

# Reflecting on Algorithmic Bias With Design Fiction: The MiniCoDe Workshops

Tommaso Turchi <sup>1</sup> and Alessio Malizia <sup>2</sup>, University of Pisa, 56126, Pisa, Italy

Simone Borsci <sup>3</sup>, University of Twente, 7500 AE, Enschede, The Netherlands

*In an increasingly complex everyday life, algorithms—often learned from data, i.e., machine learning (ML)—are used to make or assist with 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 multidisciplinary teams to perform intense, workshop-like gatherings to let potential ethical issues emerge 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 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<sup>1</sup> 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 minimize algorithmic bias in collaborative decision making with design fiction (MiniCoDe), a new board-game-like workshop methodology aimed at assisting with the ethical design of upcoming technologies that will become ingrained in daily life.

The present work extends our previous research,<sup>2</sup> which introduced the concept and preliminary structure of MiniCoDe. The previous work laid the groundwork for this methodology, focusing on its inception and theoretical 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 include 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 multidisciplinary teams addressing algorithmic bias.

---

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>  
Digital Object Identifier 10.1109/MIS.2024.3352977  
Date of publication 11 January 2024; date of current version 10 April 2024.

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 to encourage conversations about the potential for bias. It is intended for multidisciplinary 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, and nongovernmental organizations (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 ML in the coming decade. Nevertheless, ML algorithms will almost certainly harbor some form of implicit bias. For example, the academic article by Caliskan et al.,<sup>3</sup> “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 ML applications provide evidence that the word-associating agent’s flawed strategy may be used to train a resume analyzer 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 they be used to support the ethical design of those emerging AI-based services that will be impacting everyday life?

## RELATED WORKS

Workshops have played an essential role in human-computer interaction 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,<sup>4</sup> user participation in the design process,<sup>5</sup> user experiences,<sup>6,7</sup> and design fiction,<sup>8</sup> 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<sup>9</sup> 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 methodology<sup>10</sup> proposes a framework for incorporating activist social themes in game design, providing a tested methodology 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”<sup>11</sup> combined elements of design fiction and user enactments to construct an immersive theater 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,<sup>12</sup> 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 ML experts in identifying potential bias and considering mitigation solutions. Design fiction prototypes<sup>13</sup> 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<sup>14</sup> for talking about the future implications, repercussions, and effects of technology.

The recent literature has highlighted the importance of ethical considerations in technology and AI applications. Craigon et al.<sup>15</sup> 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<sup>16</sup> 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 an 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 recognized as a relevant issue in ML applications. For example, IEEE and the International Organization for Standardization (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 that humans override automated decisions where necessary, dealing with situations in which machines would struggle.<sup>17</sup> 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.<sup>18</sup>

## 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 have 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,”<sup>24</sup> and a deck of MiniCoDe ethics cards (Figure 1).

The digital version of the workshop was developed on Miro,<sup>a</sup> an online collaboration platform supporting remote and simultaneous access and editing.

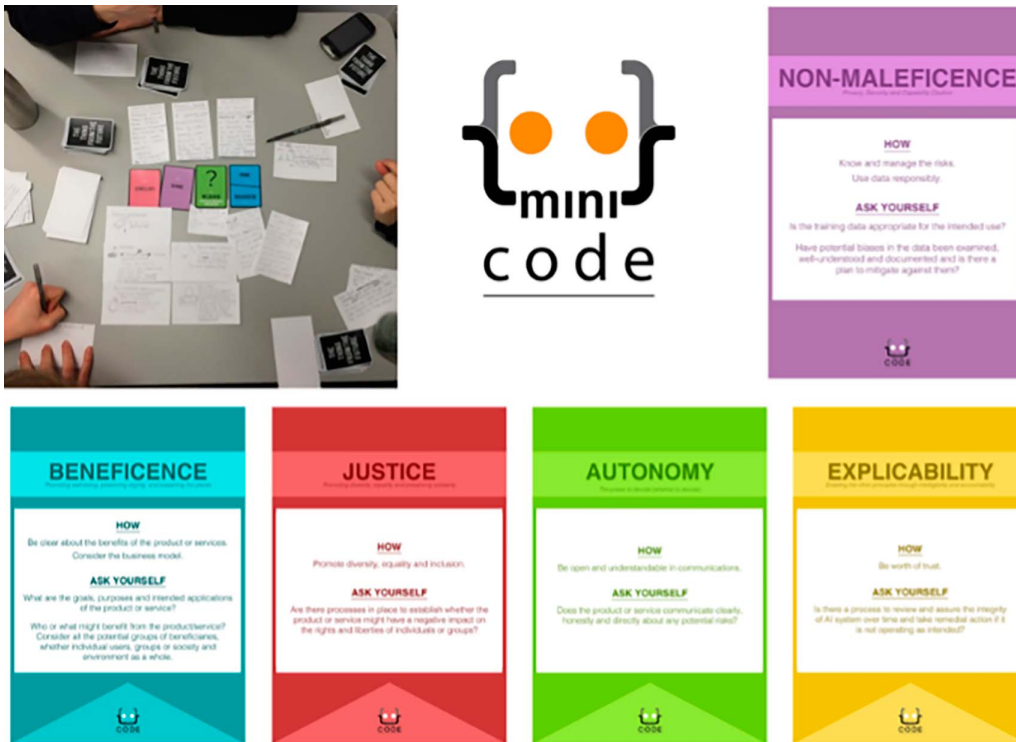
In the following, we describe each workshop phase in detail. The four distinct phases underlying our workshop approach are prepare, ideate, refine, and reflect.

### 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 three-structure narrative (challenge, climax, and ending),<sup>19</sup> 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<sup>20</sup>; 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 three elements: a brief story, a design brief, and a fictional timeline to help participants focus on the task at hand. (See Appendixes 1, 2, and 3, which present 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 appendixes are available at <https://doi.org/10.1109/MIS.2024.3352977>.)

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<sup>21</sup> are reflected in this technique. This might be considered Act I of a larger

<sup>a</sup><https://miro.com>



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

story that the participants were to compose later. This provides a narrative for 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 min to generate six to eight ideas related to the given prompt and write them on Post-It notes (digital the 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.<sup>22</sup> 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 having confident participants

who say anything that comes to mind 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 min, 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.<sup>23</sup> 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 to use only the terrain, object, and mood cards in our workshop. We discarded the arc cards since those are about imagining a future scenario, which was 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 min 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 min 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 and Cows. <sup>24</sup> We reported multiple levels of detail for each concept, <sup>25</sup> 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 min how they can inform the candidate concept using these principles:

- › *Nonmaleficence*: e.g., are 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 predefined 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 user experience (UX) designer with industry experience, two academics with a mix of design and computer science backgrounds (e.g., ML), a start-up consultant with a

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-h session with a large group of first-year Ph.D. 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 experts aware of and include in their discussion aspects associated with the 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, while 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 h each.

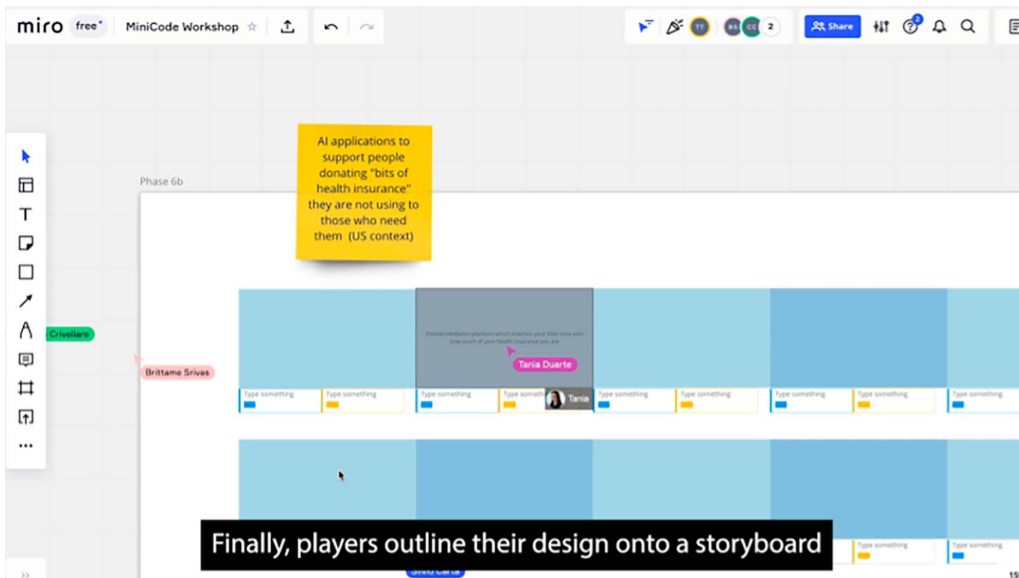
We have used a sample narrative and inspiration wall related to a fictional health insurance service tracking metabolism and the negative side effects that it can introduce to society.

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 being able to generate interesting ideas working with others; for instance, one of the experts (E2) commented,

“What would incentivize 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 start-up 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.”



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

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, [and] who own the data, but ultimately the business model dictates all ethical choices, and what was interesting in the workshop [was that] 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,”<sup>26</sup> 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 operationalize ethics principles.

Finally, the main recommendation provided by the experts concerned the final product of the workshop. They highlighted how, in its current form, MiniCoDe does not aid participants in building an artifact at the end of it, which—as this is a design-oriented activity—is quite important. To this end, they recommended that 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.

A storyboard is a visual representation of a story’s scene-by-scene progression. It is 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 is time to collaborate and make critical, creative decisions, it is 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 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 the 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 on algorithmic bias?"

## Participants

The participants in the study were 50 first-year Ph.D. students (11 female and 39 male) from different universities across Italy, as part of the National Ph.D. Program on AI and Society. Of these, 37 had a science, technology, engineering, and mathematics (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 three 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 in which all first-year Ph.D. students of the National Ph.D. Program on AI and Society participated. We carried out the workshop during the first day of the summer school as the first group activity for the entire class. The case study provided for the workshop concerned the AI Act,<sup>b</sup> 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 3 h, and, at the end, a nine-item questionnaire was administered to the participants to assess their overall engagement during the workshop. The items, presented in a five-point Likert scale (available at <http://dx.doi.org/10.13140/RG.2.2.20704.76802>) were extrapolated from Mora et al.<sup>27</sup> 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?"

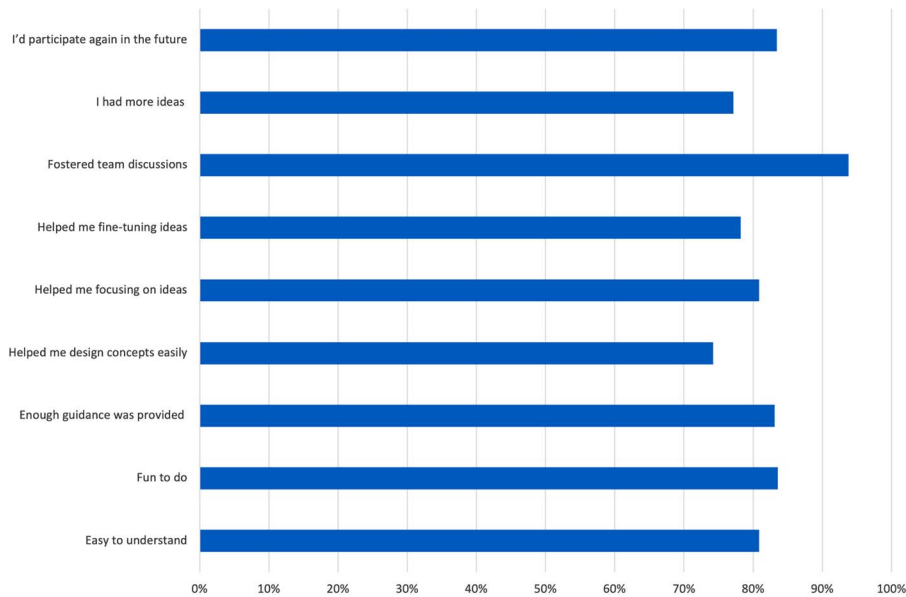
## Results

The set of items in the questionnaire we used resulted in reliability, with a Cronbach's alpha of 0.82, which is over the acceptable threshold of 0.7.<sup>28</sup> On average, the level of engagement reported by the participants was very high at 81.6% (standard deviation: 9.7%). [Figure 3](#) reports the different averages for each item of the questionnaire, suggesting that 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 participants was related to the simplification of the concept design, even though this was still very positively evaluated (74.2%).

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 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:

- ▶ *Group work for generating ideas:* Overall, 55.1% of 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).
- ▶ *Systematic approach to deal with issues:* A total of 34.7% of the participants recognized that the systematic approaches used in the workshops are useful to deal with complex issues; e.g., the use of the "deck of cards is very interesting; actually, I've discussed about new stuff that are far from my field of study" (P19).
- ▶ *Legal issues associated with AI:* Overall, 6.1% of the participants admitted that they had never reflected before on law and legal issues associated with AI: e.g., "It forced us to think on law enforcement issues that is not common for me. An interesting perspective is the necessity of a cycle between laymen and legal experts about

<sup>b</sup><https://artificialintelligenceact.eu/>



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

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 about algorithmic bias.

As suggested by Schelle et al.,<sup>29</sup> 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 strategy we employed paid off—i.e., involving a clear set of rules and a variety of card decks designed to stimulate discussion. Since the initial versions of our workshop, we have 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 Schelle et al.<sup>29</sup> regarding the importance of engagement in workshops but also sets a benchmark for further investigations.

The most 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 on algorithmic bias?"

The majority of participants seem to have developed a new perspective on algorithmic bias thanks to MiniCoDe (Q1). This, together with the relatively positive reception around the mechanisms in place to simplify the concept designs, gives us an indication that MiniCoDe can aid in reflecting on 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, which are essential to reasoning about the complicated societal impacts of new technologies.<sup>2,14,18</sup>

## Limitations

While participants generate the outcomes, we recognize that facilitators always influence the workshop. In this instance, one of the authors was moderating, but, in the future, we will test the workshop with different facilitators 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 for any small team to run it itself. 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 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 that the storyboarding phase helps capture the workshop insights.

## CONCLUSION

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 article, we have presented MiniCoDe, a design-fiction-driven workshop methodology aimed at assisting in the ethical design of ML 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 from the early stages of design and encourage more reflective practices in the design of ML applications. By bringing together multidisciplinary 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.

First, we performed 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 have included in the following iteration of MiniCoDe.

Second, we ran a MiniCoDe workshop with 50 first-year Ph.D. 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 highly positive attitude of the participants (in both dimensions) toward the methodology, also suggesting a positive effect of the workshop on the development of new perspectives regarding 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 enabling small teams to run MiniCoDe workshops on their own, without the need of an expert facilitator.

## ACKNOWLEDGMENTS

We thank Prof. Silvio Carta, Tania Duarte, and all participants in our workshops. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) NetworkPlus Not-Equal Program as part of the research project MiniCode (Project Reference NE2.001). This research was also partly sponsored by Piano Nazionale di Ripresa e Resilienza (PNRR), M4C2–Investimento 1.3, Partenariato Esteso PE00000013, “FAIR - Future Artificial Intelligence Research”, Spoke 1 “Human-Centered AI,” funded by the European Commission under the NextGeneration EU program.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval. This article has supplementary downloadable material available at <https://doi.org/10.1109/MIS.2024.3352977>, provided by the authors.

## REFERENCES

1. A. Angerschmid, J. Zhou, K. Theuermann, F. Chen, and A. Holzinger, “Fairness and explanation in AI-informed decision making,” *Mach. Learn. Knowl. Extraction*, vol. 4, no. 2, pp. 556–579, 2022, doi: [10.3390/make4020026](https://doi.org/10.3390/make4020026).
2. A. Malizia, S. Carta, T. Turchi, and C. Crivellaro, “MiniCoDe workshops: Minimise algorithmic bias in collaborative decision making with design fiction,” in *Proc. Hybrid Human Artif. Intell. Conf.*, 2022.
3. A. Caliskan, J. J. Bryson, and A. Narayanan, “Semantics derived automatically from language corpora contain human-like biases,” *Science*, vol. 356, no. 6334, pp. 183–186, Apr. 2017, doi: [10.1126/science.aal4230](https://doi.org/10.1126/science.aal4230).
4. B. Moggridge and B. Atkinson, *Designing Interactions*, vol. 14. Cambridge, MA, USA: MIT Press, 2007.

5. E. Pelle, "Scandinavian design: On participation and skill," in *Participatory Design: Principles and Practices*. Hillsdale, NJ, USA: Lawrence Erlbaum, 1993, pp. 41–77.
6. M. Bogers, A. Afuah, and B. Bastian, "Users as innovators: A review, critique, and future research directions," *J. Manage.*, vol. 36, no. 4, pp. 857–875, 2010, doi: [10.1177/0149206309353944](https://doi.org/10.1177/0149206309353944).
7. H. Karvonen, H. Koskinen, and J. Haggren, "Defining user experience goals for future concepts: A case study," in *Proc. 7th Nordic Conf. Human-Comput. Interact., UX Goals Workshop*, Tampere University of Technology, Tampere, Finland, 2012, pp. 14–19.
8. S. Ballard, K. M. Chappell, and K. Kennedy, "Judgment call the game: Using value sensitive design and design fiction to surface ethical concerns related to technology," in *Proc. Designing Interactive Syst. Conf.*, Jun. 2019, pp. 421–433.
9. B. Friedman and D. Hendry, "The envisioning cards: A toolkit for catalyzing humanistic and technical imaginations," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2012, pp. 1145–1148, doi: [10.1145/2207676.2208562](https://doi.org/10.1145/2207676.2208562).
10. M. Flanagan and H. Nissenbaum, "A game design methodology to incorporate social activist themes," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2007, pp. 181–190, doi: [10.1145/1240624.1240654](https://doi.org/10.1145/1240624.1240654).
11. M. W. Skirpan, J. Cameron, and T. Yeh, "More than a show: Using personalized immersive theater to educate and engage the public in technology ethics," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2018, pp. 1–13, doi: [10.1145/3173574.3174038](https://doi.org/10.1145/3173574.3174038).
12. M. Muller et al., "Understanding the past, present, and future of design fictions," in *Proc. Extended Abstr. CHI Conf. Human Factors Comput. Syst.*, Apr. 2020, pp. 1–8, doi: [10.1145/3334480.3375168](https://doi.org/10.1145/3334480.3375168).
13. P. Grimshaw and T. F. Burgess, "The emergence of 'zygotics': Using science fiction to examine the future of design prototyping," *Technological Forecasting Social Change*, vol. 84, pp. 5–14, May 2014, doi: [10.1016/j.techfore.2013.08.004](https://doi.org/10.1016/j.techfore.2013.08.004).
14. A. Malizia and S. Carta. "Science fiction could save us from bad technology." The Conversation. [Accessed: Jan. 19, 2024. Online]. Available: <https://theconversation.com/science-fiction-could-save-us-from-bad-technology-123418>
15. P. J. Craigan et al., "Ethics by design: Responsible research and innovation for AI in the food sector," *J. Responsible Technol.*, vol. 13, Apr. 2023, Art. no. 100051, doi: [10.1016/j.jrt.2022.100051](https://doi.org/10.1016/j.jrt.2022.100051).
16. J. Rezwana and M. L. Maher, "User perspectives on ethical challenges in human-AI co-creativity: A design fiction study," in *Proc. 15th Conf. Creativity Cognition*, Jun. 2023, pp. 62–74, doi: [10.1145/3591196.3593364](https://doi.org/10.1145/3591196.3593364).
17. M. Rovatsos, B. Mittelstadt, and A. Koene. "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?" UK Government. [Online]. Available: <http://lnhttps://www.gov.uk/government/publications/landscape-summaries-commissioned-by-the-centre-for-data-ethics-and-innovation>
18. A. Malizia. "Design fictions to mitigate social injustice in possible futures." ACM. [Online]. Available: <https://blog.ubiquity.acm.org/design-fictions-to-mitigate-social-injustice-in-possible-futures/>
19. C. Shaw and A. Corner, "Using Narrative Workshops to socialise the climate debate: Lessons from two case studies—Centre-right audiences and the Scottish public," *Energy Res. Social Sci.*, vol. 31, pp. 273–283, Sep. 2017, doi: [10.1016/j.erss.2017.06.029](https://doi.org/10.1016/j.erss.2017.06.029).
20. B. D. Johnson, *Science Fiction Prototyping: Designing the Future with Science Fiction*, 1st ed. San Rafael, CA, USA: Morgan & Claypool, 2011.
21. J. Kleinen and L. Kurz, "Exploring new technology's meaning for a sustainable future via collaborative science-fiction prototyping: A novel method for the engineering curriculum," in *Universities, Sustainability and Society: Supporting the Implementation of the Sustainable Development Goals*, vol. 335. Cham, Switzerland: Springer-Verlag, 2021, pp. 335–348.
22. B. M. Kudrowitz and D. Wallace, "Assessing the quality of ideas from prolific, early-stage product ideation," *J. Eng. Des.*, vol. 24, no. 2, pp. 120–139, 2013, doi: [10.1080/09544828.2012.676633](https://doi.org/10.1080/09544828.2012.676633).
23. S. Candy and J. Watson, *The Thing from the Future. The APF Methods Anthology*. London, U.K.: Association of Professional Futurists, 2015, pp. 18–21.
24. L. Floridi and J. Cowls, "A unified framework of five principles for AI in society," in *Ethics, Governance, and Policies in Artificial Intelligence*. Cham, Switzerland: Springer-Verlag, 2021, pp. 5–17.
25. J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. Mökander, and L. Floridi, "Ethics as a service: A pragmatic operationalisation of AI ethics," *Minds Mach.*, vol. 31, no. 2, pp. 239–256, Jun. 2021, doi: [10.1007/s11023-021-09563-w](https://doi.org/10.1007/s11023-021-09563-w).
26. D. McMillan and B. Brown, "Against ethical AI," in *Proc. Halfway Future Symp.*, Nov. 2019, pp. 1–3, doi: [10.1145/3363384.3363393](https://doi.org/10.1145/3363384.3363393).
27. S. Mora, F. Gianni, S. Nichele, and M. Divitini, "Introducing IoT competencies to first-year university students with the tiles toolkit," in *Proc. 7th Comput. Sci. Educ. Res. Conf. (CSERC)*, 2018, pp. 26–34, doi: [10.1145/3289406.3289410](https://doi.org/10.1145/3289406.3289410).

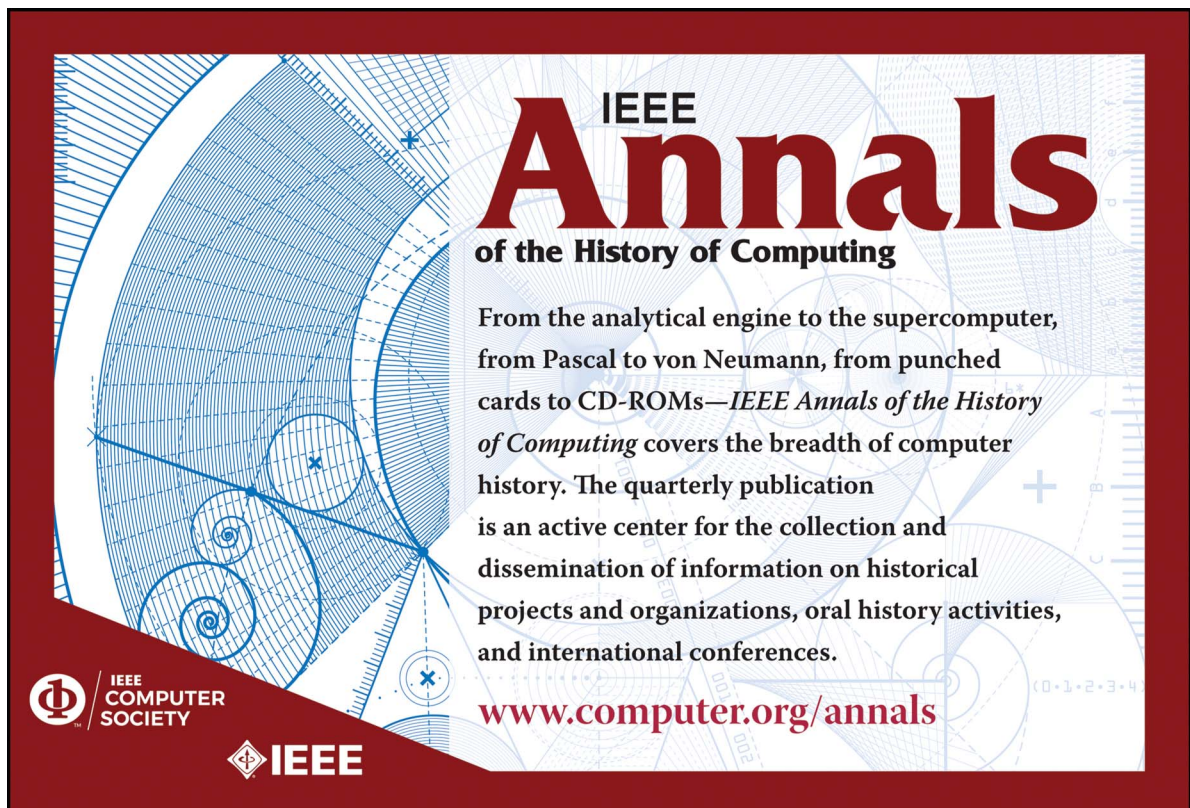
28. J. T. Shemwell, C. C. Chase, and D. L. Schwartz, "Seeking the general explanation: A test of inductive activities for learning and transfer," *J. Res. Sci. Teaching*, vol. 52, no. 1, pp. 58–83, Jan. 2015, doi: [10.1002/tea.21185](https://doi.org/10.1002/tea.21185).
29. K. J. Schelle, E. Gubenko, R. Kreymer, C. G. Naranjo, D. Tetteroo, and I. A. Soute, "Increasing engagement in workshops: Designing a toolkit using lean design thinking," in *Proc. Multimedia, Interact., Des. Innov.*, 2015, pp. 1–8, doi: [10.1145/2814464.2814481](https://doi.org/10.1145/2814464.2814481).

**TOMMASO TURCHI** is an assistant professor with the Department of Computer Science, University of Pisa, 56126, Pisa, Italy. His research interests include human–AI interaction, end-user development, and cyberphysical systems. Turchi received his Ph.D. degree in computer science from Brunel University London, U.K. Contact him at [tommaso.turchi@unipi.it](mailto:tommaso.turchi@unipi.it).

**ALESSIO MALIZIA** is an associate professor with the Department of Computer Science, University of Pisa, 56126, Pisa, Italy. His research interests include human-centered AI and design fictions. Malizia received his Ph.D. degree in physics from Sapienza University of Rome. He is a Fellow of the IEEE Computer Society and distinguished speaker of the Association for Computing Machinery. Contact him at [alessio.malizia@unipi.it](mailto:alessio.malizia@unipi.it).

**SIMONE BORSCHI** is an associate professor of human factors and cognitive ergonomics at the University of Twente, 7500 AE, Enschede, The Netherlands. His research interests include user experience, social science, and biomedical technology assessment. Borschi received his Ph.D. degree in cognitive psychology from the University of Rome "La Sapienza," Italy. He was recently nominated as an honorary senior fellow of human factors for health technology at Imperial College. Contact him at [s.borschi@utwente.nl](mailto:s.borschi@utwente.nl).

Open Access funding provided by 'Università di Pisa' within the CRUI CARE Agreement



**IEEE**  
**Annals**  
of the History of Computing

From the analytical engine to the supercomputer, from Pascal to von Neumann, from punched cards to CD-ROMs—*IEEE Annals of the History of Computing* covers the breadth of computer history. The quarterly publication is an active center for the collection and dissemination of information on historical projects and organizations, oral history activities, and international conferences.

[www.computer.org/annals](http://www.computer.org/annals)

IEEE COMPUTER SOCIETY

IEEE