This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

Reflecting on Algorithmic Bias with Design Fiction: the MiniCoDe Workshops

Turchi, T., Department of Computer Science, University of Pisa, Pisa, Italy

Malizia, A., Department of Computer Science, University of Pisa, Pisa, Italy
Faculty of Logistics, Molde University College, Molde, Norway

Borsci, S., Faculty of Behavioural, Management and Social Sciences, University of Twente, Enschede, Netherlands

Abstract — In an increasingly complex everyday life, algorithms – often learnt from data, i.e. machine learning (ML) – are used to make or assist operational decisions. However, developers and designers usually are not entirely aware of how to reflect on social justice while designing ML algorithms and applications. Algorithmic social justice – i.e., designing algorithms including fairness, transparency, and accountability – aims at helping expose, counterbalance, and remedy bias and exclusion in future ML-based decision-making applications. How might we entice people to engage in more reflective practices that examine the ethical consequences of ML algorithmic bias in society? We developed and tested a Design Fiction-driven methodology to enable multi-disciplinary teams to perform intense, workshop-like gatherings to let emerge potential ethical issues and mitigate bias through a series of guided steps. With this contribution, we present an original and innovative use of Design Fiction as a method to reduce algorithmic bias in co-design activities.

The use of Machine Learning (ML) algorithms to assist in operational decisions has become increasingly prevalent in our complex World. As our reliance on such algorithms in decision-making continues to grow, it is imperative that we consider the potential impacts of these algorithms on Society as a whole. One key concern is the issue of algorithmic bias, which refers to the systematic discrimination against certain groups or individuals. This can lead to exclusionary and unfair decision-making, with serious consequences for marginalized and disadvantaged communities.

To address this issue, the concept of algorithmic social justice has emerged as a way to promote fairness, transparency, and accountability [1] in the design of ML algorithms. However, developers and designers may not always be aware of the social justice implications of their work, or may not know how to reflect on these issues and incorporate mitigation strategies in the design process.

To address this gap, we developed MiniCoDe (Minimize algorithmic bias in Collaborative Decision Making with Design Fiction), a new board game-like workshop methodology aimed at assisting ethical design of upcoming technologies that will become ingrained in daily life.

The present work extends our previous research [2] that introduced the concept and preliminary structure of MiniCoDe. The previous work laid the groundwork for this methodology, focusing on its inception and theoretical

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

underpinnings, including the principles of fairness, transparency, and accountability that underlie its design, together with the expert evaluation we carried out.

In this current study, we build upon that foundation, detailing the development, refinement, and application of MiniCoDe as a tool to promote algorithmic social justice. We also present a quantitative analysis focused on the engagement elicited by this methodology, examining workshop data that includes participant feedback and reflections. This provides an overall picture regarding the workshop's efficacy in uncovering and mitigating algorithmic bias in emerging ML applications.

We further discuss MiniCoDe's potential as a resource for multi-disciplinary teams addressing algorithmic bias. Our investigation considers how MiniCoDe facilitates discussions about bias, offers insights for mitigation strategies, and encourages a culture of ethical consciousness among AI teams.

Additionally, we discuss comparable methodologies from the literature, distinguishing MiniCoDe through its application of Design Fiction.

During the workshop, we set up a scenario related to future ML applications with a focus on algorithmic social justice, in order to encourage conversations about the potential for bias. It is intended for multi-disciplinary teams working on the development of these services in small companies and start-ups, such as data scientists, product managers, and AI engineers. These teams may not have the resources or expertise to thoroughly evaluate the ethical implications of the solutions they are implementing, therefore they need a tool supporting them in reflecting on such fundamental issues.

We do not present MiniCoDe in opposition to existing workshops; rather, the format should be seen as a companion to other design strategies and our attempts at condensing our insights into executable steps to broaden the use of such methods and concerns. The workshop is rooted in Design Fiction, an interdisciplinary method that can allow participants (e.g., product managers, developers, NGOs) to generate scenarios (e.g., storyboards) to expose potential bias and reflect on mitigation strategies. By using scenario-based design, design fiction prototyping can provide opportunities to reveal aspects of how technology will be adopted. Therefore, design fictions are a tool to investigate the implications, ramifications, and effects of technology in the future. Although it is not easy to predict the future, we know that high-tech products, such as smart drones or driverless cars, are going to rely on machine learning in the coming decade. Nevertheless, machine-learning algorithms will almost certainly harbour some form of implicit bias. For example, Caliskan et al.'s [3] academic paper, "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases," published in the leading scholarly journal Science, described an autonomous intelligent agent associating words like "parents" and "wedding" with feminine names. In contrast, career-related terms like "professional" and "salary" were assigned to men. Several studies exploring stereotyped data used to train machine learning applications provide evidence that the word-associating agent flawed strategy may be used to train a CV-analyser service with consequences on gender balance.

The research question tackled in this work, therefore, is: *can MiniCoDe Workshops be used to uncover and mitigate algorithmic bias in novel ML applications?* In other words, can it be used to support the ethical design of those emerging AI-based services which will be impacting everyday life?

Related Works

Workshops have played an essential role in HCI for a long time as a way to engage participants with new designs or research opportunities, allowing researchers to investigate a wide range of designs and user concerns, including creativity [4], user participation in the design process [5], user experiences [6, 7], and design fiction [8] to name a few. Emerging from this tradition, an intriguing development has been the use of card-based games to stimulate dialogue about values in technology. For instance, the "Envisioning Cards" toolkit [9] incorporates key principles of value-sensitive design, encouraging attention to human values during the design process. It has been employed for diverse activities including ideation, co-design, and heuristic critique. While this provides a solid foundation for considering human values in design, our approach aims at exploring, discussing and potentially testing perceivable and potential algorithmic biases, especially in the context of ML.

Similarly, the Values at Play (VAP) methodology [10] proposes a framework for incorporating activist social themes in game design, providing a tested methodology

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

to inform designs with a stronger ethical perspective. Nevertheless, our approach takes a broader perspective by addressing the implications of algorithmic biases in real-world applications beyond gaming.

Further, a novel case study called "Quantified Self" [11] combined elements of design fiction and user enactments to construct an immersive theatre experience aimed at fostering public engagement around technology ethics. This approach, while innovative in fostering public discourse on ethics, contrasts with MiniCoDe's direct engagement with interdisciplinary teams working on the front lines of ML development.

These methodologies promote reflective thinking and discussion about the ethical and societal implications of technology, offering an accessible and interactive medium to provoke conversation around the design and deployment of new technologies. These card-based approaches effectively bring diverse participants into a co-design process, making complex concepts tangible and fostering shared understanding and innovative solutions.

Building upon these foundations, we employed design fiction as a cross-disciplinary method for designers, engineers, and product managers, among others, to reflect on the impact of technology, products and services from a human perspective and link this to possible futures.

Design Fiction is an interdisciplinary approach [12], usually implemented in the form of a participatory design workshop to enable participants to build and reimagine concepts into scenarios and, in MiniCoDe workshops, assist machine-learning experts in identifying potential bias and considering mitigation solutions. Design Fiction prototypes [13] can provide an opportunity to disclose aspects of how technology could be embraced by combining logic and fiction. As a result, Design Fiction prototypes serve as discussion starters [14] for future implications, repercussions, and effects of technology.

Recent literature highlighted the importance of ethical considerations in technology and AI applications. Craigon et al. [15] emphasize the ethical implications of digital collaboration, particularly in the food sector, advocating for a multidisciplinary approach that combines elements of design fiction with an 'ethics by design' card-based tool. Similarly, Rezwana and Maher [16] delve into the ethical challenges inherent in human-AI creative collaborations, using design fiction to explore and gather diverse user perspectives on these challenges. While both works offer valuable insights into the ethical dimensions of technology and AI, our approach uniquely focuses on facilitating reflection on these issues during the design phase of a ML application. By doing so, we aim to proactively address potential ethical dilemmas and ensure that the designed solutions are both innovative and ethically sound.

Algorithmic bias has been recognised as a relevant issue in ML applications. For example, IEEE and ISO are currently developing standards that cover algorithmic bias. A new Joint Technical Committee (ISO/IEC-SC42) has been established to develop standards related to AI. However, mitigating algorithmic bias is far from an easy task. Discursive Strategies, such as workshops and discussion forums, are an exciting class of approaches to mitigate algorithmic bias, which guarantees humans override automated decisions where necessary, dealing with situations in which machines would struggle [17]. In this work, we use Design Fiction as a method to introduce a discursive strategy for ML applications to allow participants to create and reconfigure concepts into scenarios to expose potential bias and reflect on mitigation strategies [18].

MiniCoDe Workshops

We used a set of guidance and materials as a method that combines Design Fiction with other rapid ideation techniques to create concepts and storyboards illustrating the participants' reflections on ethical and social impacts of ML applications in society. MiniCoDe was first designed to exploit physical interaction and run in-presence workshop sessions, but we've adapted the material digitally to allow for its usage in remote workshops. The materials comprise a guide board summarizing the instructions for the facilitator and a recap of each workshop phase with its expected duration, a deck of cards from "The Thing from The Future" [24], and a deck of MiniCoDe Ethics cards (Figure 1).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

FIGURE 1. The Thing from the Future (upper-left corner) and the MiniCoDe Ethics & AI original deck of cards inspired by five principles re-elaborated from primary Ethics and AI literature

The digital version of the Workshop was developed on Miro[1], an online collaboration platform supporting remote and simultaneous access and editing.

In the following, we describe each workshop phase in detail. The four distinct phases underlying our workshop approach are: Prepare, Ideate, Refine, and Reflect.

Prepare

This phase happens before the actual workshop and involves just the facilitator, who needs to set up the context of the workshop. We drafted a guide for them to follow and crafted a sample narrative about the workshop topic. We provide a guideline prepared in advance and based on the standard 3-structure narrative (challenge, climax, ending) [19], prompting the facilitator with three questions to generate characters and a story arc. By following the template and answering the questions, the facilitator can outline a brief fictional narrative for inspiration during the workshop. The short story used to set up the scenario is part of a set of materials called the Inspiration Wall [20]; inspiration walls are usually set up as a series of pictures to set participants' mood in participatory design, but we complemented it with additional materials to make the participants' experience more immersive. In MiniCoDe, the Inspiration Wall includes four elements: a brief story, a design brief, and a fictional timeline to help participants focus on the task at hand (Appendixes I-II-III respectively, presenting a real Pilot Case Study about the Metaverse, sampling the variety of inputs that can be used, e.g., a narrative, a fictional timeline, videos, fictional newspaper articles, etc.).

The Inspiration Wall also mentioned potential consequences or ramifications of the application under investigation. The first three stages of Johnson's description of developing a Design Fiction [21] are reflected in this technique. This might be considered Act I of a larger story that the participants were to compose

[1] https://miro.com

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

later. This provides a narrative to the participants to start from, supporting them in building a Design Fiction by proposing a starting point for the investigation.

Ideate

The facilitator welcomes the participants and introduces the workshop. S/he explains the different phases of the workshop, just as one would do when setting up a board game by stating rules and turns. Then s/he proceeds to read the prompt prepared in the previous phase to all participants (from the Inspiration Wall). Groups are then formed to cluster a mix of participants with different backgrounds. Each participant is given 5 minutes to generate 6-8 ideas related to the given prompt and write them on post-it notes (digital in case of using Miro).

According to research on idea generation, there is a link between producing many ideas and the number of good ideas that result [22]. The advantage of employing the 6-8-5 method was the pressure of coming up with a specific number of ideas in a limited amount of time. This activity is individual to foster contributions from all participants and avoid confident participants that speak anything that comes to mind to dominate introverts. Furthermore, by drastically empowering participants' unique and personal visions, we speed up the boundary testing and subsequent growth of the shared design. Once everyone is done, each member pitches their ideas to the group for the next 15 minutes, discussing which ideas sound promising and should be carried over.

This provides a good starting point for idea generation, as each group will finish this phase with 10-15 idea seeds that will be refined and selected later.

Refine

At this stage, participants are asked to refine the ideas generated within each group with the help of a special deck of cards: The Thing from The Future [23]. The deck aims to create interesting and thought-provoking descriptions of hypothetical things from various futures.

This prompt indicates what section of society or culture the thing-to-be-imagined comes from, describes its type and recommends an emotional reaction that it may elicit in a present-day spectator. It is initially composed of four types of cards: ARC, TERRAIN, OBJECT, and MOOD. By selecting one card for each kind, players form a prompt to generate ideas for artifacts from the future. We chose only to use the TERRAIN, OBJECT, and MOOD cards in our workshop. We discarded the ARC cards since those are about imagining a future scenario given previously by our Inspiration Wall in the design brief. Such cards aim to provide inspiration and focus the ideation on broader scenarios considering culture, society and emotional settings.

The facilitator gives each group 20 minutes to select in turn one card for each of these types and use them to enrich the ideas they have generated and form new ones. Then each group will have 10 minutes to discuss and select a single idea that will develop their candidate concept.

Reflect

Finally, at this last stage, another purposely designed deck of cards is used to aid participants in reflecting on their candidate concepts, discovering and remedying built-in bias. Each card represents a different AI Ethics concept taken from the widely popular framework by Floridi [24]. We reported multiple levels of detail for each concept [25], together with a couple of examples describing how bias affects real-world scenarios and how it can be mitigated. Each group picks two AI principle cards and discusses for 20 minutes how they can inform the candidate concept using these principles:

- Non-Maleficence: e.g., is the training data appropriate for the intended use?
- Justice: reflect on diversity, equality and inclusion
- Beneficence: consider beneficiaries of the application, whether individual users, groups or the whole society
- Autonomy: transparent communication about potential risks
- Explicability: e.g., is there any process in place to review the integrity of the AI application over time?

Finally, groups pitch their final concept design to the other groups to get their feedback. Instead of results that try to attain consensus and conclusions to solve a shared pre-defined problem, this allows us to get a more comprehensive understanding of various unique and contrasting viewpoints.

An Expert Evaluation and a follow-up study were used to test and evaluate the co-design methodology carried out during MiniCoDe Workshops. The first one involved a diverse mix of participants, including a UX Designer with industry experience, two academics with a mix of design and computer science backgrounds (e.g., machine

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

learning), a start-up consultant with financial and strategic background, an NGO director with ethical AI experience, and a developer with relevant experience using experimental research approaches.
The second was organized as a 3-hour session with a large group of first year PhD students.

Expert Evaluation

This section reports on an expert evaluation we carried out to collect initial qualitative reactions regarding the workshop by a group of experts. The aim was to collect initial feedback about the ability of MiniCoDe to make aware experts and to include in their discussion aspects associated with ethical design of emerging ML applications.

Two online MiniCoDe workshop sessions were run to gather feedback. The first included a UX Designer with Industry Experience and two academics with a mix of design and computer science backgrounds, whilst the second included as participants an academic with a design background, an NGO director with ethical AI experience, and a developer with relevant experience using experimental research approaches.

We started by first introducing MiniCoDe to participants, going over the various phases and what each entailed. The facilitator was one of the authors. Both sessions lasted about 3 hours each.

We've used a sample narrative and Inspiration Wall related to a fictional Health Insurance Service tracking metabolism and the negative side effects that it can introduce to society.

We ran the workshop with the experts as participants and gathered their informal feedback at the end.

Preliminary Findings

The six experts appreciated the experience of the overall workshop design and were invested in the whole process. Overall experts referred to MiniCoDe as a good way to provide guidelines to teams willing to investigate the impact of new technologies on society. Experts, acting as participants, reported to be able to generate interesting ideas working with others, for instance, one of the experts (E2) commented: "What would incentivise me as a business to pick this up and use it, other than people generally talking about ethical concerns and it's something I care about? But if I give a general ethical framework to a startup in my cohort they wouldn't bother to go through with it and probably think they'll get to that later, but this actually helps you think about your business model, your defensibility, robustness, if it may work".

Concurrently, another expert (E4) reported: "You're always taught about focusing on the problem first, thinking about the design part always comes in later steps, but this [workshop] could help you kill off an idea or pivot earlier, that's way more valuable to a founder.", and also: "The business model is really disconnected from all ethical choices, issues of privacy, issues of biases, debiasing, who own the data, but ultimately the business model dictates all ethical choices, and what was interesting in the workshop we kept coming back to it and that to me it's the most fundamental disconnect."

Moreover, while ethics frameworks have been criticized for 'Ethics Washing' [26] we noticed instead a different reaction from our experts. In fact, the experts appreciated the MiniCoDe pragmatic way of using cards, citing real-world examples, to operationalise ethics principles.

Finally, the main recommendation provided by the experts concerned the final product of the workshop. They've highlighted how in its current form, MiniCoDe doesn't aid participants in building an artifact at the end of it, which - being a design-oriented activity - is quite important. To this end, they've recommended we introduce a final Storyboarding phase to help participants build a concrete artifact they can use to reason upon after the workshop and to generate feedback from their peers.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

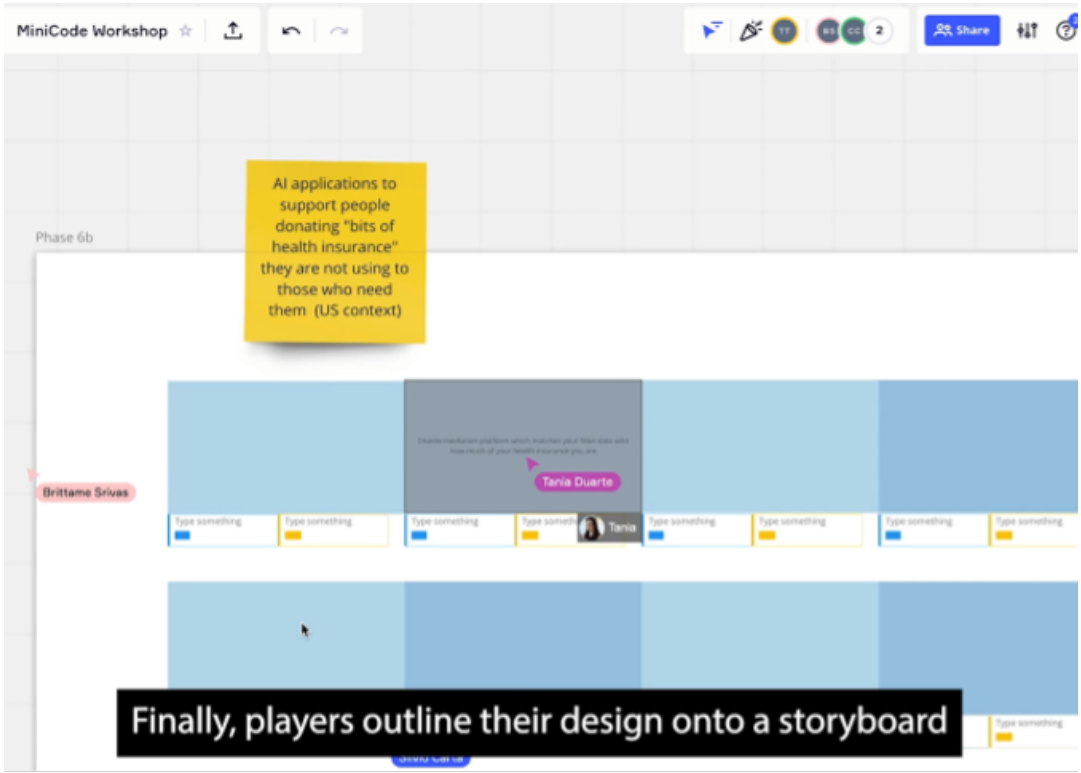

FIGURE 2. A storyboarding template helps capture the workshop insights.

A storyboard is a visual representation of a story's scene-by-scene progression. It's made up of a sequence of sketches arranged chronologically and accompanied by annotations. Storyboarding is more than just a list of the most salient information in a story. When it's time to collaborate and make critical, creative decisions, it's a method that gives team members a tangible, visual flow of a concept.

The MiniCoDe storyboard template (Figure 2) provides a simple process for creating storyboards: using a storyboarding template available in our digital material, groups can quickly build a storyboard from the notes previously produced. Alternatively, they might start with a piece of paper.

User Evaluation

This section reports on the subjective evaluation of MiniCoDe with a large cohort of participants.

Goal

The goal of this study was to evaluate whether MiniCoDe provided an engaging, useful, and collaborative way of reflecting on issues associated with algorithmic bias. The purpose is to evaluate how participants respond to a sample MiniCoDe scenario and reflect on different strategies in order to mitigate algorithmic bias.

Research Question

In response to the pervasiveness of ML-based algorithms and increasing evidence of unfairness and prejudice, new co-design methodologies are needed to assist multidisciplinary teams in designing systems that are more useful to Society.

MiniCoDe aims at combining Design Fiction with other rapid ideation techniques to create concepts and storyboards illustrating participants' reflections on ethical and social impacts of ML applications in Society, in a fun and engaging way.

The main research question derived from this context is: *"Does MiniCoDe provide an engaging and useful way of reflecting over algorithmic bias?"*.

Participants

The participants of the study were 50 first year PhD students (11 Female, 39 Male) from different universities across Italy, as part of the National PhD Program on AI

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

and Society. 37 had a STEM background, while 13 had a non-STEM degree, such as Law, or Philosophy. No prerequisite knowledge was required to attend the workshop, and only 3 participants had prior knowledge of MiniCoDe due to an introductory course on Human-AI interaction. A brief introduction to the workshop was provided to the entire group.

Context

The study took place within the University of Pisa's facilities, during a Summer School where all first-year PhD students of the National PhD Program on AI and Society participated. We carried out the workshop during the first day of the Summer School, as a first group activity for the entire class. The case study provided for the workshop concerned the AI Act[2], a proposed European law to regulate ethical AI applications. Participants were given a narrative depicting a high-risk scenario about AI-assisted courtroom decisions suffering from postcode bias and were asked to come up with mitigation strategies that could be implemented by the AI Act through the full MiniCoDe workshop.

Procedure

Participants were asked to form eight groups, six groups composed of six members each, and two groups composed of seven members. The entire workshop lasted three hours, and at the end a 9-item questionnaire was administered to the participants to assess their overall engagement during the workshop. The items, presented in a five-point Likert scale (see http://dx.doi.org/10.13140/RG.2.2.20704.76802) were extrapolated from [27]. Moreover, participants were invited to answer two open questions regarding the usefulness of the workshop to support their awareness of AI bias, and about the general usefulness of the workshop, specifically: (Q1) "Has your perspective on algorithmic bias been altered by the workshop?" and, (Q2) "What did you find most useful about the workshop?".

Results

The set of items in the questionnaire we used resulted reliable, with a Cronbach's alpha of 0.82, which is over the acceptable threshold of 0.7 [28]. On average the level of engagement reported by the participants was very high at 81.6% (SD: 9.7%). Figure 3 reports the different averages per each item of the questionnaire, suggesting that for the participants the most valued aspect of MiniCoDe for participants was providing a systematic way to foster team discussion (93.8%), while the less appreciated aspect of this methodology for the participants was related to the simplification of the concept design, even though this was still very positively evaluated (74.2%).

[2] https://artificialintelligenceact.eu/

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

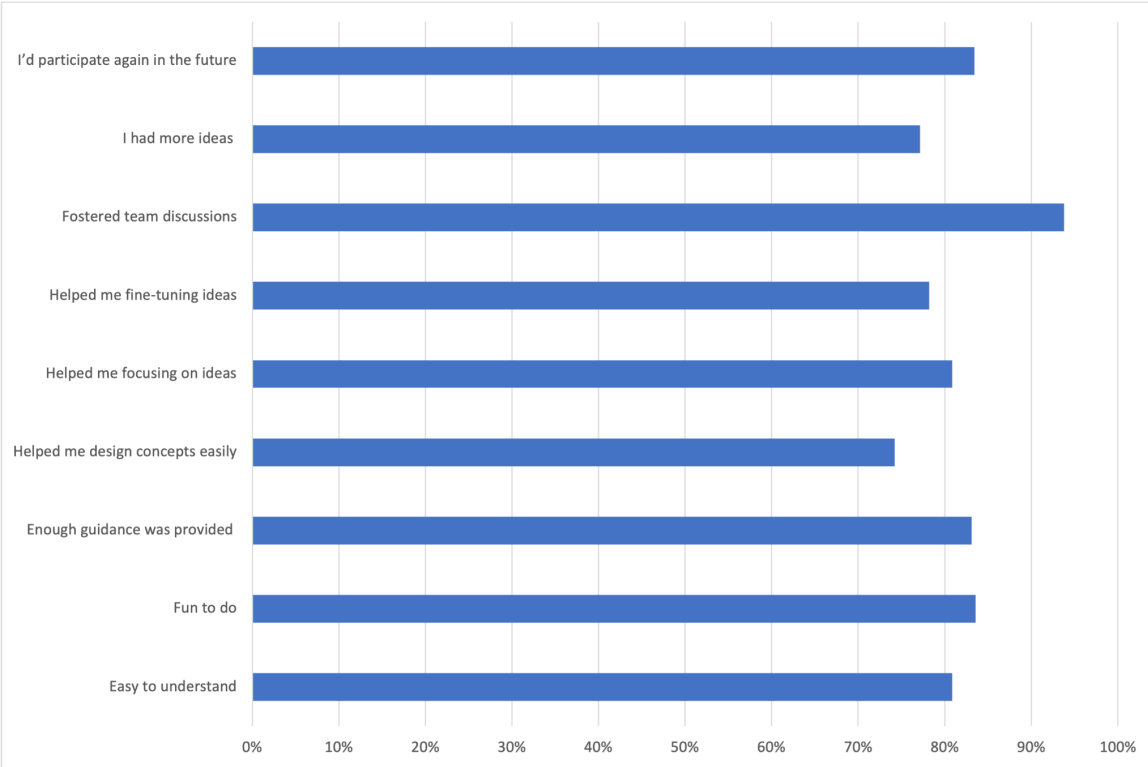

FIGURE 3. Resulting averages for all items of the questionnaire.

Regarding the usefulness of MiniCoDe as a tool to enhance awareness regarding bias (Q1) 61.2% of the participants declared that this methodology helped them to gain a new perspective and to learn more about aspects associated with algorithmic bias. Regarding the question about the general usefulness of the workshop (Q2) a thematic analysis of the answers of the participants suggested three main themes: (i) Group work for generating ideas. 55.1% of the participants declared that the workshop helped them to understand the importance of teamwork and of exchanging ideas when it comes to dealing with complex systemic topics. For instance, a participant suggested that the workshops made clear the importance of a "Discussion between group members to foster thinking about issues that are directly connected to my research topic" (P6). (ii) Systematic approach to deal with issues. 34.7% of the participants recognised that the systematic approaches used in the workshops are useful to deal with complex issues e.g., the use of "deck of cards is very interesting, actually I've discussed about new stuff that are far from my field of study" (P19). (iii) Legal issues associated with AI. 6.1% of the participants admitted that they never reflected before on law and legal issues associated with AI e.g., "It forced us to think on law enforcement issues that it's not common for me. An interesting perspective is the necessity of a cycle between laymen and legal experts about how algorithmic biases influence normal people's lives" (P49). Finally, the remaining participants confirmed the usefulness of the workshop talking about this as an approach to help people to freely express ideas, or as a sort of gamified way of dealing with complexity e.g., "Graphical support instruments and gamification can be used to foster discussion and debate" (P35).

Discussion

The main research question of this study was to investigate how MiniCoDe supports reflections over algorithmic bias.

As suggested in [29], engagement is a critical element to be considered when designing workshops, influencing both the effectiveness and the lasting impact on participants. Our study's results corroborate this view, demonstrating a significant level of engagement through both the questionnaire responses and the open-ended feedback gathered at the conclusion of our workshop. In line with this outcome, it seems that the gamification

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

strategy we employed paid off i.e., involved a clear set of rules and a variety of card decks designed to stimulate discussion. Since the initial versions of our workshop, we worked on improving our approach by including a board game-like experience for multidisciplinary teams. Moving forward, we will continue to leverage this game-like approach, recognizing its effectiveness in fostering engagement. This approach not only aligns with the insights from [29] regarding the importance of engagement in workshops but also sets a benchmark for further investigations.

The utmost important set of results we can gain from our study closely relates to the main Research Question we posed at the beginning: *"Does MiniCoDe provide an engaging and useful way of reflecting over algorithmic bias?"*

The majority of participants seem to have developed a new perspective on Algorithmic Bias thanks to MiniCoDe (Q1). This, together with the relative positive reception around the mechanisms in place to simplify the concept designs, gives us an indication that MiniCoDe can aid in reflecting over algorithmic bias at design time, even though more research is needed to identify how these reflections are embedded in the final designs.

Finally, the selected theme around which the workshop revolved prompted participants to reflect on issues far from their usual field of study (Q2), which is indeed evidence of how MiniCoDe can gather insights across multiple disciplines and combine expertise that are essential to reason over complicated societal impacts of new technologies [2, 14, 18].

Limitations

While participants generate the outcomes, we recognise that facilitators always influence the workshop. In this instance, one of the authors was moderating, but in the future we'll test the workshop with different facilitators in order to evaluate its robustness.

As we highlighted in the discussion, our aim is to develop this methodology further and package it into a toolkit that we can provide any small team to run it themselves. This will have to be carefully analyzed and tested for the generalizability of the methodology.

In the current study, groups were formed autonomously, which might have generated inner biases by skewing some of the groups and limited the general validity of the results. Moreover, participants were mostly males, novices, and from an academic background, thus more studies are needed in order to test the workshop with expert participants with different backgrounds to generalize our findings.

Finally, the workshops are demanding to facilitate; the workshops require a firm commitment from the participants at the outset. It is challenging to maintain a fully open flow structure as both the facilitator and the participants become invested in the outcomes. Insights can be elusive and challenging to capture. To mitigate such issues, we noticed that the materials positively stimulate participants' engagement and the storyboarding phase helps capture the workshop insights.

Conclusions

The use of ML algorithms in decision-making has the potential to lead to biased and unfair outcomes if proper attention is not paid to issues of social justice. To address this, in this paper we've presented MiniCoDe, a design fiction-driven workshop methodology aimed at assisting in the ethical design of machine learning applications. By using scenario-based design and prototyping, MiniCoDe allows participants to explore potential bias and reflect on mitigation strategies. Through this innovative approach, we hope to enable a broad spectrum of knowledge about potential bias to emerge since early stages of design, and encourage more reflective practices in the design of ML applications. By bringing together multi-disciplinary teams and facilitating intense, workshop-like gatherings, we aim to create a space for the emergence of potential ethical issues and the development of strategies to mitigate bias.

Firstly, we carried out an expert evaluation carrying out a pilot workshop with two groups of experts, coming from both Academia and Industry, with a mixed background of AI, UX, and Ethics. They tested the methodology and offered their feedback, which we've included in the following iteration of MiniCoDe.

Secondly, we run a MiniCoDe workshop with 50 first-year PhD students in an AI program, issuing a follow-up questionnaire based on reliable items and two open-ended questions to check perceived usefulness and engagement with MiniCoDe. The findings reported a high positive attitude of the participants (in both dimensions) toward the methodology, also suggesting a positive effect

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

of the workshop on the development of new perspectives over algorithmic bias and linked issues.

In conclusion, our work suggests a high perceived potential value of MiniCoDe and Design Fiction as a method to reduce algorithmic bias in co-design activities and promote algorithmic social justice.

In the future we will further develop the methodology, packaging it in a toolkit containing materials that we can distribute and enable small teams to run MiniCoDe workshops on their own, without the need of an expert facilitator.

Appendix I: Story

Mark is a freshman at Harvard in Cambridge, Massachusetts. Like most freshmen, he is short of money and therefore begins mining cryptocurrencies; after all, it is 2037, and the "real" action is happening in Metaverse 2.0. Three long years passed; he is not a freshman anymore. He has been saving all this time to organise a dinner at the exclusive "The Distracted Globe", a famous zero-gravity dance club, to re-enact the scene from "Ready Player One" and dance floating in virtual air with his girlfriend Clarissa; just the perfect night to propose: New Year's Eve, 1st Jan 2040.

Sitting in his dark room on the 31st Dec 2039, with only a table light on, he is staring at his 180-degrees curved monitor, ready to tap on the virtual table tableau and book the dinner of a lifetime. The Distracted Globe is so exclusive that it only allows booking a minute before the event. Clarissa and Mark are in the waiting room, but time is relative in the Metaverse, where milliseconds are considered slow. The moment the transaction is accepted, their avatars will be teleported on the dancing floor.

$3 Billion cryptocurrency and three years of NFT tokens mining worth are gone in an instant. His heart skips a beat waiting for the transaction to be accepted; The Distracted Globe is so unique, Mark and Clarissa are going to tell their children about that unforgettable night! The screen returns an error: "not enough NFT tokens", a message blinks on the screen, CNET news – 1st Jan 2040, 0:01 AM: "NFTs tokens crisis on New Year's Eve, your tokens are worthless! Burst trillions of dollars in a picosecond!".

Appendix II: Design Brief

Concept – A digital service for cryptocurrency decision-making.

You are asked to develop a concept for a digital service using Virtual Assistant Technology (e.g. Chatbots or voice-activated apps like Alexa or Siri) to teach and explain cryptocurrency concepts to the layperson and assist in decision making. Like the classic "angel on your shoulder" metaphor, Virtual Assistants will form a companionship bond with the consumer suggesting the right actions to take according to consumers' literacy about cryptocurrency.

Your proposal should be highly imaginative and take account of current and near-future, cutting edge approaches to Virtual Assistants (refer to the inspirational wall), sensing and artificial intelligence or more advanced technologies such as Brain-Computer Interfaces or the Metaverse (characters like a butler or librarian as in Ready Player One).

You are asked to adopt a systemic approach to this project, which considers the challenges involved in digital services for cryptocurrency decision-making. Please think about an engaging, inclusive and meaningful experience for the target audience which encourages further technology adoption. Your Virtual Assistant aims to foster the consumers' awareness of making decisions relating to cryptocurrencies and what potential consequences such behavior might have on their lives and society.

Furthermore, can we imagine scenarios where people might want to switch it off? Why?

Appendix III: Timeline

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

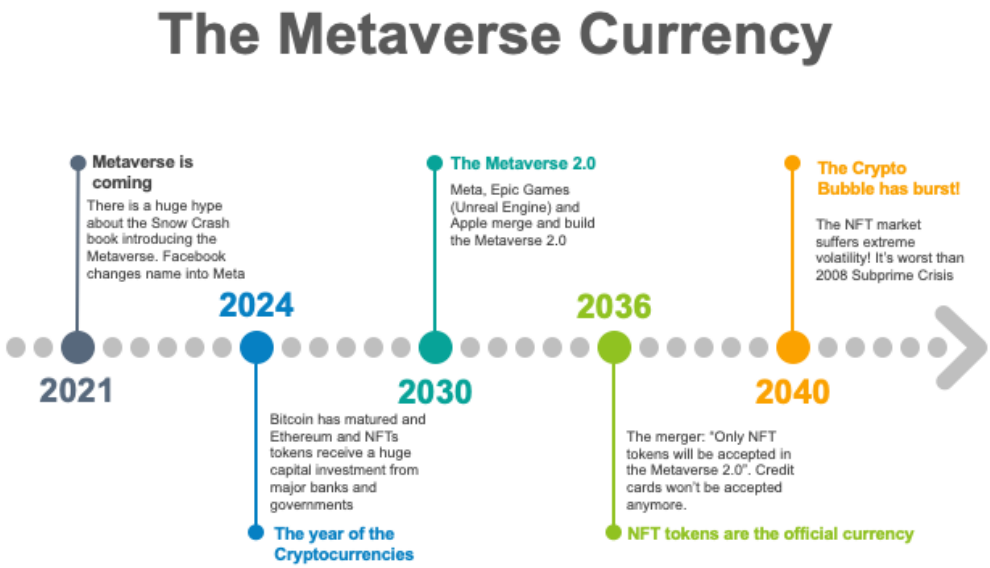

FIGURE 4. Example of the inspiration wall's timeline on using future ML applications to mine cryptocurrency in the Metaverse

ACKNOWLEDGMENTS

We thank Prof. Silvio Carta, Tania Duarte, and all participants in our workshops. This work was supported by the EPSRC NetworkPlus Not-Equal Program as part of the research project MiniCode (Project Reference NE2.001).

REFERENCES

1. Angerschmid, A., Zhou, J., Theuermann, K., Chen, F. & Holzinger, A. 2022. Fairness and Explanation in AI-Informed Decision Making. Machine Learning and Knowledge Extraction, 4, (2), 556--579, doi:10.3390/make4020026
2. Malizia, A., Carta, S., Turchi, T., & Crivellaro, C. 2022. MiniCoDe Workshops: Minimise Algorithmic Bias in Collaborative Decision Making with Design Fiction. In Proceedings of the Hybrid Human Artificial Intelligence Conference.
3. Caliskan, A., Bryson, J. J., & Narayanan, A. 2017. Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), 183-186G.
4. Bill Moggridge and Bill Atkinson. 2007. Designing interactions. Vol. 14. MIT press Cambridge, MA.
5. Pelle Ehn. 1993. Scandinavian design: On participation and skill. Participatory design: Principles and practices (1993), 41–77.
6. Marcel Bogers, Allan Afuah, and Bettina Bastian. 2010. Users as innovators: a review, critique, and future research directions. Journal of management (2010).
7. Karvonen, H., Koskinen, H., & Haggrén, J. (2012). Defining user experience goals for future concepts: A case study. In 7th Nordic Conference on Human-Computer Interaction, NordiCHI2012, UX Goals Workshop (pp. 14-19). Tampere University of Technology.
8. Ballard, S., Chappell, K. M., & Kennedy, K. (2019, June). Judgment call the game: Using value sensitive design and design fiction to surface ethical concerns related to technology. In Proceedings of the 2019 on Designing Interactive Systems Conference (pp. 421-433).
9. Friedman, B., & Hendry, D. 2012. The envisioning cards: a toolkit for catalyzing humanistic and technical imaginations. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1145-1148). https://doi.org/10.1145/2207676.2208562
10. Flanagan, M., & Nissenbaum, H. 2007. A game design methodology to incorporate social activist themes. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 181-190). https://doi.org/10.1145/1240624.1240654
11. Skirpan, M. W., Cameron, J., & Yeh, T. 2018. More than a show: Using personalized immersive theater to educate and engage the public in technology ethics. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-13). https://doi.org/10.1145/3173574.3174038
12. Muller, M., Bardzell, J., Cheon, E., Su, N. M., Baumer, E. P., Fiesler, C., ... & Blythe, M. (2020, April). Understanding the Past, Present, and Future of Design Fictions. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-8).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Intelligent Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MIS.2024.3352977

13. Grimshaw, P., & Burgess, T. F. (2014). The emergence of 'zygotics': using science fiction to examine the future of design prototyping. Technological Forecasting and Social Change, 84, 5-14.
14. Malizia, A. and Carta, S., 2019. Science fiction could save us from bad technology. The Conversation.
15. Craigon, P.J., Sacks, J., Brewer, S., Frey, J., Gutierrez, A., Jacobs, N., Kanza, S., Manning, L., Munday, S., Wintour, A. and Pearson, S., 2023. Ethics by design: Responsible research & innovation for AI in the food sector. Journal of Responsible Technology, 13, p.100051.
16. Rezwana, J. and Maher, M.L., 2023, June. User Perspectives on Ethical Challenges in Human-AI Co-Creativity: A Design Fiction Study. In Proceedings of the 15th Conference on Creativity and Cognition (pp. 62-74).
17. Rovatsos, M., Mittelstadt, B., & Koene, A. (2019). Landscape Summary: Bias In Algorithmic Decision-Making: what is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it?.
18. Malizia, A. (2019). Design Fictions to Mitigate Social Injustice in Possible Futures. Blog@Ubiquity, ACM.
19. Shaw, C., & Corner, A. (2017). Using Narrative Workshops to socialise the climate debate: Lessons from two case studies–centre-right audiences and the Scottish public. Energy Research & Social Science, 31, 273-283.
20. Johnson, B.D. 2011, Science Fiction Prototyping: Designing the Future with Science Fiction, 1st edn, Morgan & Claypool Publishers, US.
21. Kleinen, J., & Kurz, L. (2021). Exploring New Technology's Meaning for a Sustainable Future via Collaborative Science-Fiction Prototyping: A Novel Method for the Engineering Curriculum. Universities, Sustainability and Society: Supporting the Implementation of the Sustainable Development Goals, 335.
22. Kudrowitz, B.M. and Wallace, D., 2013. Assessing the quality of ideas from prolific, early-stage product ideation. Journal of Engineering Design, 24(2), pp.120-139.
23. Candy, S., & Watson, J. (2015). The thing from the future. The APF Methods Anthology. London: Association of Professional Futurists, 18-21.
24. Floridi, L., & Cowls, J. (2021). A unified framework of five principles for AI in society. In Ethics, Governance, and Policies in Artificial Intelligence (pp. 5-17). Springer, Cham.
25. Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI Ethics. Minds and Machines, 1-18.
26. McMillan, D., & Brown, B. (2019, November). Against ethical AI. In Proceedings of the Halfway to the Future Symposium 2019 (pp. 1-3).
27. Mora, S., Gianni, F., Nichele, S., & Divitini, M. (2018). Introducing IoT Competencies to First-Year University Students With The Tiles Toolkit. In Proceedings of the 7th Computer Science Education Research Conference (CSERC '18) (pp. 26–34).
28. Shemwell, J.T., Chase, C.C. and Schwartz, D.L. 2015. Seeking the general explanation: A test of inductive activities for learning and transfer. J Res Sci Teach, 52: 58-83. https://doi.org/10.1002/tea.21185
29. Schelle, K.J., Gubenko, E., Kreymer, R., Naranjo, C.G., Tetteroo, D. and Soute, I.A., 2015. Increasing engagement in workshops: designing a toolkit using lean design thinking. In Proceedings of the Mulitimedia, Interaction, Design and Innnovation (pp. 1-8).

Tommaso Turchi, is an Assistant Professor at the Department of Computer Science, University of Pisa, Italy. His current research interests include Human-AI Interaction, End-User Development, and Cyber-physical Systems. He received the Ph.D. degree in Computer Science from Brunel University London, UK. Contact him at tommaso.turchi@unipi.it.

Alessio Malizia, is an Associate Professor at the Department of Computer Science, University of Pisa, Italy. His current research interests include Human-Centred AI and Design Fictions. He received the Ph.D. degree in physics from University. He is a Fellow of the IEEE Computer Society and distinguished speaker of the ACM. Contact him at alessio.malizia@unipi.it.

Simone Borsci, is an Associate Professor of Human Factors and Cognitive Ergonomics at University of Twente. His current research interests include User Experience, Social Science, and Biomedical Technology Assessment. He received the Ph.D. degree in cognitive psychology from the University of Rome "La Sapienza", Italy. He was recently nominated Honorary Senior Fellow of Human Factor for Health Technology at Imperial College. Contact him at s.borsci@utwente.nl.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/