1):7.

[19] Crockett K, Latham A, Wood M, Abberley L, Rawsthorne M, Attwood S. Building Trust–The People's Panel for AI. In2023 IEEE Conference on Artificial Intelligence (CAI) 2023 Jun 5 (pp. 168-170). IEEE.

[20] S. Brown, Consequence Scanning Manual Version 1. London: Doteveryone.

[21] People's Panel for AI: Terms of Reference [online] Available: https://www.turing.ac.uk/people-powered-ai Date accessed [20-01-24]

[22] Manchester City Council, 'Household and ward data | Census 2021 - Demography Summary' [online], available: https://www.manchester.gov.uk/info/500388/census_2021/8588/census_2021_-_demography_summary/2

[23] Salford City Council, 'Ward profiles' [online], available: https://www.salford.gov.uk/people-communities-and-local-information/my-local-community/ward-profiles/

[24] The Tatton Café and Community Hub , available: https://www.facebook.com/p/The-Tatton-previously-Ordsall-Cafe-Project-Ltd-100057699458287/

[25] The Levenshulme Inspire Centre Community Hub, available: https://www.lev-inspire.org.uk/

[26] LaGrandeur, K. The consequences of AI hype. AI Ethics (2023). https://doi.org/10.1007/s43681-023-00352-y

[27] Leslie, D. Briggs, M., (2019), Explaining decisions made with AI: A workbook (Use case 2: Machine learning for children's social care) , DOI 10.5281/zenodo.4624732

[28] ACAS, Equality and diversity monitoring form template, [online], available at: https://www.acas.org.uk/equality-and-diversity-monitoring-form-template

[29] Crown copyright, English indices of deprivation 2019: mapping resources, [online], available at: https://www.gov.uk/guidance/english-indices-of-deprivation-2019-mapping-resources

[30] Freshminds, Culture on Demand: Ways to engage a broader audience, 2007, [online], available at: https://www.culturehive.co.uk/wp-content/uploads/2013/04/Culture-On-Demand-full-report1.pdf