

Learning AI Ethics through Game Design as Ethical Inquiry

Fuji Lyu

Edinburgh Futures Institute, University of Edinburgh, UK

S2465267@ed.ac.uk

Abstract: The Artificial Intelligence (AI) responsibility gap is an abstract and uncertain ethical problem. Judgements of responsibility may change as the situation, available information, and the conditions for action change. Pre-designed ethics cases give learners a concrete setting for discussion, but the problem is usually framed in advance. This study examines a facilitated participatory game-making process as a way to explore questions of responsibility. Fifteen students from computing-related disciplines at a UK university took part in a qualitative participatory study involving case discussion, collaborative tabletop game design, prototype playtesting, and reflection. The analysis focused on how participants translated their understanding of responsibility into game design and how they reconsidered that understanding during later playtesting and discussion. The findings show how participants moved from broad responsibility judgements to more situated reflection as they reconstructed an initially framed problem through collaborative design, playtesting, facilitation, and debriefing. Some previously implicit assumptions and uncertainties became visible in the prototypes and returned as topics for further discussion. The study presents game making as part of an integrated process of ethical inquiry through which participants can articulate, test, and reconsider their judgements of responsibility.

Keywords: AI ethics, Learning through game design, Participatory game design, Tabletop games, Ethical inquiry

1. Introduction

AI ethics is becoming more common in computing and related curricula. One difficult topic is the responsibility gap. When autonomous or learning AI systems contribute to an outcome, human actors may not be able to fully predict or control the system's behaviour. This can make traditional forms of responsibility attribution difficult (Matthias, 2004). AI systems are also developed and used by many technical and social actors. Judgements of responsibility may therefore depend on the situation and on what different actors know, control, and are able to decide. Understanding the responsibility gap involves more than asking who should be responsible. It also requires attention to why responsibility can be difficult to judge clearly and consistently in a particular situation.

Pre-designed ethics cases can place abstract issues in a concrete setting and give learners a starting point for discussion and judgement. In these cases, the situation and main actors are usually provided, and learners interpret and judge a problem that has already been framed (Thiel et al., 2013). In the present study, case discussion and game making were used as connected stages rather than as alternatives. The prepared cases established an initial ethical problem, while the subsequent design activity asked participants to reconstruct and interrogate that problem by deciding how responsibility should take form in a playable system. By turning the initial responsibility problem into a game, participants had to make their own judgements about who was responsible, what different actors could do, and how these relationships should be represented.

Making a game asks learners to turn their understanding into a design that can be run, tested, and revised, rather than only experiencing a game designed by someone else (Kafai, 2006; Kafai and Burke, 2015). Abstract ethical ideas must take form through concrete design choices. Those choices can then be discussed and revised after the prototype is played. For the responsibility gap, this brings an abstract problem into a playable situation built by the participants themselves and allows it to be explored through both design and playtesting.

This paper examines a participatory tabletop game design activity focused on the AI responsibility gap. Following an initial discussion of researcher-prepared AI cases, participants worked together to turn their understanding of responsibility into playable game designs and explored these ideas further through prototyping, playtesting, and reflection. The paper focuses on how this integrated process brings an abstract responsibility problem into a sequence of design, experience, and reconsideration. It asks: How do participants explore and rethink the context-dependent and uncertain nature of AI responsibility by designing and testing games?

2. Ethical Inquiry through Game Making

2.1 From Playing Games to Making Games

Serious games and simulations can give learners an experiential space in which they act within rules and constraints, face consequences, and reflect on decisions. Players make choices within a working system and can observe how those choices shape what happens next. Debriefing connects action in the game with later

discussion and reflection, turning the experience into material for reconsidering judgement (Crookall, 2010). Role-play can also place participants in different positions within a shared fictional situation, where their choices produce consequences that can be experienced and discussed (Barab, Gresalfi and Ingram-Goble, 2010).

Game making changes the learner's relationship with the game. From a constructionist perspective, learners can organise and express their understanding by creating a game artefact as well as by interacting with an existing game system (Kafai, 2006). Kafai and Burke (2015) further show how game making involves learners in constructing rules, representations, and systems of interaction. Ideas that may have remained verbal must take form as design choices and become part of an artefact that others can experience, test, and discuss. Learning can take place while deciding how the game should work, explaining design choices, and revising them. Making also requires learners to decide how an understanding of the problem can work in practice. Some assumptions that were previously implicit become visible in the artefact and available for later discussion.

Participatory design adds a collaborative dimension to this process. Participants can take part in how a problem is represented and made concrete, rather than only responding to a problem that has already been defined (Spinuzzi, 2005; Wanick and Bitelo, 2020). While co-designing a game, participants can externalise their ideas and assumptions in provisional designs, then revisit them through negotiation, prototype testing, and revision. Game making creates an artefact while also providing a shared activity through which different understandings can be expressed and compared.

2.2 Why the Responsibility Gap can be Explored through Game Design

The responsibility gap was first closely associated with the challenge that autonomous and learning systems pose to traditional conditions for assigning responsibility. Matthias (2004) argues that when the behaviour of a learning system cannot be sufficiently predicted or controlled by human operators, approaches that link responsibility to control and foreseeability can become difficult to apply. The central question is whether a human actor can still meet traditional conditions of responsibility when an AI system contributes to a morally significant outcome, and how that responsibility should be understood.

Later work has made the issue more complex. A responsibility gap does not simply mean that no one is responsible after an AI-related outcome. Santoni de Sio and Mecacci (2021) distinguish several forms of responsibility, including blameworthiness, moral accountability, public accountability, and active responsibility. Questions about who deserves blame, who must explain and answer for an outcome, who bears public accountability, and who has a duty to prevent harm are therefore not identical. Königs (2022) also examines what kind of problem an AI responsibility gap represents and how different conditions of responsibility shape our judgement of the gap.

More recent work has questioned the idea of the gap itself. Kiener (2025) argues that AI may create situations with many responsible actors rather than too little responsibility. Noh (2025) examines responsibility in the AI era from the perspective of distributed responsibility. In complex AI sociotechnical systems, questions of responsibility cannot always be reduced to finding a single responsible actor.

Responsibility still has to be judged in context, and the conditions behind that judgement are not always complete or fixed. What an actor can know, influence, and do in a particular situation can shape how responsibility is understood (Santoni de Sio and Mecacci, 2021). As the situation develops or relevant conditions change, earlier judgements may need to be reconsidered. The educational challenge is therefore to make the uncertainty, contextual dependence, and revisability of responsibility judgements available for concrete exploration.

These features make the responsibility gap well suited to exploration through game making. Because responsibility depends on the situation and on the knowledge, control, and capacity for action of different actors, participants must consider how to turn abstract and uncertain relationships of responsibility into a game situation that can actually run. Game making turns the responsibility gap from a concept to interpret into a problem that participants have to construct and test. Their understanding of responsibility becomes concrete enough to discuss and can be revisited after the game is played.

3. Methods

3.1 Research Design and Participants

This study used a qualitative participatory research design combining case discussion, collaborative game design, prototype playtesting, and reflection. It examined how participants expressed, made concrete, and reconsidered their understanding of AI responsibility while making games.

Fifteen students from computing-related disciplines at a UK university took part. Participants were recruited purposively, with additional snowball recruitment, to include different stages of study and levels of practical experience. The first three groups contained three undergraduate students each, spanning Years 2 to 4, while the fourth group contained six master's students. Their programmes included computer science, artificial intelligence, human-computer interaction, game design, data science, and related fields. Participants reported different levels of previous exposure to AI ethics, including experience through relevant ethics courses.

3.2 Activities and Data

Each group completed a facilitated session lasting roughly three hours. The main sequence included around 30 minutes of case discussion, about one hour of collaborative game design, approximately 40 minutes of prototype playtesting, and around 20 minutes of debriefing and reflection. The remaining time was used for instructions, transitions, and related discussion. The initial discussion used researcher-selected news reports concerning autonomous-driving accidents to provide a shared ethical situation and to open the question of who should be responsible when an AI-related outcome goes wrong. During this discussion, the researcher also asked whether participants had previously encountered AI ethics through relevant courses or other study. Participants then developed the responsibility problem into a playable tabletop game. The researcher asked broad questions to help participants clarify the type and direction of the game and, where useful, referred to familiar tabletop games such as Monopoly and Splendor as examples of possible game formats. These references were used only as general design prompts and did not prescribe a particular game type, rule set, or responsibility mechanism. Roles, rules, interactions, and other game elements were discussed and developed by the participants themselves. When a group became stuck, the researcher occasionally used an open question or a brief suggestion to help the discussion continue. These interventions were recorded as part of the facilitated co-design process and considered during analysis.

The data included recordings of group discussions, researcher observation notes, participant-generated game materials, records of prototype playtesting, and post-activity reflections. Recordings were transcribed and the materials were anonymised before analysis. Game artefacts were examined alongside the discussion and playtesting records, so that claims about design or reconsideration could be traced across more than one form of evidence. This made it possible to examine how participants translated their understanding of responsibility into design choices and how those ideas were discussed, tested, and reconsidered in later interaction.

3.3 Data Analysis

The data were analysed using thematic analysis following Braun and Clarke (2006), supported by comparison across different stages of the activity. Transcripts, observation notes, game materials, and playtesting records were organised in NVivo. Initial coding was primarily data-led and focused on moments where participants articulated responsibility, negotiated design choices, encountered problems during play, or reconsidered earlier judgements.

Codes were compared, grouped, and refined into broader themes through repeated movement between the coded material and the full dataset. The analysis then traced how ideas expressed in discussion entered participant-generated designs and what happened to those ideas during playtesting and reflection. Cross-stage comparison was used to identify sequences in which an earlier judgement was maintained, made more specific, questioned, or reformulated. The themes reported in this paper were selected because they captured this movement from abstract judgement to design, testing, and reconsideration.

Because the same researcher facilitated the sessions and conducted the analysis, researcher intervention was treated as part of the interpretative context. Prompts and scaffolding recorded in the observation notes were considered when assessing how a design idea developed. Reconsideration was identified when a change in judgement could be traced across stages or supported by more than one source of evidence.

4. Findings

4.1 From Abstract Judgements of Responsibility to Playable Designs

During the case discussion, participants could express general views about AI responsibility, such as the idea that a contested outcome might involve several actors. Once these ideas had to become part of a tabletop game, saying that several parties shared responsibility was no longer enough to make the game work. Participants had to decide which actors should appear in the game, what positions they occupied, and what choices players could make from those positions.

In the early discussion of one group, participants understood responsibility as involving several actors. Moving into design made this judgement more specific. An open facilitator prompt was followed by a participant proposing that different roles should have different speaking or action orders. Another participant then developed the idea further. A general concern with multiple responsible actors became a design question about who could act, when they could act, and from what position. The interaction shows how researcher scaffolding and participant discussion worked together to make an initially broad judgement more specific through design.

This translation did more than reproduce the earlier discussion. Participants had to find a form in which their understanding could actually work, and the design exposed assumptions that had not been stated before. Giving a role a particular choice, for example, also assumed that the role had the capacity to take that action in the situation. Design decisions like this turned the general question of who can do what into something visible and experienceable in the prototype.

The groups produced different prototypes and did not represent responsibility in the same way. Across the designs, however, participants had to turn ideas first expressed in words into something that others could see, experience, and discuss. Game making left behind a concrete artefact and made some previously implicit assumptions more visible.

4.2 Playtesting Brought Responsibility Back into Discussion

Prototype playtesting brought participants back to choices they had made during design. Some ideas that seemed reasonable during discussion and development produced unexpected results once the game was played. Participants then had to deal with what happened when those designs were put into practice.

One group used temporary numerical values to make differences in responsibility discussable. During playtesting, one participant asked how a particular value should be set. Another participant replied that no value had yet been decided, and the group considered alternative ways of proceeding. The discussion then moved beyond whether the prototype could run to whether such values could be justified as representations of responsibility. The temporary numerical values gave participants a concrete way to discuss differences in responsibility. Playtesting exposed uncertainty about how those values could be justified and represented.

Playtesting changed how participants related to what they had made. The prototype was no longer simply something to finish. It became something that could be run, questioned, and revised. While dealing with problems that appeared during play, participants also returned to the ideas about responsibility that had shaped the design.

4.3 Design Difficulties as Resources for Ethical Reflection

Game making did not always proceed smoothly. Some ideas were difficult to turn directly into workable rules, and designs that seemed clear in discussion sometimes created new problems during making and playtesting. These difficulties pushed participants to explain more clearly what they wanted the game to represent and whether their earlier judgements about responsibility still worked in the design.

The difficulty of moving from ethical discussion to workable design was visible across the activity. In one group, the researcher asked how resources and data could be translated into playable rules. The group paused, after which one participant stressed the need for balance while another commented that the numerical design was difficult. The episode shows how the design task, group negotiation, and facilitation worked together to make the representational problem explicit.

Some design difficulties later opened new ethical discussions. When a prototype did not work as expected, or when group members could not agree on a design choice, participants returned to earlier judgements and explained again what relationships of responsibility they wanted to represent. Sometimes the rules needed to

change. At other times, the understanding behind the rules was reconsidered. Design problems and ethical questions became closely connected.

These unfinished prototypes became temporary objects for thought. They made previously unspoken assumptions visible and gave participants something they could manipulate, question, and revise through making and playtesting. The value of a prototype did not depend only on whether it became a polished game. Designs that did not work smoothly could still leave concrete material for later ethical discussion.

5. Discussion

5.1 Game Making as Ethical Inquiry into the Responsibility Gap

Case discussion and game making played connected but different roles in this study. The researcher-prepared cases provided a shared problem situation and an initial frame for responsibility. The prepared cases gave participants a shared starting point, which they then reconstructed and interrogated through participatory game design by deciding how the problem could be represented and experienced in a playable system.

Research on game making has positioned learners as creators of games as well as players of existing ones (Kafai, 2006; Kafai and Burke, 2015). In this study, that process involved making an ethical problem without a fixed answer work as a game. When participants tried to make the responsibility gap playable, they had to find concrete forms for abstract ideas about responsibility. The prototypes then became shared objects for discussion. Participants could see how their ideas worked in practice and reconsider them after playtesting and discussion.

This matters for the responsibility gap because judgements of responsibility depend on context and may change as relevant conditions change. Participatory game design makes these judgements concrete by asking participants to turn their assumptions into playable choices. Playtesting and debriefing then bring the consequences and tensions created by those choices back into discussion. Across this process, participants can externalise their understanding of responsibility, examine how it works in a playable situation, and reconsider it with others. Game making therefore provides a participatory form of ethical inquiry in which responsibility can be expressed, tested, and revised.

5.2 Design Tensions as Resources for Reflection

Difficulties in game design can raise ethical questions as well as practical ones. A design may fail to express the relationship of responsibility that participants intended, or it may produce an unexpected result when played. At these moments, the design problem can lead back into ethical discussion.

Incomplete prototypes can therefore support reflection. They make unresolved parts of an earlier understanding more visible and prompt participants to explain why a design feels inadequate, unreasonable, or in need of change. Previous research has also shown that student-generated designs and fictional prototypes can provide concrete objects for reflecting on technology and the assumptions behind it (Rapp, 2020). In this study, that reflection extended to judgements of responsibility. For contested ethical problems, design tension can provide a way back into the question itself.

Debriefing is important in this process. Discussion can move from what happened in the game to why the game worked that way and what ethical assumptions were expressed through its design choices. Reflection can address both actions within the fictional situation and the design of the game itself (Crookall, 2010). The prototype can then be questioned and revised as an ethical representation rather than treated as a final account of responsibility.

5.3 Implications for Learning through Game Making

Game-making activities do not need to aim for a highly polished commercial game. For ethical inquiry, a simple prototype that can express a problem, be playtested, and support revision can already provide material for reflection. Educators can begin with a contested ethical case, ask learners to reconstruct it through an initial design, and then use playtesting and debriefing to revisit the assumptions expressed in that design.

In ethical contexts, the making process can give temporary form to participants' understanding of a problem. Participants have to decide which relationships belong in the game and how to represent them, then face those choices again once the prototype is played. The process supported articulation, reflection, and reconsideration of participants' responsibility judgements during the activity.

This approach is particularly useful for problems that cannot be fully addressed through a fixed answer. A game made by learners can serve as a temporary object for thought. It gives concrete form to their understanding at a particular moment while leaving it open to experience, questioning, and revision. The prototype gives participants a shared form through which they can examine their current understanding.

For the AI responsibility gap, uncertainty itself can enter the process of making and reflection. Participants have to decide how to represent different actors, conditions for action, and relationships of responsibility. Problems that emerge when the game is played can bring those decisions back into discussion. Through making, playtesting, and revising the game, learners take part in constructing, representing, and rethinking the responsibility problem.

6. Limitations

This study involved 15 students from computing-related disciplines at one UK university, so the findings are situated within a specific context. Participants from other disciplines, educational settings, or professional contexts may develop different design processes and understand AI responsibility in different ways. Future research can examine game making as ethical inquiry with more diverse groups and in other settings.

The study focused on how participants expressed, made concrete, and reconsidered their understanding of responsibility during an integrated sequence of case discussion, facilitation, game design, playtesting, and reflection. Because all participants experienced this sequence, the study cannot isolate the effects of game design from researcher facilitation, playtesting, or debriefing. It also did not use standardised pre- and post-tests or longitudinal follow-up. The paper therefore reports observable articulation and reconsideration during the activity rather than sustained learning outcomes or transfer. Future research could compare different activity structures and examine whether these processes relate to longer-term ethical learning.

7. Conclusion

This study examined how facilitated participatory tabletop game making can support learners in exploring the AI responsibility gap. When participants had to turn their understanding of responsibility into a game that could run, ethical judgements took form in concrete design choices. Playtesting and discussion then made it possible to question some of those choices and return them to collective reflection.

The process required participants to keep making decisions. They had to decide which actors belonged in the situation, what those actors could know and choose, and what consequences should follow from different choices. When a prototype did not work as expected, these design decisions could return to discussion. The game connected ethical judgement with practical design and gave participants a system of their own making through which to continue engaging with uncertainty around responsibility.

For responsibility problems that do not have a single answer, game making allows learners to reconstruct and interrogate an ethical problem rather than only respond to a finished case. Across case discussion, participatory game design, playtesting, facilitation, and reflection, responsibility judgements become concrete enough to be expressed, tested, and reconsidered. Game making therefore creates a temporary and revisable form through which contested judgements of responsibility can be examined with others.

Ethics Declaration: Ethical approval for this study was granted by the University of Edinburgh (Ref: FLYU03022025) on 4 March 2025. All participants provided informed consent before participation.

AI Declaration: Generative AI tools were used to provide limited support with English phrasing and language editing. All substantive content, analysis, interpretation, and conclusions were developed and verified by the author.

References

- Barab, S.A., Gresalfi, M. and Ingram-Goble, A. (2010) 'Transformational play: Using games to position person, content, and context', *Educational Researcher*, 39(7), pp. 525–536. <https://doi.org/10.3102/0013189X10386593>
- Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3(2), pp. 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Crookall, D. (2010) 'Serious games, debriefing, and simulation/gaming as a discipline', *Simulation & Gaming*, 41(6), pp. 898–920. <https://doi.org/10.1177/1046878110390784>
- Kafai, Y.B. (2006) 'Playing and making games for learning: Instructionist and constructionist perspectives for game studies', *Games and Culture*, 1(1), pp. 36–40. <https://doi.org/10.1177/1555412005281767>

- Kafai, Y.B. and Burke, Q. (2015) 'Constructionist gaming: Understanding the benefits of making games for learning', *Educational Psychologist*, 50(4), pp. 313–334. <https://doi.org/10.1080/00461520.2015.1124022>
- Kiener, M. (2025) 'AI and responsibility: No gap, but abundance', *Journal of Applied Philosophy*, 42(1), pp. 357–374. <https://doi.org/10.1111/japp.12765>
- Königs, P. (2022) 'Artificial intelligence and responsibility gaps: What is the problem?', *Ethics and Information Technology*, 24, article 36. <https://doi.org/10.1007/s10676-022-09643-0>
- Matthias, A. (2004) 'The responsibility gap: Ascribing responsibility for the actions of learning automata', *Ethics and Information Technology*, 6(3), pp. 175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- Noh, H. (2025) 'Beyond the responsibility gap: Distributed non-anthropocentric responsibility in the AI era', *Topoi*. <https://doi.org/10.1007/s11245-025-10302-4>
- Rapp, A. (2020) 'Design fictions for learning: A method for supporting students in reflecting on technology in Human-Computer Interaction courses', *Computers & Education*, 145, article 103725. <https://doi.org/10.1016/j.compedu.2019.103725>
- Santoni de Sio, F. and Mecacci, G. (2021) 'Four responsibility gaps with artificial intelligence: Why they matter and how to address them', *Philosophy & Technology*, 34(4), pp. 1057–1084. <https://doi.org/10.1007/s13347-021-00450-x>
- Spinuzzi, C. (2005) 'The methodology of participatory design', *Technical Communication*, 52(2), pp. 163–174.
- Thiel, C.E., Connelly, S., Harkrider, L., Devenport, L.D., Bagdasarov, Z., Johnson, J.F. and Mumford, M.D. (2013) 'Case-based knowledge and ethics education: Improving learning and transfer through emotionally rich cases', *Science and Engineering Ethics*, 19(1), pp. 265–286. <https://doi.org/10.1007/s11948-011-9318-7>
- Wanick, V. and Bitelo, C. (2020) 'Exploring the use of participatory design in game design: A Brazilian perspective', *International Journal of Serious Games*, 7(3), pp. 3–20. <https://doi.org/10.17083/ijsg.v7i3.358>