

Building Trust – The People’s Panel for AI

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

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

Melissa Wood
Technology Consultant
HD Labs
Stockport, UK
melissa.wood@hilltopdigitallab.com

Luke Abberley
Senior Data Scientist
Synectics Solutions Ltd
Staffordshire, UK
luke.abberley@synectics-solutions.com

Mat Rawsthorne
Head of Research, HD Labs
Stockport, UK
mat.rawsthorne@hilltopdigitallab.com

Sam Attwood
School of Psychology and
Computer Science
University of Central Lancashire
Preston, UK
ORCID: 0000-0001-8140-6157

Abstract—This paper describes The People’s Panel for AI – a mechanism to build public trust in AI products and services from conceptualization to deployment. To increase public awareness of how AI and data-driven systems are affecting the lives of ordinary people, a series of Artificial Intelligence Roadshows were delivered in community centers. Community members were recruited to the People’s Panel and completed two days of training about key aspects of data, AI and ethics, including learning a technique for exploring ethical aspects of new technologies (consequence scanning). As part of a pilot study, four People’s Panel sessions were held where tech businesses and researchers pitched their ideas and discussed questions and concerns of the panel members. Through participating in the panel, panel members reported an increase in confidence in being able to question businesses and businesses heard a diverse stakeholder voice on the ethical impacts of their products / services, leading to change.

Keywords—Trustworthy AI, Citizen engagement, consequence scanning

I. INTRODUCTION

Citizen trust in AI applications and data driven technologies is at the forefront of ethical guidelines, principles, practices and future legalisation [1..3]. It is internationally recognised that both the design of products and services embedding AI and policy itself should be inclusive and there should be a “*creation of mutual language and understanding of AI*” [4]. Results of a recent study (Eurobarometer 92.3) on 30,000 EU citizens highlighted that public trust is influenced by discrimination, accountability and responsibility [6]. However, the practicalities of building trust through directly engaging with citizens in hard-to-reach communities, those with low digital literacy or suffering from digital poverty, or simply a lack of skills and knowledge of data and AI is a significant challenge. To date, actual public engagement in AI Policy has been very limited, yet national country-specific AI strategies, such as in Finland, state that “*inclusion of the public in defining ethics for AI would increase levels of trust and predictability in AI regulatory system*” [5]. This paper presents a concept known as The People’s Panel for Artificial Intelligence (PPfAI). The purpose of the panel is to 1) identify people’s concerns and viewpoints

about the use of AI in their everyday lives through interactions with products and services; 2) provide a sounding board for new innovative business solutions and university-based research in Artificial Intelligence and Data Driven Technologies and 3) engage in continuing dialogue with different communities about new technology. In this paper, we highlight the method used to establish the PPfAI and present initial results of a pilot study on business engagement. The main contribution of this work can be summarized as a practical method to establish a People’s Panel for AI and links to a terms of reference document which will allow others to convene and facilitate such panels.

II. METHODOLOGY - HOW TO GROW A PEOPLES PANEL?

A. What is a public panel?

A public panel is a group of citizens who are brought together to learn about, discuss and give their views on a particular issue. In this work, our theme was the ethical and responsible use of data and AI in people’s everyday lives. The purpose is to enable and stimulate discussion by considering the consequences of new products and services developed by business and researchers. This gives people in communities a voice and provides opportunities to feed back to business and researchers – allowing everyday citizens, especially from traditionally marginalized communities, to potentially influence change.

B. Community Roadshows

Community Roadshows on the theme of data and AI were developed to raise general awareness of the project and to give an opportunity for education/upskilling of citizens in Greater Manchester from two traditionally marginalised communities in the UK with little to no knowledge of AI. The AI Roadshows were interactive and covered a) What is Artificial Intelligence and why does it matter? b) examples of existing applications using AI in people’s everyday lives c) real-world AI case studies for debate d) introduce the People’s Panel and what is involved. Pre and post roadshow surveys showed an increased understanding of AI of citizens by 33% ($n=29$). The roadshows were also a mechanism to recruit citizens who were interested in

becoming panel members and were asked to sign up for two free PPfAI training days.

C. Citizen Training

Citizens who volunteered to become panellists attended two interactive training days which involved learning about the use of data and AI through debate and discussion. Training day 1 provided a deeper dive into a) Data b) How an AI model is built and learns from data through patterns c) Ethical AI (Bias, fairness, explainability, accountability) d) A risk-based approach to AI e) the law f) Consequence Scanning technique. Training day 2 covered a) Practicing and applying the consequence scanning technique [7], b) Harms Modelling technique [8] c) roles and responsibilities of the PPfAI d) a practice panel (mock). At the end of each training day an evaluation was undertaken to allow panel members to discuss and reflect on the panel training, their motivation and confidence in taking part. A specific challenge was ensuring that the training material was inclusive to those with low digital literacy, and ensuring all examples were relevant to the lived experiences of the community. Figure 1 shows an example of explaining machine learning and an example of Human in the loop decision making in health care.

Consequence Scanning, originally developed by doteveryone [7] is defined as an “*agile practice for responsible innovators*”, that can be integrated throughout the development lifecycle. It is designed to facilitate a conversation about the impact of a potential new products and services and assigns responsibility to act. Consequence scanning first seeks to identify all stakeholders who may be impacted by the product or service and requires thinking out of the box. The following three questions are considered:

1. What are the intended and unintended consequences of a product or feature?
2. What are the positive consequences we want to focus on?
3. What are the consequences we want to mitigate?

The power of the technique is in its simplicity which makes it suitable for diverse citizens to start a deeper dive into their understanding of AI and data driven technologies.

III. PEOPLE’S PANEL FOR AI FRAMEWORK

A PPfAI Panel Session comprises a facilitator, a panel chair, 5-7 trained panel members and 1-3 business representatives. A typical People’s Panel lasts for two hours and includes

- Welcome and Introductions (Panel and Business)
- Aim and Objectives of the Panel
- Business Pitch (20 minutes)
- Panel Q&A to the Business
- Panel retire for Consequence Scanning and the discussion is documented by the Facilitator (25 minutes)
- Panel feedback to Business provided by the Panel Chair
- Business response, panel discussion and evaluation

Following the panel session, written feedback is sent to the business that includes questions raised by the panel, intended/unintended consequences, positive consequences and suggestions for mitigations. Further questions and comments can be fed back to the business following reflection by panel

members after the panel session via the facilitator. To close the loop, after the panel session the business has the opportunity to feed back any changes to their processes, research or products based on suggestions of the panel by contacting the panel facilitator directly. The facilitator will then share information on these adjustments with the panel members to measure the longer-term impact of the PPfAI. A co-produced draft of the Terms of Reference can be found here [9].

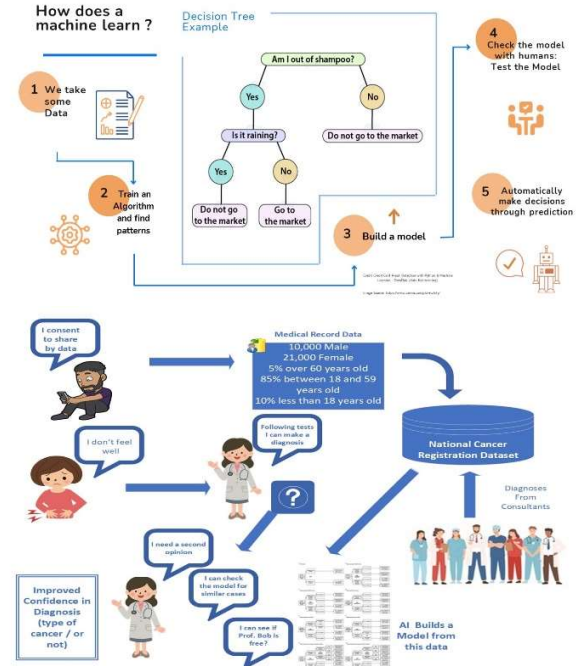


Fig. 1: Example of Citizen Training slides



Fig. 2: Citizen Certificate of Accomplishment

IV. BUSINESS PILOT STUDY

This section provides brief background and feedback from two of the businesses, HD Labs [10] and Synectics Solutions [12], who took part in the first pilot of the panel in November 2022. Following each two-hour panel, each business was emailed a short report summarising questions asked by the panel and a summary of the consequence scanning. HD Labs [10] are a digital services start-up focused on improving citizens experience of accessing public services. They focus their strengths on the enablement of personalisation by working to

make personal data easier to reach at the right time and place, ensuring clients and their users trust the products and platforms they build through ethical disciplines and standards. They push to ensure services are as relevant as possible by smart use of available data and expertise where data will assist in a positive outcome. HD Labs was the first company to participate in a panel session. Melissa Wood, Technology Consultant at HD Labs, said: “We presented to GM People’s Panel for AI (PPfAI) one of the critical automation components we’re building as part of our SMART Interventions service. It was an invaluable and insightful opportunity for us as a business to get different perspectives. More important to us was how engaged the panel was and the quality of ideas contributed to risk mitigation and product improvement - evidencing the business benefit of empowering people by providing transparency of our innovation. We recommend that other businesses developing data-driven technologies dedicate the time and commitment to proactively provide transparency as part of their own ethical assurance. Fantastic service, we look forward to seeing how PPfAI scales.”

A. Synectics Solutions was the second business to pitch to the PPfAI panel. This was a different experience for the panel as the product was developed in house and mature yet constantly evolving. Synectics Solutions [11] has 30 years’ of experience of using data and insight to power their award winning AI product known as Precision, which is used to detect and prevent fraud within the financial sector. Senior Data Scientist, Luke Abberley stated *“I advocate for ensuring Artificial Intelligence is implemented ethically by guaranteeing that the real-world data I use and the models I develop are fair and unbiased towards underrepresented groups of people. This is achieved by conducting advanced analytics on the data and ensuring that all models developed are transparent through the use of explainability techniques at the enquiry level. Therefore, I feel it is vital to be able to explain to laypersons of different backgrounds how AI can be used to benefit all.”*

V. CONCLUSION AND NEXT STEPS

One of the motivations for this project was to build people’s confidence to ask questions about how their data and AI is being used by businesses through an increased understanding of the field. This was achieved significantly through this pilot study and will be the subject of a forthcoming publication e.g., from the roadshows alone, responses to the question ‘*I feel confident about how to question a company that used my data to make an automated decision about me*’ increased by 29% (mean); 40% (median). All citizens who took part received a tangible certificate of accomplishment (Fig.2). The brief qualitative feedback presented in this paper shows that the

panel was also valued by participating businesses. Initial analysis indicates that the panel has influenced their research/product development with 57% indicating that the PPfAI has influenced change in ethical practices and product design. Further work will provide a full quantitative analysis across those business who took part. This small pilot shows what is possible on a small research budget (10k + in kind time) over a 6-month period. The aim is to find both national and international People’s Panel for AI champions who will take ownership to deliver community roadshows and training, widening participation and reach. Champions can be from academia, business, government, or the communities that are being engaged. The key is ensuring panelists feel valued and are compensated for their time, so they are treated as equals.

ACKNOWLEDGMENT

We would like to thank those citizens who volunteered to be trained up to be People’s Panellists and took part in the live panels. We would also like to thank Hannah Berry from GMVCO who acted as a sounding board for the PPfAI, and The Alan Turing Institute who funded and supported this work.

REFERENCES

- [1] EU AI Act, 2023, [online], Available: <https://artificialintelligenceact.eu/>, Date accessed [20-02-23]
- [2] White House, 2022, The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, [online], Available: <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>, Date accessed [20-02-23]
- [3] United Nations Educational, Scientific and Cultural Organization, 2021, [online], Recommendations on the Ethics of Artificial Intelligence, available: [Recommendation on the Ethics of Artificial Intelligence - UNESCO Digital Library](https://unesco.org/en/articles/artificial-intelligence-experts-and-policy-makers-discuss-inclusion-ai-policy-development), Date accessed [20-02-23]
- [4] UNESCO, 2022, Artificial Intelligence experts and policymakers discuss inclusion in AI policy development, [online], Available: <https://www.unesco.org/en/articles/artificial-intelligence-experts-and-policy-makers-discuss-inclusion-ai-policy-development>, Date accessed [20-02-23]
- [5] Wilson, C., 2022. Public engagement and AI: A values analysis of national strategies. *Government Information Quarterly*, 39(1), p.101652.
- [6] Omrani, N., Riviuccio, 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 & Social Change*, 181, p.121763.
- [7] Brown S. (2019) *Consequence Scanning Manual* Version 1 . London: Doteveryone.
- [8] Microsoft, Harms Modelling, (2021), [online], available: <https://docs.microsoft.com/en-us/azure/architecture/guide/responsible-innovation/harms-modeling/>, Date accessed [20-02-23]
- [9] Peoples Panel for AI: Terms of reference – draft [online], Available - <https://www.mmu.ac.uk/research/research-centres/cfacs/projects/gm-people-panel-ai>
- [10] HD Labs, (2023), [online], Available: <https://hdlabs.co.uk/>, Date accessed [20-02-23]
- [11] Synectics Solutions, (2023), [online], Available: <https://www.synectics-solutions.com/>, Date accessed [20-02-23]