

NOT-EQUAL

EPSRC NetworkPlus: Social Justice through the Digital Economy

Project Final Review Form

Please submit this form within one month of completing your project 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 outline the research challenge and question your project aimed to address, in less than 100 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 developed 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.

2. APPROACH

Please provide a summary of the approach of your research project, including any deviations from your work plan, the reasons for this and how you addressed any issues.

We carried out a literature review surveying existing Design Fiction methods (Lindley J., 2015, Johnson B. D., 2011) 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 analysing 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) conceptualised; 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've then operationalised each phase to define which input and output it produces and its goal.

We've pilot tested the toolkit with a small group of academics. After refining it, we ran a final workshop to see whether a small industry team could effectively adopt it to validate AI service designs.

As a consequence of the C-19 pandemic and some delay from our legal department related to the T&Cs for this project, we started our project in March 2021, adapting our schedule as outlined below*:

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	UH	█							
3 WP3 - BACKGROUND REVIEW	UH	█	█						
4 WP4 - TOOLKIT DESIGN AND DEVELOPMENT	UH			█	█	█	█		
5 WP5 - EVALUATION AND TESTING	UH							█	█
6 WP6 - DISSEMINATION	UH			█	█	█	█	█	█

Figure 1: Updated GANTT (UH -University of Hertfordshire)

All WPs have now been completed (considering month 1 starting March 2021).

*To note that our initial proposal included Northumbria University as co-applicant. The Co-I from Northumbria encountered some difficulties before starting the project. We informed NOT-EQUAL of the issues with our partner and agreed a modified plan where we replaced food banks with our partners WeandAI and Data Reply. In addition, we included a new partner suggested by the Network, the Digital Catapult, to help with Evaluation and Testing. We have also been granted an extension to the end of September 2021 to complete Evaluation and Testing, and Dissemination.

Lindley, J. (2015, June). Researching design fiction with design fiction. In Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition (pp. 325-326).

Johnson, B. D. (2011). Science Fiction Prototyping: Designing the Future with Science Fiction , Morgan & Claypool.

3. ACTIVITIES & OUTPUTS

Please list any outputs from your project to be entered in the Not-Equal Researchfish submission. These include events, publications, workshops, webinars, invited talks, media coverage and tools (please include links to open source, git-hubs if relevant) that have resulted from your project.

Please include the following for each entry:

Title: MiniCoDe Kick-off Meeting



Date: 10/06/2021

Type of Event: Meeting

Number of People Reached: 5

Primary Audience: Academics / Industry / Partners

Key Outcomes/Impact: We run a kick-off meeting with all the project participants, including representatives from Data Reply and WeandAI. During the meeting, we introduced the project to the end-users and presented its objectives. We agreed to involve partners in the next stage of the toolkit design, as Data Reply decided to run a workshop within their workforce to test the toolkit.

Title: MiniCoDe Workshop Pilot Study

Date: 23/09/2021

Type of Event: Workshop

Number of People Reached: 2

Primary Audience: Academics / Industry / Partners

Key Outcomes/Impact: We run a Pilot Workshop Study within the MiniCoDe Academic Team together with an academic from another institution and a Professional UX Designer.

Title: Bright Night

Date: 24-25/09/2021

Type of Event: Public event

Number of People Reached: about 100

Primary Audience: Various

Key Outcomes/Impact: We showcased a poster outlining the MiniCoDe Toolkit during the Researchers' Night in Pisa, a public event where a broad audience interacts with academics, receiving exciting feedbacks.

URL: <https://www.bright-night.it>

Title: MiniCoDe Workshop Study

Date: 07/11/2021

Type of Event: Workshop

Number of People Reached: 5

Primary Audience: Academics / Industry / Partners

Key Outcomes/Impact: We run our first Workshop Study with the MiniCoDe Partners. The video of the event will be used for dissemination purposes.

URL: Forthcoming

4. INSIGHTS & IMPACT

Please describe the findings of your project and their significance in relation to potential or actual social impact.

Although it is not easy to predict the future, we know that high-tech products and systems are going to rely on algorithms that will almost certainly harbour some form of implicit bias in the coming decade. This interdisciplinary project employed a design fiction approach to developing a toolkit in 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.



This project was centred on algorithmic social justice, stimulating reflections and actions focused on the fairness of machine learning solutions embedded in socio-technical systems (e.g., decision-making systems) that might have disruptive implications for people and society at large.

We aimed at developing practical responses to social justice issues by experimenting with a new approach to design socio-technical systems that help meet social aspirations and goals in the form of a Design Fiction Toolkit that:

- It helps practically-minded developers apply social justice principles at design time during the ML development pipeline and to signal to researchers where further work is needed.
- Informs the discussion and recommendations to anticipate the impact of Machine Learning applications embedded in Socio-Technical Systems by involving communities such as the WeandAI network (fostering awareness and understanding of AI in the Society).

Our Toolkit responds to the needs of product managers, developers, and data scientists of ML applications (at Data Reply, for example) to mitigate bias, e.g., social, racial, etc. Companies, such as Data Reply, employing Data Scientists trained in using our Toolkit will learn to avoid bias at design time before introducing socially unjust services into society.

The scenarios and requirements based on WeandAI experience on increasing awareness and understating of AI together with the Design Fiction toolkit will help future developers of ML prospective employers (start-ups and more mature companies) to deliver fairer applications and services (e.g. non-ethnically biased credit score apps).

5. REFLECTIONS & FUTURE DIRECTIONS

Please list the key highlights from your project, summarize any lessons learned from this work and outline any future directions or plans to continue activities beyond this project.

- We developed our Design Fiction Toolkit to work offline and online, successfully employed and tested it in two online scenarios.
- We intend to develop the Toolkit further and adapt it to two use cases that emerged during our research:
 - Educational – Our toolkit could be easily adapted to be employed in educational scenarios, such as within Service Design Modules, as part of an ethics awareness activity by embedding elements of Design Thinking.
 - Digital Teams – The toolkit can be refined and used by small heterogeneous teams within innovative startups interested in ethically designing new AI-based features.
- We plan on releasing the Project Website together with an explanatory Video outlining the toolkit, how it works and how it was employed in the final workshop we carried out.
- We plan on disseminating our work in different venues:
 - The outputs coming from our pilot will be submitted as a Late Breaking work to the ACM CHI 2022 – Deadline January 13, 2022.
 - We are planning to submit a proposal for an interactive workshop or demo session within the CRAFT (Critiquing and Rethinking Accountability, Fairness and Transparency) session at the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) 2022, to test the toolkit further and receive feedback from an expert audience – Deadline TBC, traditionally in January.
 - After we received feedback from the CHI Community, we would like to extend the paper and include the results from the first workshop and submit a full paper to the Springer Journal on AI and Ethics.
 - We also plan to engage the general public by publishing a Feature Article on Medium (an online magazine) and on The Conversation describing the toolkit before the end of the year, see for example Malizia, A. Carta, S. (2019): “Science fiction could save us from bad technology”. The

Conversation is a high-impact independent source of news and views, sourced from the academic and research community and delivered directly to the public.

- Future plans to test with companies Data Reply and Digital Catapult (propose a date, like Jan 2022) to introduce the toolkit as a planned case study in data analytics course at UH and Pisa (Data Analysis in the digital humanities course thought by Alessio Malizia)

Further Information

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

