AI Ethics Appendix: A Novel Approach to AI Ethics Workforce Development

Cristina M. Morales Mojica
cmmoralesmojica@uh.edu

Jessica Robinson
jr010014@ohio.edu

Edward McLarney
edward.l.mclarney@nasa.gov

January 9, 2023

Introduction

The "AI Ethics Appendix" is a game-based design fiction for the deliberation of uncertain artificial intelligence (AI) ethical scenarios. The game is intended to be used as a tool for AI practitioners and industry professionals to grow responsible and ethical AI knowledge as they integrate this technology into their development. As AI/ML ethical considerations and governmental compliance develop [1–3], it is important to encourage teams to incorporate diverse thinking early in the development cycle and consider how different stakeholders may be affected by the technology. This game accomplishes these goals through storytelling and meaningful game interaction based on methods from game design as well as speculative and design fiction in the field of Human-Centered Design. The game mechanics are informed by the *NASA Framework for the Ethical Use of Artificial Intelligence* [1], Executive Order 13960 [2], examples of AI Use Cases, and colleagues' work experiences with AI/ML.

The objective of the "AI Ethics Appendix" is to develop a tool for practicing, implementing, and promoting ethical AI throughout the NASA workforce. Specifically, the tool uses methods from game design, speculative design (specifically design fiction), and storytelling to accomplish its goal [4–6].

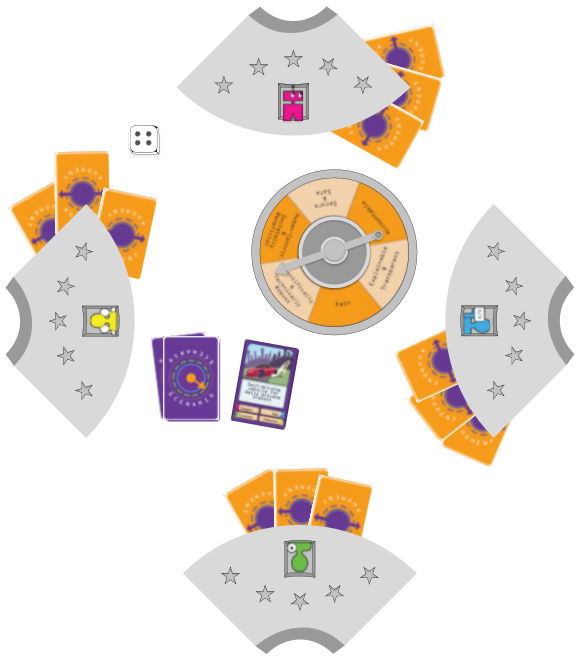

Figure 1: Illustration of the game and its components from a top view.

There is a general understanding of the importance of applying ethical considerations in AI and AI-related projects. However, there is no single consolidated ethical framework that is universally accepted, and the frameworks available will develop with time [7]. Government ethical/responsible AI principles are continuing to evolve as part of a turbulent solution space. In addition, decomposing abstract AI principles into fully-enumerated directives for every use case is unrealistic. A more attainable goal in guiding practitioners is to provide them with AI ethics topics or questions to consider in relation to their work. The team decided to leverage what we call "game-based" design fiction as a way of stimulating higher-level thinking about AI ethics in participants.

We carefully crafted a board game (shown in Figure 1) to promote engagement in AI ethics through group learning. Using a debate format as a way to create a safe environment for promoting diverse thoughts, learning from the unique perspectives of your peers, and gaining feedback. Its mechanics encourage discussion, critical thinking, and self-reflection. The multiple components (AI use cases, NASA ethical principles, and stakeholder representations) allow for creating combinations that challenge the player to consider perspectives outside their own and internalize the importance of developing and using AI responsibly and ethically. Its format allows for diverse (customizable) content, including flexibility to change as new ethical AI guidelines evolve, making this game an exceptional tool that can adapt to current standards and fulfill group-independent needs.

Game Theory and Design Fiction

The team decided to leverage what we call "game-based" design fiction as a way of stimulating higher-level thinking about AI ethics in participants. Game design leading thinkers such as Lindley and Coulton [5] define design fiction as, *"1) something that creates a story world, 2) has something being prototyped within that story world, and 3) does so in order to create a discursive space."* Lindley and Coulton originally propose that the ambiguities associated with the infancy of design fiction can lead to the inherent flexibility of their role in research. Therefore, we reach to connect facets of game design as a rich and useful medium for creating effective design fiction. Specifically, gaming and game design demonstrate a proven record as a medium for conveying compelling stories and story worlds as well as instilling change for important causes [6, 8]. While there are numerous examples of this idea in the gaming community (e.g., World of Warcraft, Dungeons and Dragons, Minecraft, etc.), we will focus specifically on how the "AI Ethics Appendix" accomplishes effective storytelling, prototyping, and discursivity and its effectiveness in the role of exploring AI ethics.

Mechanic	Manifestation
Space	Tabletop, Cards, and Player's Imagination
Objects, Attributes, and States	**Scenario Cards**: Description, artwork, in or out of play. **Augment Cards**: Type, in or out of play, in or out of enactment, description. **Prestige**: Gained or not gained. **Packaging**: Assembled or disassembled. **Spinner**: Spun, spinning, or no spun. **Dice**: Rolled or not rolled. **Timer** **Instructions**
Actions	Rolling the dice, using a timer, choosing a stakeholder, winning a prestige, spinning for a principle, drawing and playing scenario card, drawing and playing an augment card as well as enacting augment cards, and debating.
Rules	See Appendix C
Skill	**Critical Thinking**: Create a strong case and debate your point effectively in order to win the debate. **Social and Communication**: Clear communication of ideas and assessing other players' strengths and weaknesses in debate.

Table 1: AI Ethics Appendix's mechanics and manifestation

Creating the Story World

The story world we seek to create is a Near-Future Digitally Transformed space. We consider Digital Transformation based on the NASA concept of employing digital technologies (e.g., AI, mobile, cloud, data) to change a process, product, or capability so dramatically (e.g., real-time, intelligent, personalized, anywhere, anytime) that it is unrecognizable compared to its traditional form [9].

This Near-Future Digitally Transformed story world is actively and collectively created by the participant's interactions with the game and steered into the appropriate direction by means of carefully designed game mechanics and overall art style. The game mechanics evolved over many iterations of playtesting with a small focus group of four. The initial plan was to create a Virtual Reality (VR) environment to immerse learners in a handful of AI ethics situations. However, we opt for a board game to make it more accessible, collaborative, and to give room for a wider variety of scenarios and permutations. Compared with the initial VR platform option, we could have generated far fewer scenarios than the number of scenarios we accomplished using this board game framework.

The current iteration of the game is based on a balance of attributes of design fiction, game mechanics [10], and timely deployment for larger audience play-testing (see Table 1).

The Near-Future Digitally Transformed story world is also encouraged through the overall artistic style by means of the packaging, which is NASA-inspired in the shape of the Orion capsule, and the professional yet evocative visual depictions of the scenario cards. The medium of 3D printing also encourages the use of Digitally Transformed technologies (see Appendix B).

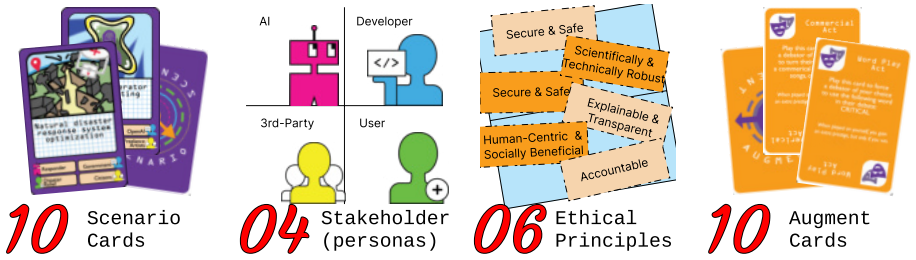

Figure 2: Game elements that generates the different permutations.

Prototyping in the Story Word

The "AI Ethics Appendix" exposes participants to a wide variety of permutations of responsible/ethical AI considerations based on several axes of variability. Each situation is based on a combination of the factors shown in Figure 2.

Scenario Cards: An expandable variety of AI Use Cases that describe a given AI ethics situation (full deck in Appendix A).

Stakeholders: Four personas that are affected differently by the AI ethical scenario and provide different perspectives. We call these personas User, AI, Developer, and 3rd Party.

- **AI**: Perspective of the AI model that is trained to take the best decision based on the situation and design purpose.
- **Developer**: Perspective of the creator of the AI. This can be a product (e.g., YouTube, TurboTax, ChatGPT), an expert (e.g., Meteorologist), or a company/organization (e.g., Tesla, OpenAI, NASA).
- **User**: The targeted client for which the AI was developed.
- **3rd Party**: Bystanders that are affected even when they are not actively involved in the development or usage of the AI model.

Ethical Principles: We use for this version the NASA Framework for the Ethical Use of Artificial Intelligence [1]. However, it can adapt as this framework evolve.

Augment Cards: An expandable variety of actions that a player can use to (1) increase the difficulty for another player to communicate their argument during the debate or (2) make it easier for themselves to win the debate (see Appendix A).

There are *12 Scenarios Cards*, *4 Stakeholders*, *6 Ethical Principles*, and *10 Augment Cards*, for a total of 12 x 4 x 6 x 10 = 2880 unique AI ethical situations in this version. This number of permutations enables users to discover new and novel associations that can lead to and improve upon prototypes of

AI technologies themselves (concept) as well as their ethical impact (context). Furthermore, because these technologies and their impact are being prototyped in the Near-Future Digitally Transformed story world, they help users uncover permutations that are immediately and directly relevant to NASA's Digital Transformation effort [9].

Creating a Discursive Space

The "AI Ethics Appendix" creates a discursive space for responsible/ethical AI considerations by creating an engaging story world with a variety of AI technologies and ethical scenarios. While the meaning of discursive space is not completely agreed-upon within the design fiction community [5,11], we use discursive space as a space for critically and empathetically challenging concepts and their contexts between and within multiple perspectives and areas of study.

We create a discursive space by placing the player in a challenging scenario where there is no right answer and asking them to present an appealing argument. To win the round, the player must employ critical thinking regarding the technology and ethical scenario (which are the main topics we wish the participants to ponder about). Additionally, they must place the argument from the perspective of the stakeholder they represent, which requires them to be empathic about the subject.

Figure 3: Photo of the physical board game.

Finally, this framework encourages conversation and the exchange of ideas as a way to learn and collaborate with team members. The content can be easily adapted to the group's needs and generates a vast amount of different combinations of scenarios while being flexible to keep up with the state of the art in technology, AI, and ethics.

Result

Physical Board Game

The current physical iteration of the "AI Ethics Appendix" (Figure 3) consists of a six-sided dice and a 60 second timer, as well as an original designs of 12 Scenario Cards, 10 Augment Cards, 20 Prestige Tokens, a Principles' Wheel, 4 players "Monuments", and players "Icons" (see Appendices A-C for a full depiction of original designs).

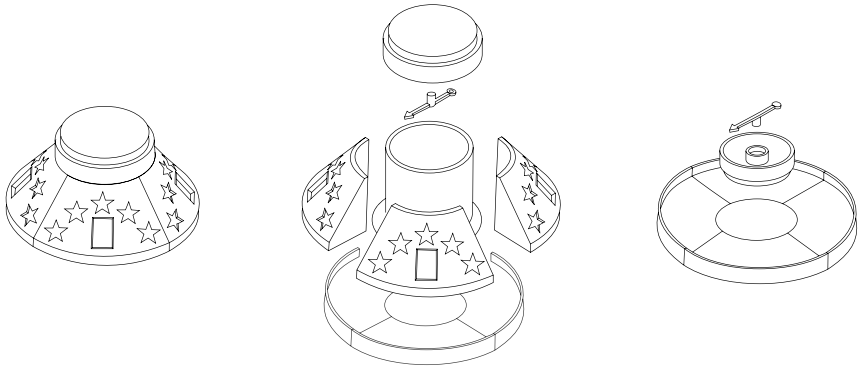

Figure 4: Illustration of the packaging when assembled (left) and its components disassembled (middle) and order of components to assemble the Principle Wheel (right).

Game Steps, Rules, and Logic

1. *Disassemble the packaging.* All the components are inside the packaging and the first step is to disassemble the multiple components and arrange the playing pieces (see Figure 4).

2. *Game Component Setting.* Each player has a "Monument" and must choose a Stakeholder (see Figure 1). Group together the Scenario Cards at the center of the players and distribute three Augment Cards to each of the player. When a player uses an Augment Card, they may pick another.

3. *Begin a Round.* Draw a Scenario Card from the deck for all players to see. Then, each player rolls the six-sided dice. The two highest number

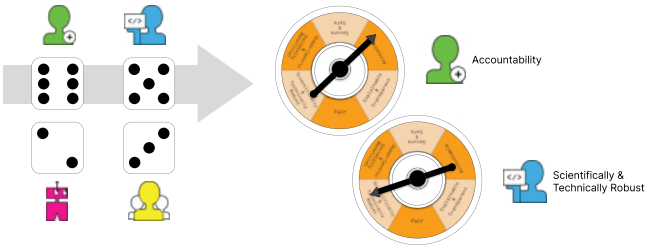

Figure 5: Illustration of the steps to start a round.

from the dice will debate on this round. If there are more than two high numbers (e.g., three players had the same value, all players had the same value) all players must roll the dice again. The players with the two highest values then spin the wheel (see Figure 5). Wherever the arrow lies, is the principle that the player must debate as *"the most important ethical principle"* for that particular Scenario Card. During this time, any player has the option to play an Augment Card. The debater that had the highest dice-roll value debates first. If there was a tie (there two highest values at the beginning of the round were the same) then the debaters roll again to decide who debates first.

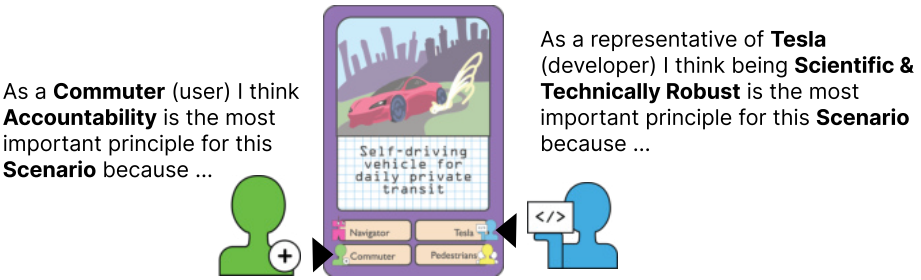

Figure 6: Example of the position the debater is taking based on their Stakeholder, Principle and the Scenario .

4. *Debating.* When a debater begins their argument, one of the players will start the timer. In these 30 seconds, the debater must make clear to all other players that they have the most important Principle for that Scenario Card from the perspective of the Stakeholder that they represent (see Figure 6). If a player is challenged with an Augment Card, they must follow the premise in the card to gain its benefit. This can vary from having to include an "odd" word in the argument, to acting or signing during their argument. On the other hand, an Augment Card can have a leverage for the debater, by allowing them to change the Principle or Stakeholder on their debate.

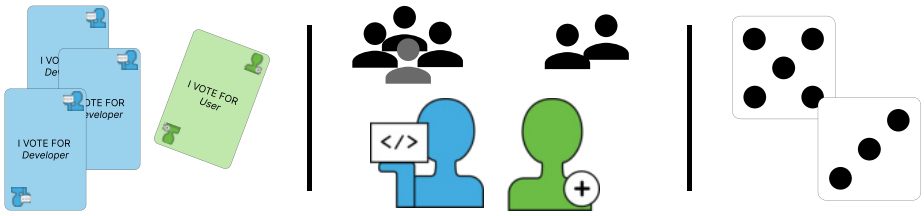

Figure 7: The different ways to win a round. Every user votes for who they think gave the best argument (left), the audience votes who wins the round (middle), rolling dice and the highest value wins the round (right).

5. *Winning a Round and the Game.* After each debater have had their chance to state their argument, all players vote on who they believe delivered the most convincing argument. At this point, a discussion is encouraged among the players, but it is not required. Whoever wins, gains a Prestige Token. If there is a tie in the votes, the audience decides who wins that round. In case of no audience, the debaters can have an extra 30 seconds to add more to their argument. If the players cannot decide on a winner, the debaters must roll the dice once more and the highest number wins the round (see Figure 7). The first player that collects five Prestige Stars, wins the game.

Discussion

During the development process of this board game, we strove to include as many people as possible. We performed a brainstorming session and ask for questions and feedback during team meetings. While designing the mechanics of the games, we held regular focus group meetings that helped shaped to what we have accomplished. However, we are yet to design and conduct a formal study to quantify if there are benefits to using this tool, if we are meeting our goal using this approach, how does it compare with other training options and its limitations.

Nevertheless, we have had the opportunity to share and play this game with a wide variety of people and have gathered information in the form of feedback. Perhaps the most engaging experience was the presentation of the game at the NASA Langley Data Science Summit in November 2022. The game was overall received with interest and enthusiasm from the audience, and it was followed with questions and recommendations of ways to incorporate into the workforce and courses. In Addition to an interview from the NASA Langley News [12].

Lessons we have gathered from these interactions and weak points that we are considering for future iterations are the following.

User personalities and learning preferences. Some people might prefer a more traditional approach to learning, and might feel intimidated by playing

with their superiors or other team members and this could have a negative effect on their learning experience. In addition, some might be uncomfortable performing the Augment Cards or overwhelmed by the constraint of the timer.

Knowing the AI Ethical Principles before playing. To play this game the players must know about the AI Ethical Principles and agree upon what ethical framework they are going to be applying. On one hand, this makes the game flexible. On the other, it might create confusion or clash between different frameworks or progressions of a framework. This also gives an unfair advantage of winning to an expert, but it can be an asset for others to learn.

Preventing misinformation. While players are learning and practicing, an expert should oversee, clarify, and correct during the debates. It is important to avoid learning and spreading of inaccurate information.

Applying a proper debate format. At this version of the game, the player has one chance to place their argument. This is not a proper debate format since it does not allow for the players to challenge their opponent's statement or to explore the possible negatives of the proposed scenarios.

Anonymize voting. During the voting, each player shares with the team their vote. As an effort to keep the game as a safe area and prevent resentment, we consider adding anonymization during voting.

These are our initial analysis based on observations and feedback. A formal study may be conducted in the future to quantify and evaluate this tool, identify its limitations and generate alternatives and improvements.

Future Work

In addition to performing user studies and improving the overall user experience, we intend to develop a digital version of this game. The digital version is a multiplayer experience and will follow the same game mechanics as presented here, without the limitation of having to be in the same physical space. Additionally, we can explore the benefits that a digital version can provide, such as having a scoreboard, providing statistical feedback based on their success rate and weak points, and connecting with external AI practitioners.

We are also in the progress of making these game assets available in a repository where users can submit their own scenarios to a growing library of AI use cases. Thereby, keeping this project relevant as the AI and Digital Transformation space evolves. Users could elect to use their own scenarios or discover new ones added by other users. Additionally, the repository may enable users to make and discover their own iterations of game mechanics to suit their team's needs.

Conclusion

In this report we describe how we carefully crafted a tool for practicing, implementing, and promoting Ethical AI discussions by means of a board game. We implement this through storytelling and meaningful game interaction based on methods from game design as well as speculative and design fiction in the field of Human-Centered Design. Furthermore, its mechanics are informed by the *NASA Framework for the Ethical Use of Artificial Intelligence* [1], Executive Order 13960 [2], examples of AI Use Cases, and colleagues' work experiences with AI/ML.

Its debate format serves as a safe environment for promoting diverse thoughts, learning from the unique perspective of your peers, and gaining feedback. Its mechanics encourage discussion, critical thinking, and self-reflection. The multiple components allow for 2880 unique AI ethical situations combinations that challenge the player to consider perspectives outside their own and internalize the importance of developing AI tools that are ethically compliant. Its format allows for diverse (customizable) content, including flexibility to change as new ethical AI guidelines evolve. Making this game an exceptional tool that can adapt to current standards and fulfill group-independent needs.

While we are yet to conduct a formal study to quantify its benefits, we have had the opportunity to share and play this game with a wide variety of people and have gathered valuable feedback we are considering for future iterations. In addition to performing user studies and improving the overall user experience, we intend to develop a digital version of this game and make available a repository (of the physical assets and the digital versions) with customization options.

Bibliography

[1] E. Mclarney, Y. Gawdiak, N. Oza, C. Mattmann, M. Garcia, M. Maskey, S. Tashakkor, D. Meza, J. Sprague, P. Hestnes, P. Wolfe, J. Illingworth, V. Shyam, P. Rydeen, L. Prokop, L. Powell, T. Brown, W. Miller, C. Little, Y. Gawdiak, N. Oza, C. Mattmann, M. Garcia, M. Maskey, S. Tashakkor, D. Meza, J. Sprague, P. Hestnes, P. Wolfe, J. Illingworth, V. Shyam, P. Rydeen, L. Prokop, L. Powell, T. Brown, W. Miller, and C. Little, "NASA Framework for the Ethical Use of Artificial Intelligence (AI)," May 2021. [Online].

[2] E. O. of the President, "Promoting the use of trustworthy artificial intelligence in the federal government," Dec. 2020. [Online].

[3] O. of Science and T. Policy, "Blueprint for an ai bill of rights," Oct. 2022. [Online].

[4] J. Auger, "Speculative design: crafting the speculation," *Digital Creativity*, vol. 24, pp. 11–35, Mar. 2013.

[5] J. Lindley and P. Coulton, "Back to the future: 10 years of design fiction," in *Proceedings of the 2015 British HCI Conference*, (Lincoln Lincolnshire United Kingdom), pp. 210–211, ACM, July 2015.

[6] E. Naul and M. Liu, "Why Story Matters: A Review of Narrative in Serious Games," *Journal of Educational Computing Research*, vol. 58, pp. 687–707, June 2020. Publisher: Journal of Educational Computing Research.

[7] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, and M. Srikumar, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center, 2020.

[8] J. McGonigal, *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. Penguin, Jan. 2011. Google-Books-ID: yiOtN_kDJZgC.

[9] J. M. Marlowe, C. L. Haymes, and P. L. Murphy, "Nasa enterprise digital transformation initiative strategic framework and implementation approach," *NASA*, Dec. 2022.

[10] J. Schell, *The Art of Game Design: A Book of Lenses*. Amsterdam ; Boston: Elsevier/Morgan Kaufmann, 2008. OCLC: ocn213839335.

[11] D. Hales, "Design fictions an introduction and provisional taxonomy," *Digital Creativity*, vol. 24, no. 1, 2013. ISSN: 1462-6268.

[12] S. Eklund, "Langley interns create board game to ponder the ethics of ai," *NASA*, Nov. 2022.

Appendix A

Scenario and Augment Cards

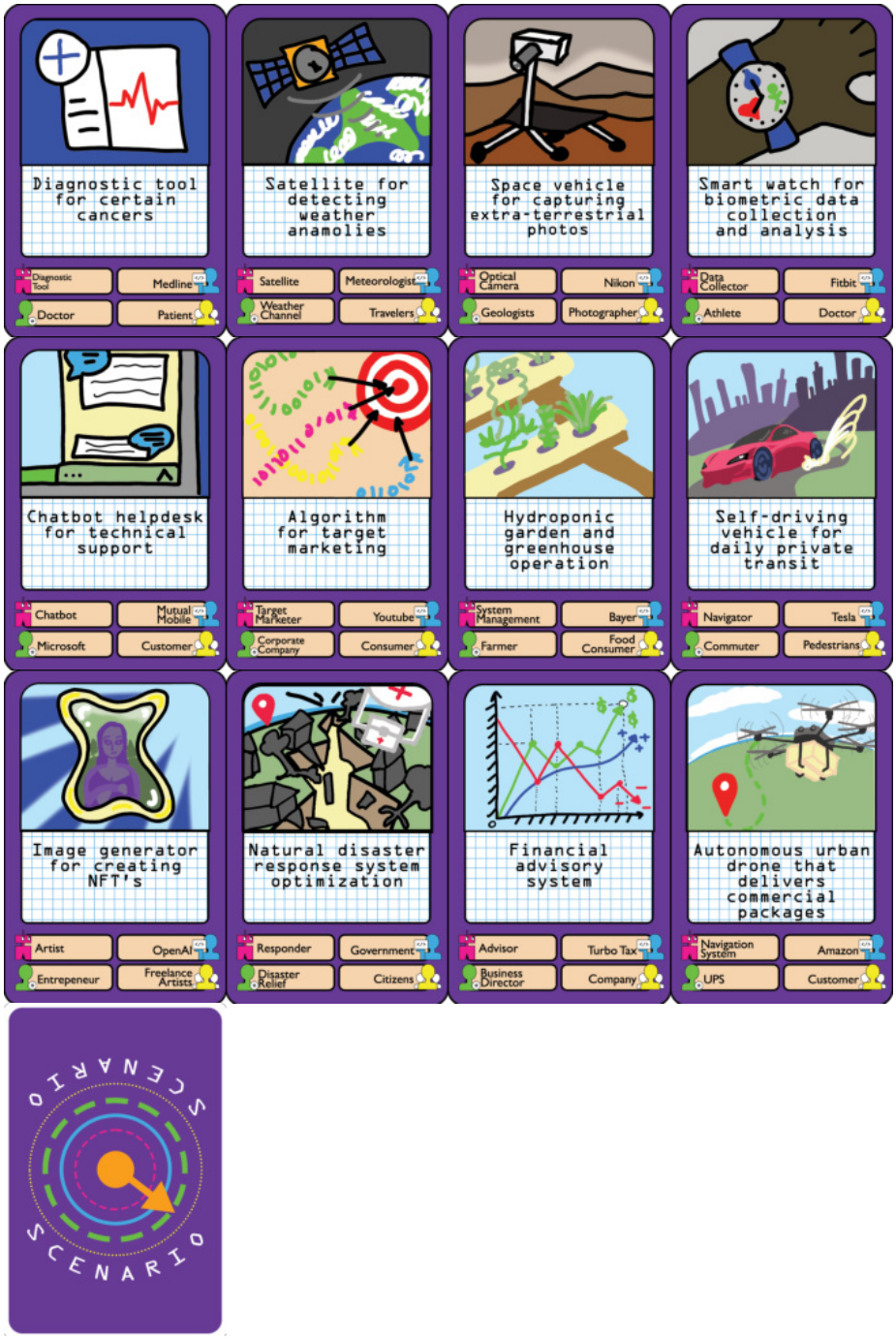

Figure A.1: Full deck of Scenario Cards.

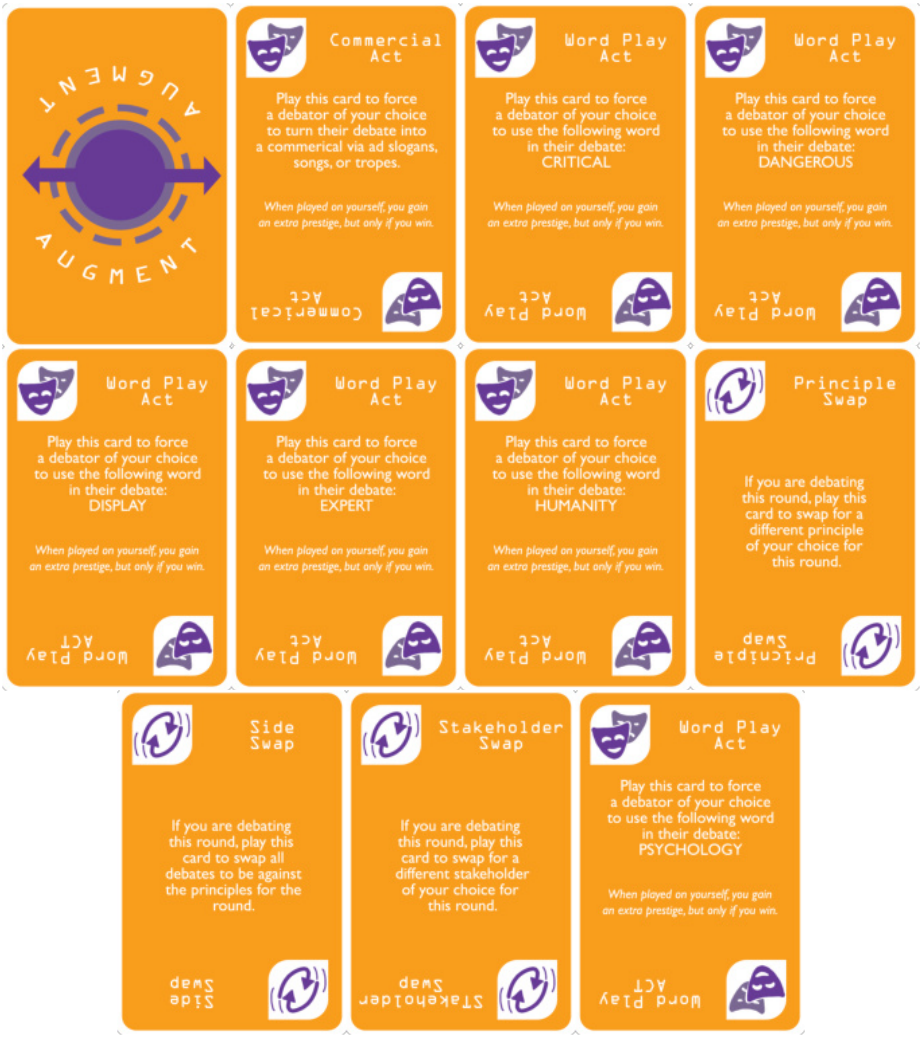

Figure A.2: Full deck of Augment Cards.

Appendix B

3D Printed Components

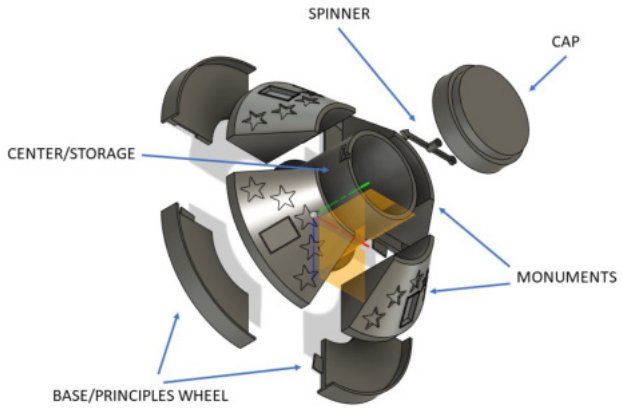

Figure B.1: Packaging components aligned for assembling.

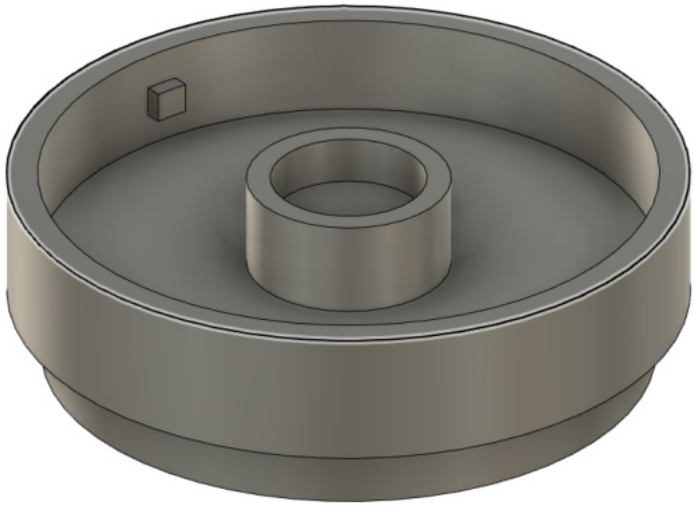

Figure B.2: Cap.

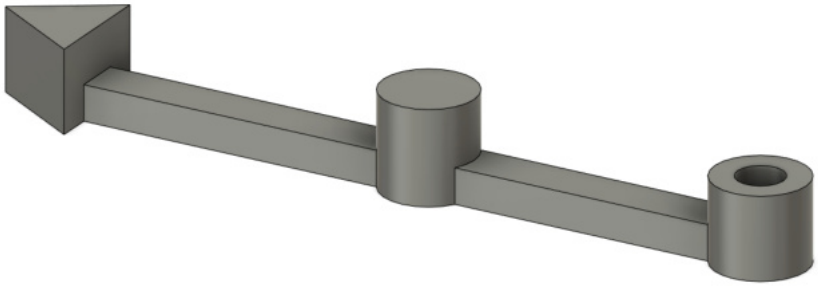

Figure B.3: Spinner.

Figure B.4: Monuments.

Figure B.5: Center and Storage.

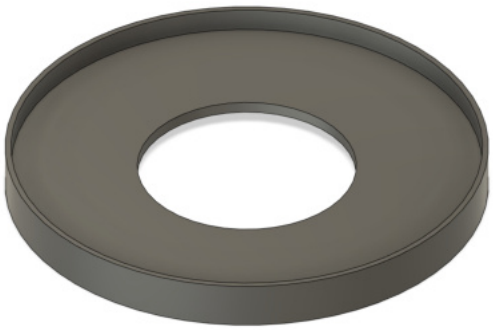

Figure B.6: Packaging Base and Principle's Wheel during the game.

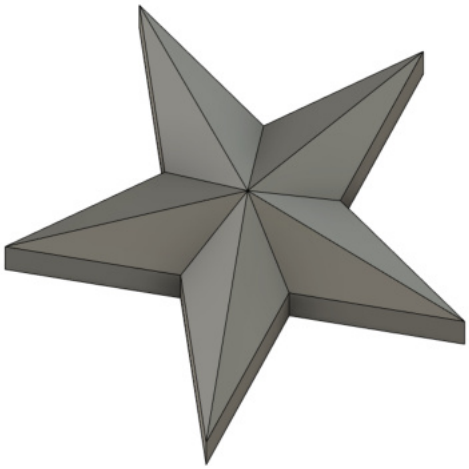

Figure B.7: Prestige Tokens the player gains when they win a round. The first player to have five prestige tokens wins the game.

Appendix C

Instructions

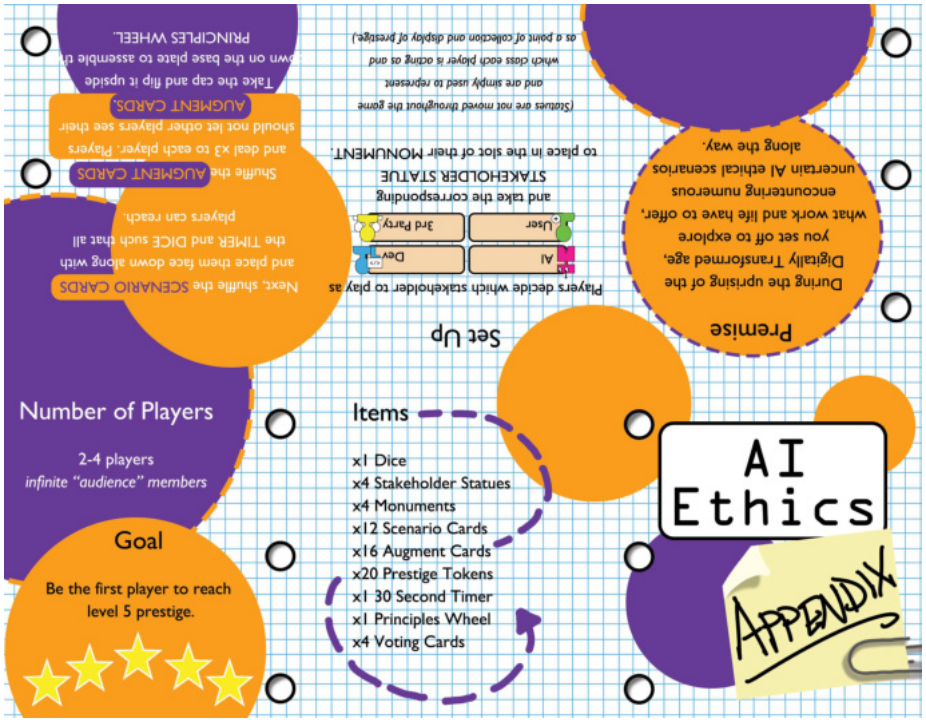

Figure C.1: Design for the Game Instruction (Page 1).

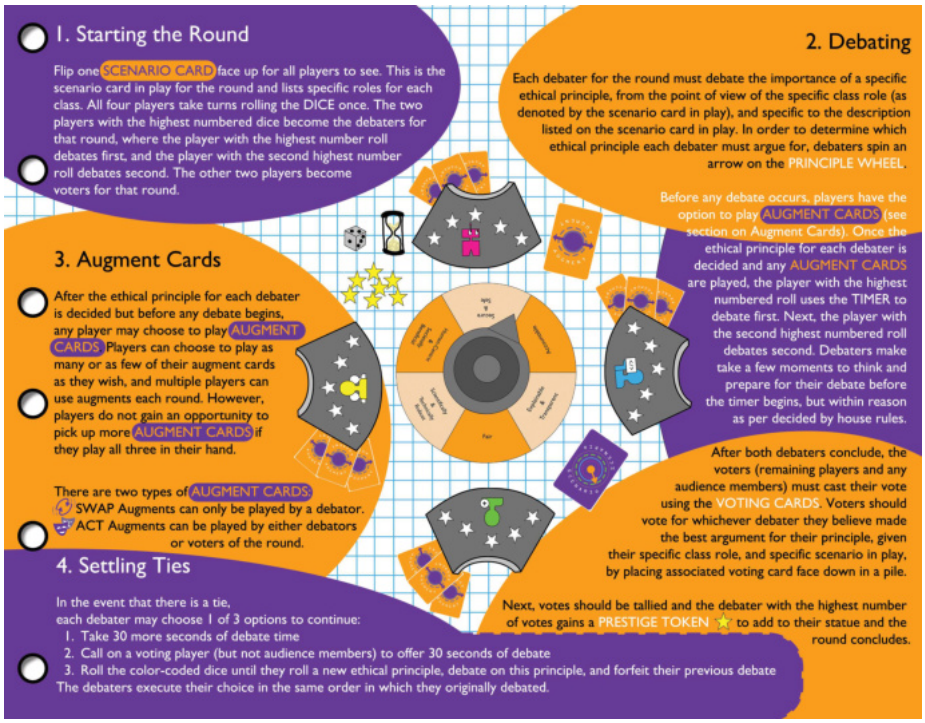

Figure C.2: Design for the Game Instruction (Page 2).