NOT-EQUAL

EPSRC NetworkPlus: Social Justice through the Digital Economy

Project Review Form - Mid-Term Review Pilot Project

Please submit this form **to notequal@ncl.ac.uk**.

GENERAL INFORMATION	
Lead Applicant (PI): Prof. Alessio Malizia **Email address: a.malizia@herts.ac.uk** **Job Title: Professor** **Department: School of Creative Arts** **Organisation: University of Hertfordshire**	**Co-Investigators (names and organisations): Dr Silvio Carta, University of Hertfordshire** **Supporting Partner(s): WeandAI, Data Reply** **Project Title: MiniCoDe** **Project Reference Number: NE2. 001**

1. SUMMARY

Please provide a summary of the activities and/or initial findings of your research project to date. This also includes events, engagement activities with non-academic partners and any other activities. Please include any images or website links that could be used for dissemination purposes (at least 500 words).

This project aims to tackle social injustice in future algorithmic-based decision-making applications, namely devise strategies to expose, counterbalance, and remedy bias and exclusion built into algorithms, considering fairness, transparency, and accountability. We are developing a Design Fiction Toolkit (DFT) in the form of a collaborative workshop session with supporting materials to be used by stakeholders to experiment with scenarios to expose potential bias and reflect on mitigation strategies at design time.

We carried out a Literature Review surveying existing Design Fiction methods and toolkits from Google Scholar, Scopus, and Elsevier, focusing on narrative design and communication to devise an appropriate supporting strategy for the workshop facilitator.

We've also scoped several papers analyzing the different types of bias that might inherently be embedded in algorithms and datasets that will constitute a valuable guideline to design the future experiments we will carry out.

Starting from the literature, we designed an initial workshop plan which runs in seven different phases: (1) an inspirational narrative is prompted to participants to communicate the design brief, (2) participants are clustered in groups, and each group starts the idea generation, (3) the ideas get refined and later (4) enriched, then the best candidate idea selected within each group is (5) conceptualized; the resulting concept is then (6) analysed in light of a set of ethics principles embedded in scenarios (in the form of cards) to expose its potential biases, and finally, each group (7) reports its findings to the others to get final feedback. We're now operationalizing each phase to define which input and output it produces and its goal. We will then select the most appropriate technique to be used within each phase. The following table summarizes the findings.

	Phase	Description	Input	Output	Goal	Method
1	Inspirational Prompt	Inspirational Wall: video, timeline, *newspaper from the future*	Innovative Service	Inspirational Narrative	Inspire audience and get them into the right mindset (Design Fiction)	Narrative Development
2	Seed Generation	Read the design brief and each group member starts generating ideas	Design Brief/Narrative	Initial Ideas	Generate *various* ideas implementing the design brief	6-8-5
3	Ideas Refinement	Inform previous design or generate a new idea with cards	Initial Ideas	Refined Ideas	Refine ideas and select the most promising	Card Sorting
4	Ideas Enrichment	Enrich ideas with concepts	Refined Ideas	2/3 Enriched Ideas	Enrich ideas with card decks	Card Decks (PLEX/Things for the Future)
5	Conceptualisation	Collaboratively create a concept of service from the selected idea	2/3 Enriched Ideas	Best Candidate Idea	Turn one idea into a service	Co-Design
6	Ethical Augmentation	Inform the concept with two ethics principles per group	Best Candidate Idea	Ethically Augmented Concept	Introduce ethic principles (mitigation strategies?) into design	Cards with ethics principles
7	Collaborative Decision Making	Present concepts back to other groups and receive feedback	Ethically Augmented Concepts	Feedback	Obtain feedback on concepts with different ethical principles	Sticky Notes

In line with our work plan, we run a kick-off meeting with all the participants of the project, including representatives from Data Reply and WeandAI. During the meeting, we introduced the project to the end-users and presented its objectives. We summarised the current status of the toolkit and agreed to involve partners in the next stage of the toolkit design: WeandAI is going to be involved in establishing ethics principles and mitigation strategies to be included in phase 6, whilst Data Reply agreed to run a workshop within their workforce to test the toolkit once it's ready.

Please indicate if these details can be shared in a blog post on the Not-Equal website   **YES**   ~~NO~~

2. WORK PLAN

Please explain any deviations from your work plan, the reasons for this and plans to address the issue (up to 250 words)

Due to contingency related to the pandemic and legal issues with signing the contract, we started our project in March 2021. Our main partner, Northumbria University, had problems signing legal documents and the PI going on maternity leave. Thanks to a reasonable effort from our team at UH, we are mainly on track with the work plan depicted in the proposal:

	WP / TIME (month)	WP LEAD	MONTHs 1-3			MONTHs 3-6			MONTHs 7-8	
			1	2	3	4	5	6	7	8
1	WP1 - PROJECT MANAGEMENT	UH								
2	WP2 - RECRUITMENT	NI								
3	WP3 - BACKGROUND REVIEW	UH								
4	WP4 - EXPLORATION OF WORK	NI								
5	WP5 - TOOLKIT DESIGN AND DEVELOPMENT	UH								
6	WP 6 - EVALUATION AND TESTING	NI								
7	WP 7 - DISSEMINATION	UH								

Figure 1: original GANTT (UH -University of Hertfordshire, NI – Northumbria University)

WP1 is well underway, while WP2 and WP3 have now been completed (considering month 1 starting March 2021).

WP4 did not progress due to the issue experienced by Northumbria (NI was the leader). WP5 is currently in progress, while WP6 will be executed in the last two months as planned. WP7 is in progress.

We informed not-equal of the issues with our partner Northumbria. We agreed on a modified plan that skipped WP4 and replaced food banks with our partners WeandAI and Data Reply plus one suggested by the network, the Digital Catapult, to help on WP6. We also asked for an extension to the end of September to complete WP6 and WP7 appropriately.

In relation to the evaluation of the DFT (WP6), we are working alongside our partners WeandAI and Data Reply to define a suitable case study for the pilot experiment. We plan on organizing an online workshop event with Data Reply in September to evaluate our toolkit in terms of benefit, usability, and acceptability, including some concepts from the SMACTR auditing framework (Raji et al., 2020).

References

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ...& Barnes, P. (2020, January). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (*FAT* '20*). Association for Computing Machinery, New York, NY, USA, 33–44.

Further Information
If you have any further questions regarding this form, please contact notequal@ncl.ac.uk