

Activity 3.4: Researching with communities in and beyond AI

Duration: 60 minutes

Aims:

- To understand the role of data in the AI Lifecycle
- To explore the role of communities in data collection and curation
- To discuss what are the Problems / Challenges / Benefits of citizen engagement

Requirements: Flip chart, markers, post-its, computer and screen, chairs and tables

Activities:

- The facilitator can utilise the presentation provided to for this activity. Start by using slides 3 overview the aims of the session.
- Slides 4 and 5 are used to perceive the typical paths to public engagement – the one way sharing of information and knowledge (slide 4) compared with the two-way exchange (slide 5).
- Slide 6 just hights that engagement with AI adds further complexities. Slide 7 is an example from the British Medical Journal that illustrates the impact of data representativeness in the AI lifecycle.
- Slides 8 and 9 look summarise data and AI challenges and the in the context of the AI Lifecycle.
- Exercise (Slide 10) asks the participants to explore the question “What could be the role of Communities in Data Collection / Curation in the context of Artificial Intelligence?”. This can be done in groups or as a class. Use post it-notes to collect ideas. If delivered online use Menti or Padlet. Slide 11 can be used to conclude this discussion.
- The facilitator should lead on slides 12 – 13 that introduce participatory AI.
- Exercise (slide 14) asks the participants to explore the question “Does introduction of community voice into data collection / curation and labelling introduce more bias or better data representation? Why?”. This can be asked to the entire group.
- The facilitator should lead on slides 15 – 18 that look at a snapshot of challenges with regards to data. Please not this is not complete nor comprehensive – the level of detail will depend on the background of participants and be used to prompt discussion.
- Exercise (Slide 19) asks the participants “What are the Problems / Challenges / Benefits of citizen engagement in your own research?” This is an individual activity or can be done in groups. Allow up to 10 minutes with 5 minutes to reflect together as a group and look at whether certain challenges can be mitigated.

Facilitators notes:

- Slides 4 and 5 are used to illustrate that the most common public engagement activities are researcher/ instructor teaching led with information often flowing one way (slide 4); whereas co-production allows a more balanced sharing of information and knowledge (slide 5);
- Slide 7 Provides a great example from the British Medical Journal [Does “AI” stand for augmenting inequality in the era of covid-19 healthcare? | The BMJ](#). This example is used to demonstrate the impact of using data to train a model which is not representative. To quote

from this article: “Data representativeness: The datasets used to train, test, and validate AI models are too often insufficiently representative of the public. For instance, datasets composed of electronic health records, genome databases, and biobanks often under sample those who have irregular or limited access to the healthcare system, such as minoritised ethnicities, immigrants, and socioeconomically disadvantaged groups. The increased use of digital technologies, like smartphones, for health monitoring (e.g., through symptom tracking apps) also creates potential for biased datasets. In the UK, more than 20% of the population aged 15 or older lack essential digital skills and up to 10% of some population subgroups do not own smartphones. Datasets from pervasive sensing, mobile technologies, and social media can under-represent or exclude those without digital access. Whether originating from medical data research facilities or everyday technologies, biased datasets that are linked—such as in biomedical applications that combine pervasive sensing data with electronic health records will only exacerbate unrepresentativeness “.

- Slide 8 provides a summary of high-level AI and Data challenges and slide 9 is used to discuss how bias propagates in the AI lifecycle.
- Slide 11 and 12 introduce the Nesta - A framework for operationalising participatory AI.
- Slides 13 – 18 that look at a snapshot of challenges with regards to data. Please note this is not complete nor comprehensive – the level of detail will depend on the background of participants and be used to prompt discussion

Resources:

- PEAs in Pods Training Activity 3.4 - resource - researching with communities in and beyond AI.pdf

Additional resources:

- Franzen, M., Kloetzer, L., Ponti, M., Trojan, J., Vicens, J. (2021). Machine Learning in Citizen Science: Promises and Implications. In: , *et al.* The Science of Citizen Science. Springer, Cham. https://doi.org/10.1007/978-3-030-58278-4_10
- [Participatory AI for humanitarian innovation \(nesta.org.uk\)](https://www.nesta.org.uk/participatory-ai-for-humanitarian-innovation)
- [Research with potentially vulnerable people – UKRI](https://www.ukri.org/research-with-potentially-vulnerable-people)
- [Consent – UKRI](https://www.ukri.org/consent)
- [Data protection in research: An overview | Manchester Metropolitan University \(mmu.ac.uk\)](https://www.mmu.ac.uk/data-protection-in-research)
- [Policies and guidelines | Research ethics at The University of Manchester](https://www.manchester.ac.uk/policies-and-guidelines)
- [Research Data Management | University of Salford](https://www.salford.ac.uk/research-data-management)