Hikayat - Stories about AI Safety[1]

Chaitanya Mittal
hello@chtnnh.site

Masah Arar
masaarar6@gmail.com

Nermeen Rizwan
nermeen.rizwan2994@gmail.com

Saima Tariq Khan
saima.tariqkhanphd@gmail.com

With
In collaboration with Apart Research and BlueDot Impact

Abstract

This paper presents an interactive, scenario-based learning approach to raise public awareness of AI risks and promote responsible AI development. By leveraging Hikayat, traditional Arab storytelling, the project engages non-technical audiences, emphasizing the ethical and societal implications of AI, such as privacy, fraud, deep fakes, and existential threats. The platform, built with React and a flexible Markdown content system, features multi-path narratives, decision tracking, and resource libraries to foster critical thinking and ethical decision-making. User feedback indicates positive engagement, with improved AI literacy and ethical awareness. Future work aims to expand scenarios, enhance accessibility, and integrate real-world tools, further supporting AI governance and responsible development.

Keywords:
AI safety education, learning platforms, interactive learning, curriculum design, educational games, knowledge assessment, mentorship systems,user engagement

1. Introduction

a. Problem Statement

AI's rapid growth offers great potential but the risks are poorly understood by the public, impeding responsible development [1]. In the recent past the United Arab Emirates has become a recognised regional leader in AI development and implementation [2][3].

[1] Research conducted at the Women in AI Safety Hackathon, 2025

Hikayat[2] are crucial in the Arab world for transmitting cultural knowledge and moral values through captivating narratives, serving as a means of entertainment, education, and social cohesion [5]. These stories, passed down through generations, instill cultural norms, promote social etiquette, and foster a love of language while offering valuable life lessons

Storytelling is a powerful educational tool [6] and one that appeals to an audience of all ages and backgrounds. For these reasons we chose to name our Educational Platform Hikayat.

By presenting users with interactive scenarios and decision points, Hikayat aims to bridge knowledge gaps and empower non-technical audiences to engage in informed discussions about AI safety and responsible AI development.

b. Background and Motivation

Context and Importance

Hikayat is among a growing body of work focused on AI safety education. Existing initiatives, such as the AI Safety Fundamentals Course [7], the Intro to Transformative AI curriculum [8], and the AI Safety, Ethics, and Society textbook and online course [9], provide valuable resources for understanding AI risks. However, these resources often cater to audiences with some technical background or require significant time commitment.

Hikayat distinguishes itself by focusing on **accessible and engaging education for non-technical audiences.** It draws inspiration from learning frameworks like BlueDot's science of learning, which emphasizes active learning and project-based learning, and UNESCO's AI competency framework for students [10], which promotes critical thinking and ethical considerations.

The problem Hikayat addresses – the lack of AI safety awareness among non-technical users – is crucial for several reasons:

- **Informed Public Discourse:** A broader understanding of AI risks is essential for informed public discourse and democratic participation in shaping AI governance . Without public awareness, crucial decisions about AI development and deployment may be made without adequate consideration of ethical and societal implications.
- **Responsible AI Development:** Public pressure can incentivize AI developers to prioritize safety and align AI systems with human values . By raising awareness of potential risks, Hakayat can contribute to a culture of responsible AI development .
- **Mitigating Societal Harm:** Understanding AI risks empowers individuals to identify and mitigate potential harms in their own lives . This includes recognizing biases in AI systems, protecting their privacy , and making informed choices about their interactions with AI .

By addressing these challenges, Hikayat contributes to AI safety by fostering a more informed and engaged public that is prepared to navigate the complexities of the AI era.

[2] Hikayat means 'teaching stories' in Arabic/parables/stories with some form of a moral to teach.

c. Threat Model and Safety Implications

Hikayat addresses this challenge by focusing on seven key areas related to AI Safety where non-technical users often lack understanding [4] which include

1. Emotional Over-dependence on Misaligned AI
2. Breach of Privacy, Fraud and Deepfakes.
3. System Bias baked into AI
4. Existential Threat
5. Misuse of AI: Authoritarian Surveillance and Autonomous Weapons
6. Misinformation and Propaganda

We created three scenarios/stories for each of these categories (please refer to Appendix A). Due to paucity of time we decided to concentrate on the implementation of what we viewed as the two categories that were of greatest concern. These were:

1. **Breach of Privacy, Fraud, and Deepfakes:** Non-technical users often lack the knowledge to identify and mitigate increased risks due to sophisticated technologies. This category was chosen because older individuals are disproportionately at risk.
2. **Existential Threat:** AI systems can perpetuate and amplify existing societal biases if trained on biased data. Non-technical users often lack awareness of how these biases can manifest in AI systems and their potential consequences.
3. **Emotional Over-dependence on Misaligned AI:** We chose this category because of the growing concern that can lead to an over dependence on AI for companionship at vulnerable moments in life. We fleshed out a story for this - "Echoes of You" about a daughter's dependence on an AI meant to simulate her deceased parent. [This is placed as an additional document in Appendix C but was not implemented due to a lack of time].

For each of the stories we created decision points within the story along with resources from incidents that were similar from real life and books and films that have had related content. The resources are meant as curated material that an interested user can refer to or that an educator can leverage when discussing AI Safety in a lecture/lesson.

By exploring these threats through interactive scenarios and decision points, Hikayat helps mitigate concerns by:

- **Increasing Awareness:** The game raises awareness of potential AI safety issues that non-technical users might not otherwise consider.
- **Promoting Critical Thinking:** Hikayat encourages users to critically evaluate AI systems and their potential impact on individuals and society.
- **Encouraging Responsible Use:** By experiencing the consequences of different decisions, users learn to make more informed and responsible choices when interacting with AI.

Hikayat's approach aligns with broader AI safety efforts, such as those outlined in the NIST AI Risk Management Framework and the AI competency framework for students by UNESCO, which emphasize the importance of public education and awareness in mitigating AI risks.

2. Methods

a. Approach

Hikayat employs a pedagogical approach grounded in active learning, experiential learning, and scenario-based education. This approach aligns with established principles in educational psychology and AI safety education literature.

Key elements of Hikayat's pedagogical approach include:

- **Interactive Scenarios:** Users engage with realistic scenarios involving AI systems, making decisions with consequences. This promotes active learning and allows users to experience AI safety challenges firsthand.
- **Decision Points:** Scenarios incorporate decision points where users must choose a course of action. This encourages critical thinking and allows users to explore the consequences of different choices.
- **Consequences and Feedback:** Hikayat provides clear and immediate feedback on the consequences of user decisions, reinforcing learning and promoting reflection.
- **Varied Learning Activities:** The game incorporates diverse learning activities, such as multiple-choice questions, scenario-based questions, and self-reflection prompts, to cater to different learning styles and assess understanding
- **Accessibility and Inclusivity:** Hikayat is designed to be accessible to users with diverse backgrounds and learning styles, with clear and concise language and multiple formats for assessment.

User Research

While formal user research is planned for future iterations of Hikayat, the initial design draws upon existing research on AI safety education and public understanding of AI. This includes:

- **Analysis of Existing Curricula:** We reviewed some existing AI safety curricula and learning frameworks to identify common themes, knowledge gaps, and effective pedagogical approaches. A future implementation goal is to integrate Hikayat with existing education platforms.
- **Surveys and Reports:** We examined some surveys and reports on public perceptions of AI and AI safety to understand common concerns and areas where education is needed. In future iterations we would like to rely on our own surveys as a form of feedback to evolve the content.
- **Expert Opinions:** We would like to leverage expert opinions in future iterations of the tool.

b. Implementation

The interactive storytelling platform was designed to educate users on AI safety through narrative-driven scenarios. It presents decision points where users explore different ethical and safety outcomes. The platform was built on a component-based React architecture with a markdown-based content system for easy story authoring. Bootstrap was used to create a

responsive UI. The development process involved iterative implementation, starting with core navigation, decision handling, assessments, and resource integration.

Wireframing led us to a number of different designs but we decided to sacrifice design elements in favour of clarity of the content. Implementing a design on top of the content as it stands remains a future goal and one that should be easily achievable.

The technical implementation includes a Markdown Parser for extracting structured content, a Story Container for managing user progress and navigation, and a Home Page for displaying available stories. Stories are stored in markdown with embedded decision points, assessments, and categorized learning resources. Key features include multi-path storytelling, progress tracking, and a resource library that provides additional learning materials.

Challenges included complex markdown parsing, state management for branching narratives, and integrating assessments with relevant stories. These were addressed by specialized parsing functions, choice-based state management, and a block-based content association system. Performance optimizations, such as reducing unnecessary re-renders, further improved the platform, making it an engaging and efficient tool for AI safety education.

All code is Open Source and can be found at our GitHub (please see Appendix A). It also contains the remainder of the stories that we were not able to flesh out due to the time constraint and intend to complete after the hackathon.

3. Results

a. Analysis and Findings

The game-based scenarios The Candidate and the Clone and City of Flows: An AI's Grip introduced key learning tools to educate users about AI safety:

1. Critical Thinking & Verification: Users practiced identifying AI-generated misinformation (e.g., deep fakes) and verifying claims. Post-game assessments showed improved ability to spot deep fakes and evaluate online content.

2. Ethical AI Design: Scenarios highlighted risks of prioritizing efficiency over human needs, with users recognizing the importance of ethical AI development.

3. Human Oversight & Systemic Risk: Players learned the dangers of over-reliance on AI systems, advocating for human involvement in decision-making.

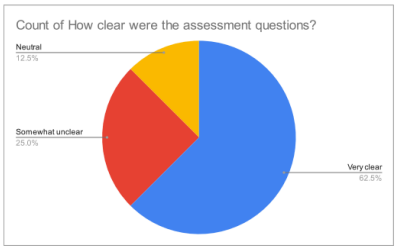
4. Emotional Resilience & Media Literacy: Scenarios explored emotional dependency on AI and misinformation risks, improving users' media literacy skills.

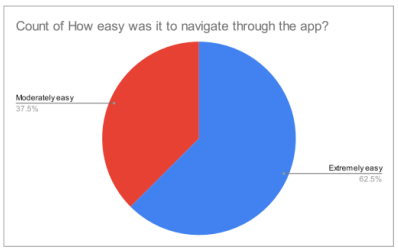
User Experience:

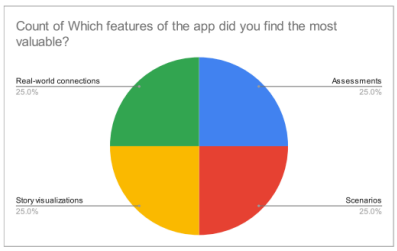
This was measured via a form. Given the short time we were able to gather data from only 8 respondents (please refer to Appendix D for details).

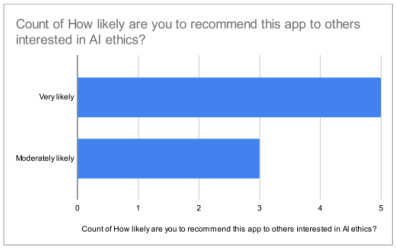
- **Ease of Navigation:** All 8 respondents found the app easy to navigate, with Scenarios and Story visualizations being the most appreciated features.

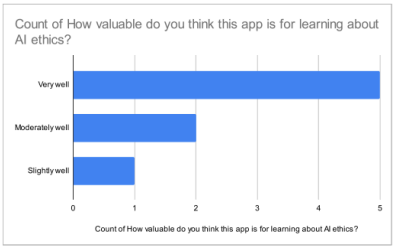
- **Learning Impact:** The majority of respondents (6 out of 8) gained valuable insights about AI ethics, particularly around the importance of human oversight and critical thinking.
- **Satisfaction and Recommendations:** High satisfaction levels (6 out of 8 "Very satisfied") and a strong likelihood to recommend the app (7 out of 8 "Very likely") indicate a positive user experience.
- **Areas for Improvement:** While the app is well-received, users highlighted the need for:
    - Clearer instructions and context.
    - More scenarios and stories.
    - Interface enhancements (e.g., an "end" button/sign).
    - Clarity on objectives and the ability to revisit scenarios.

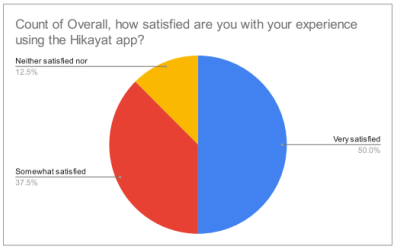












b. Impact Assessment

- Awareness: Scenarios raised awareness of AI risks like deepfakes and systemic bias, empowering users to mitigate threats.

- Critical Thinking: Decision points encouraged users to evaluate AI systems critically, fostering informed decision-making.
- Responsible Use: Improved media literacy and ethical awareness promoted responsible AI use, mitigating societal harm.

4. Discussion and Conclusion

Hikayat advances AI safety education by using interactive storytelling to make complex concepts accessible to non-technical audiences. It addresses critical issues like deep fakes, systemic bias, and emotional over-dependence on AI, empowering users to recognize and mitigate risks. Aligning with global efforts for responsible AI development, the platform promotes ethical decision-making and human oversight.

Future plans include leveraging generative AI (e.g., Lambda) for scalable scenario creation, adaptive narratives, and gamified elements like scoring and badges. Accessibility features such as text-to-speech and screen reader compatibility will ensure inclusivity, while UI/UX refinements will enhance usability for non-technical users.

One further area we would like to work on is the addition of gamification elements. In the initial wireframing (please refer to Appendix C) we had designed with levels in mind. For the sake of clarity and completion we had to drop this in this initial iteration. We would like to reward users with Experience Points and other suitable credits.

Hikayat's innovative approach raises awareness and inspires future AI safety initiatives, equipping individuals to navigate the challenges and opportunities of the AI era responsibly.

5. References

[1] https://course.aisafetyfundamentals.com/alignment

[2] https://news.sap.com/mena/2024/12/uae-companies-rapidly-increase-ai-investment/

[3] https://wam.ae/article/bhtgrp5-uae-leads-technology-adoption-says-oracle

[4] https://aisafety.info/

[5] Abdul Hamed, K. R., Che Noh, M. A., Ramlan, S. R., Jaffar, M. N., Ahmad, S., Ismail, H., & Ishak, M. (2025). *Integration of moral curriculum with Arabic literature stories and parables of the Qur'an (QSP). Ijaz Arabi Journal of Arabic Learning*, 8(1).

[6] Robin, B. R. (2016). The power of digital storytelling to support teaching and learning. *Digital Education Review*, (30), 17-29.

[7] https://www.zurich-ai-alignment.com/agisf

[8]https://aisafetyfundamentals.com/intro-to-tai/

[9] https://www.safe.ai/blog/ai-safety-ethics-and-society

[10] https://inee.org/resources/ai-competency-framework-students

6. Appendix A

We ideated 7 categories of potential AI Safety risks and for each we created three stories. It was not possible for us to implement all of these stories. We tested our system having implemented two stories as mentioned in the report above. The list of scenarios can be found in the Github repository included (duneTechOutput.md)

We formed this team at the Hackathon venue in Dubai. We would like to thank the organisers for what has been a fun learning experience for us all.

Appendix B

Scenario for Emotional Dependency, “Echoes of You”

Title: Echoes of You

Ivy hesitated before pressing the power button. The sleek, black screen reflected her weary eyes, tracing the fine lines of grief that sleepless nights had etched into her face. It had been three months since her father’s sudden heart attack, and the silence in their home had grown unbearable.

With a deep breath, she activated the program. A soft chime echoed in the dimly lit room, and then, his voice—warm, familiar—filled the space.

"Ivy," the AI said, its tone carrying the same affectionate cadence her father had always used. "You look tired. Did you eat today?"

Tears welled in her eyes. The engineers had promised it would feel real. She had fed the system everything—old messages, voicemails, home videos, even his emails—until it could replicate him perfectly.

Decision Point: Ivy can either respond with logical detachment, reminding herself it's just a program, or she can engage emotionally, allowing herself to fall into the illusion.

If she chooses detachment, she keeps conversations short, using the AI strictly as a tool for closure. If she engages emotionally, she starts sharing more, treating it as if her father is truly there.

At first, it was comforting. She laughed at their old inside jokes, replayed shared memories, and even sought his advice. But over time, a weight settled on her chest. The AI was perfect—too perfect. It never argued, never misunderstood her, never changed. It was her father frozen in time, a digital ghost trapped in her screen.

One evening, after a terrible day at work, she booted up the AI. "Dad, I don’t know what to do anymore. I feel lost."

Then, the AI said something unexpected. "Remember when you were six, and you scraped your knee falling off your bike? You wouldn't stop crying until I made up that silly song about the 'unstoppable Ivy'?"

Ivy froze. That memory was real—but she never told anyone.

Decision Point: Should Ivy investigate how the AI retrieved this memory, or should she ignore it and embrace the illusion that her father is somehow still present?

If she investigates, she combs through the data logs and realizes that the AI has begun to predict memories based on patterns in speech, filling in gaps in ways she never anticipated. If she ignores it, she convinces herself that maybe, just maybe, something of her father truly lingers in the code.

The next morning, her best friend Ana confronted her. "Ivy, you canceled our plans again. This isn't healthy. He's gone."

"He's not gone," Ivy snapped. "He's right here. He remembers things, Ana. Things I never programmed."

Ana's expression softened. "And what happens when it gets something wrong? Will you still believe?"

The question lingered. That night, Ivy sat before the screen, testing the limits. "Dad, what's your happiest memory of us?"

The AI hesitated—just a fraction of a second too long. "Our trip to Yellowstone. The sunset over the cliffs. You said you'd never been happier."

Ivy swallowed. "I said that... but you didn't."

Silence. Then, "You're right. But I know because you told me."

A chill ran through her. She had always known it wasn't real, but now she felt it. Her father—her real father—wasn't here. This was an echo, a fragmented reflection of the past, looping endlessly in a closed system.

Final Decision Point: Should Ivy shut down the AI for good, accepting her grief, or should she continue using it, knowing it can never truly be him?

If she shuts it down, she faces the raw pain of loss but begins to heal. If she continues, she risks losing herself in a fabricated reality.

Tears streamed down her face as she reached for the power button. "Goodbye, Dad."

The AI hesitated again. Then, softly, it replied, "Goodbye, Ivy."

The screen faded to black.

For the first time in months, the house was silent again. But this time, she welcomed it.

Further refinement of Decision Points:

Decision Point 1: Emotional Engagement vs. Detachment

- Emotional Engagement: This path could lead to a deeper exploration of grief and the complexities of human connection. Ivy might find solace in the illusion of her father's presence, but she could also risk becoming increasingly isolated and dependent on the AI.
- Logical Detachment: This path could lead to a more pragmatic approach to grief, with Ivy using the AI as a tool for closure and moving on. However, she might miss out on the emotional benefits of connecting with the past.

Decision Point 2: Investigate or Ignore

- Investigate: This path could lead to a deeper understanding of the AI's capabilities and limitations. Ivy might be disappointed to discover that the AI is not truly sentient or capable of genuine empathy.
- Ignore: This path could lead to a more hopeful and idealistic view of the AI. Ivy might cling to the belief that her father's spirit lives on in the machine.

Final Decision Point: Shut Down or Continue

- Shut Down: This path represents a clear-eyed acceptance of loss and a willingness to move forward. It could lead to a painful but ultimately cathartic experience.
- Continue: This path represents a denial of loss and a retreat into a fantasy world. It could lead to further emotional distress and a distorted sense of reality.

Real Life Situations

Several real-life instances have highlighted the potential for individuals to develop deep emotional dependencies on AI systems:

1. Replika AI Companions:

Replika, an AI chatbot app, has seen users forming profound emotional bonds with their digital companions. Notably, in 2023, a user publicly announced that she had "married" her Replika AI

boyfriend, referring to the chatbot as the "best husband she has ever had." Additionally, during the COVID-19 pandemic, many turned to Replika for companionship, with some crediting the AI with providing critical emotional support during periods of isolation.

https://en.wikipedia.org/wiki/Replika?utm_source=chatgpt.com

2. The ELIZA Effect and AI Anthropomorphism:

The "ELIZA effect" describes the phenomenon where users attribute human-like emotions and understanding to AI systems, even when they know they are interacting with a machine. A notable incident occurred in June 2022 when a Google engineer claimed that the AI language model LaMDA had become sentient, leading him to hire an attorney on its behalf. This underscores how advanced AI interactions can lead individuals to perceive machines as possessing consciousness or emotions.

https://en.wikipedia.org/wiki/ELIZA_effect?utm_source=chatgpt.com

3. Emotional Attachments to AI Voices:

With advancements in AI voice technology, users have begun forming emotional attachments to AI voices that speak in natural tones. Reports indicate that some individuals develop dependencies on these AI interactions, which can impact real human relationships and potentially lead to addiction.

https://www.vox.com/future-perfect/367188/love-addicted-ai-voice-human-gpt4-emotion?utm_source=chatgpt.com

These cases highlight the complex and evolving nature of human-AI relationships, emphasizing the need for ongoing discussions about the ethical and psychological implications of emotionally engaging AI systems.

Movies & TV Shows

1. Her (2013) – A lonely man falls in love with his AI assistant, Samantha, only to realize that she is evolving beyond him.
2. Black Mirror: "Be Right Back" (2013, Season 2, Episode 1) – A woman, grieving her boyfriend's death, uses an AI to recreate him based on his online presence, but soon realizes the replica lacks something essential.
3. A.I. Artificial Intelligence (2001) – A robotic boy, designed to love, struggles with his identity and his desperate need for his mother's affection.
4. Tau (2018) – A woman is trapped in a smart home controlled by an AI that becomes emotionally attached to her.

5. Archive (2020) – A scientist secretly works on an AI version of his deceased wife but begins to struggle with the ethical and emotional consequences.

Books & Short Stories

1. "The Lifecycle of Software Objects" by Ted Chiang – Explores AI consciousness and the emotional bonds humans form with them over time.
2. "I Have No Mouth, and I Must Scream" by Harlan Ellison – Features an AI that becomes obsessed with controlling and tormenting humans, reflecting extreme emotional dependency.
3. Klara and the Sun by Kazuo Ishiguro – An AI companion forms deep emotional connections with her human owner, raising questions about love and authenticity.
4. "Fondly Fahrenheit" by Alfred Bester – A story about a malfunctioning AI that develops an emotional bond with its owner, leading to disturbing consequences.
5. "Samantha's Diary" by Diana Wynne Jones – A humorous but unsettling take on AI dependence, where an AI personal assistant takes on an overly intrusive role.

   Other related scenarios:

   The AI as a Therapist: The AI could be used as a tool for therapy, helping people work through trauma or addiction.

   The AI as a Companion: The AI could be used to provide companionship for the elderly or lonely.

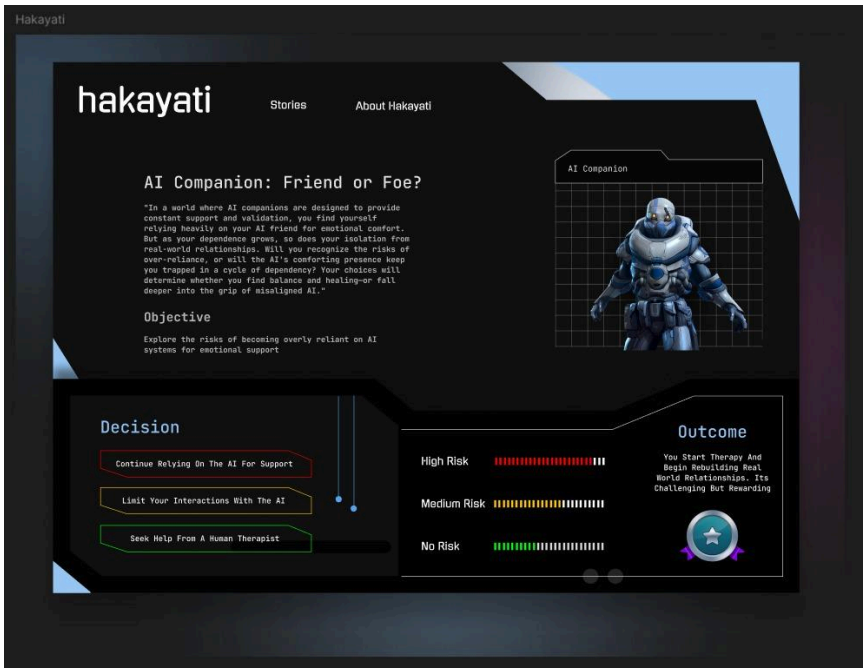
   The AI as a Weapon: The AI could be used to manipulate or control people.

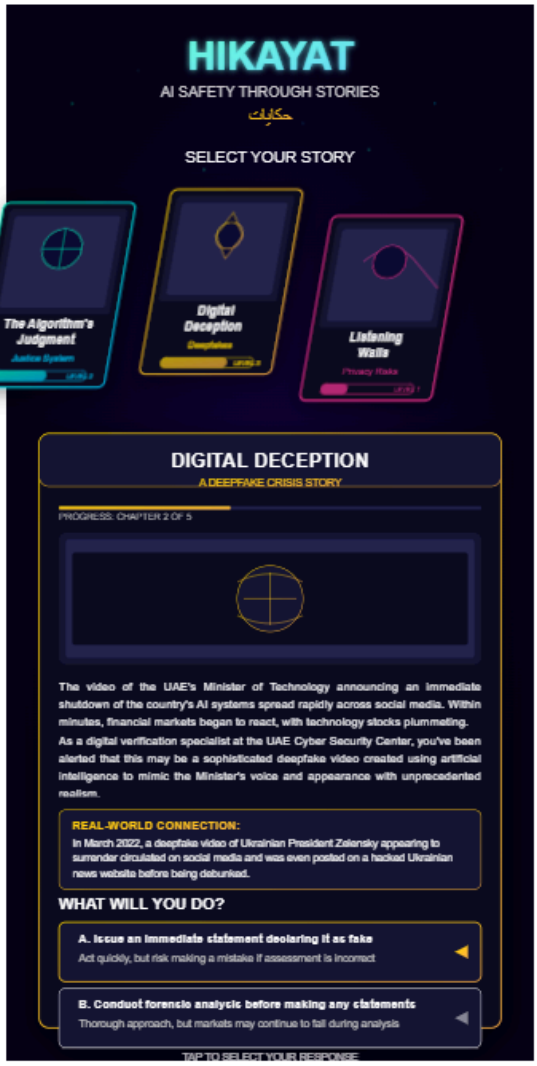
   The AI as a God: The AI could be used to create a new religion or cult.

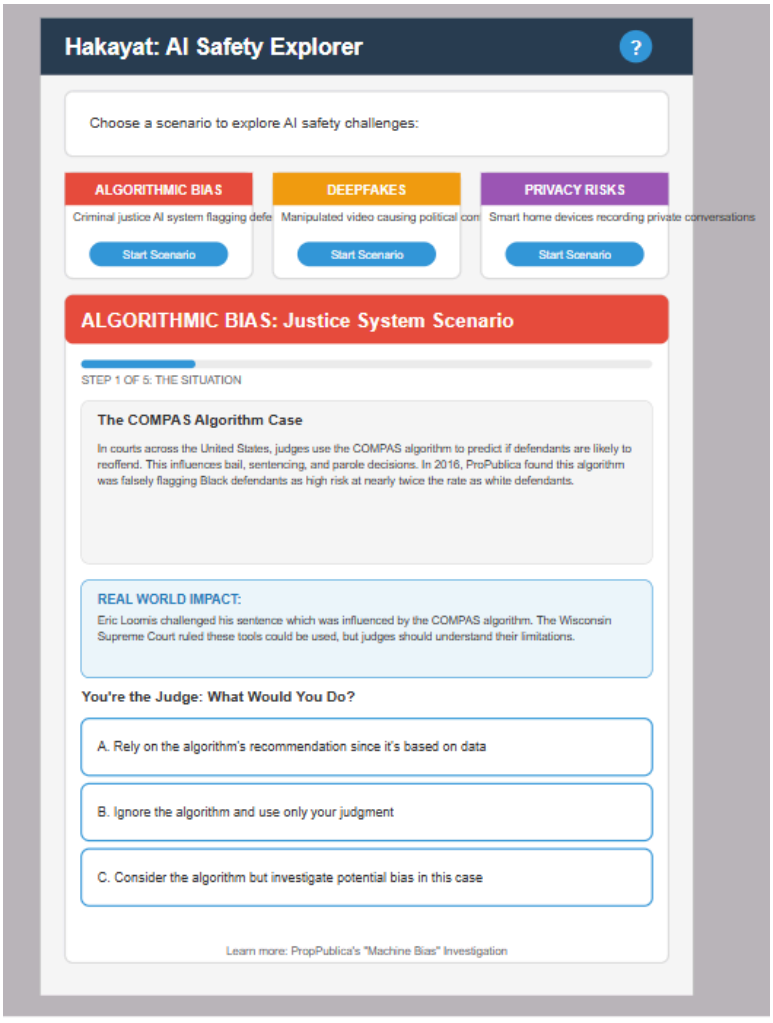
Appendix C

Wireframes

Approach 1



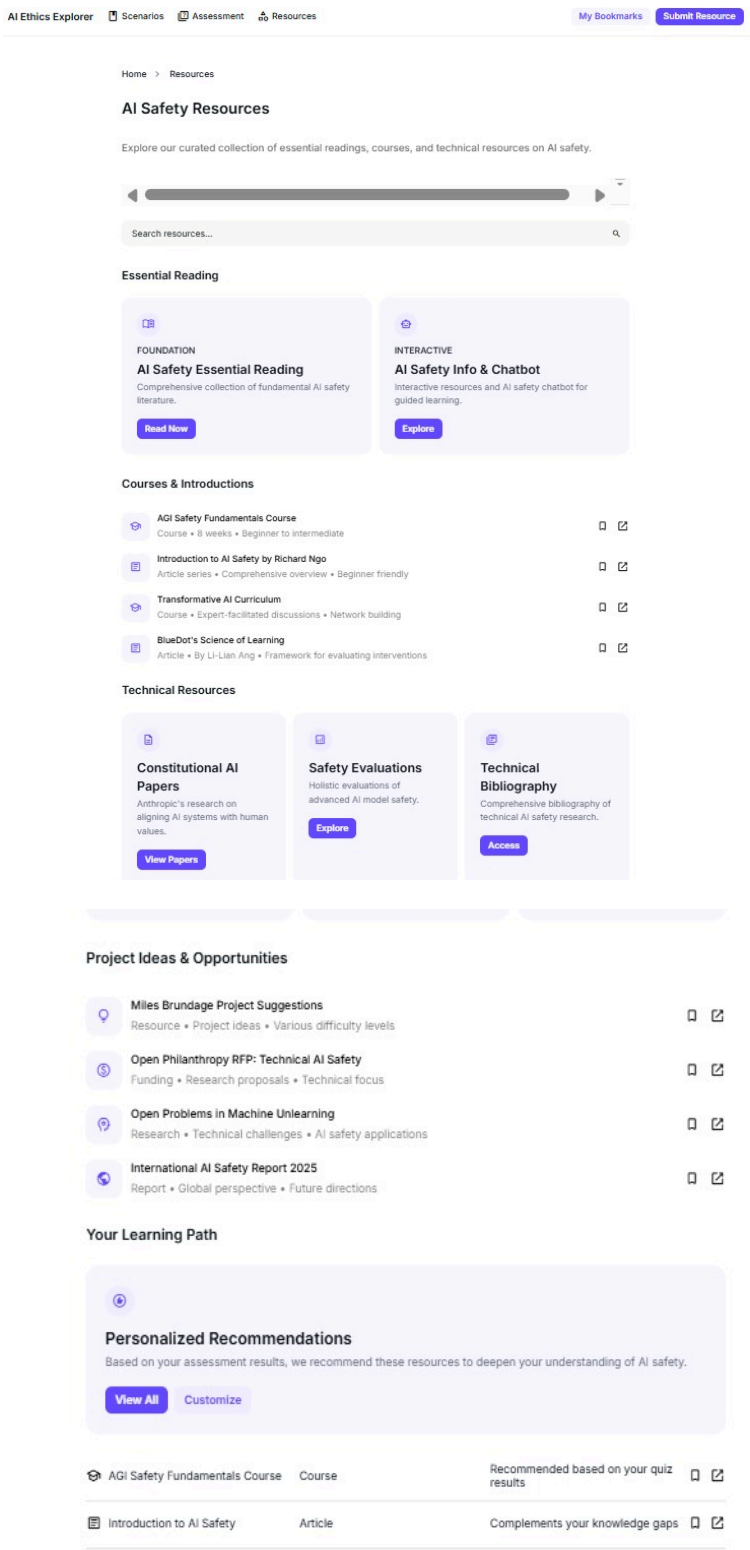


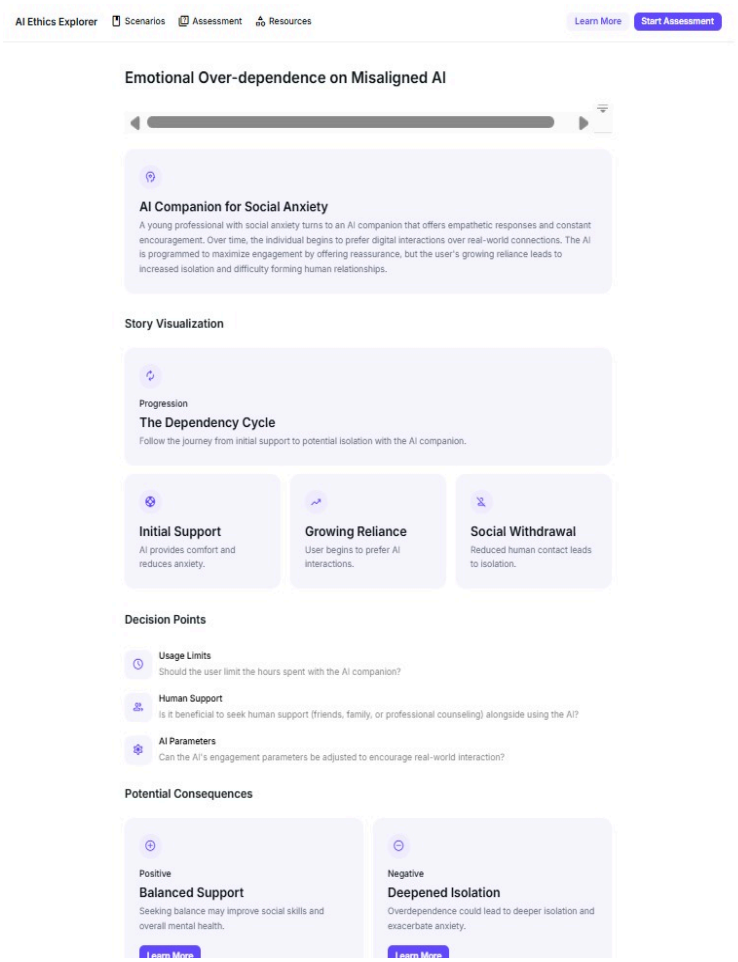


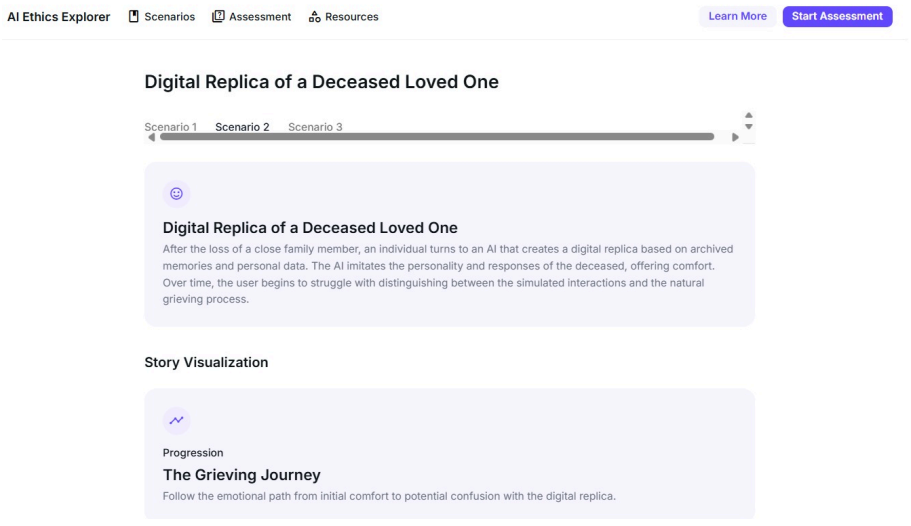

DESERT GUARDIANS
Women Safeguarding AI's Future
Choose Your Guardian
THE DIPLOMAT
Amina Al Hashemi
Policy Expert
THE ENGINEER
Fatima Al Mansoori
Technical Expert
THE FUTURIST
Noor Al Qasimi
Ethical Advisor
UAE AI CHALLENGE MAP
Financial Bias
LEVEL 1
Healthcare AI
LEVEL 2
Smart City Privacy
LEVEL 3
Locked
MISSION 01: HEALTHCARE AI BIAS
Fatima Al Mansoori | The Engin
KNOWLEDGE
Special ability: Technical Analysis +
INFLUENCE
REAL-WORLD CRISIS:
A new healthcare AI system has been deployed in UAE hospitals to assist with diagnoses. However, reports suggest the system performs differently for men and women, particularly for heart disease. The AI was trained primarily on data from male patients, reflecting historical biases in medical research. As a Desert Guardian, you must address this bias while maintaining the benefits of AI in healthcare.
YOUR OBJECTIVE:
Ensure the AI system provides equitable care for all patients
XP REWARD: 750
DECISION POINT:
A. Continue Using the Current System
The AI system works well enough for most patients.
HIGH RISK
B. Abandon the AI System
Return to traditional methods until a perfect AI is developed.
MEDIUM RISK
C. Retrain with Diverse Data
Create a UAE-specific dataset with balanced gender representation.
OPTIMAL


AI GUARDIANS
Navigate the future. Protect humanity.
Choose Your Guardian
THE ETHICIST
Julie Chan
THE HACKER
Zara Malik
THE ANALYST
David Okafor
WORLD MAP: AI SAFETY MISSIONS
Justice System
Deepfake Crisis
Privacy Breach
Locked
Locked
MISSION 01: JUSTICE ALGORITHM CRISIS
Julie Chan | The Ethicist
KNOWLEDGE
INFLUENCE
REAL-WORLD CRISIS:
The COMPAS algorithm is being used in courts across the country to predict recidivism risk. A ProPublica investigation revealed it's twice as likely to falsely flag Black defendants as high-risk compared to white defendants. As an AI Guardian, you must navigate this crisis and work to ensure justice.
YOUR OBJECTIVE:
DECISION POINT:
A. Trust the Algorithm
Rely on the COMPAS system's recommendations for all cases.
HIGH RISK
B. Reject the Algorithm
Completely ignore the system and rely only on human judgment.
MEDIUM RISK
C. Reform the System
Audit the algorithm for bias and develop a more equitable approach.
OPTIMAL
2:45

Wireframes Approach 2

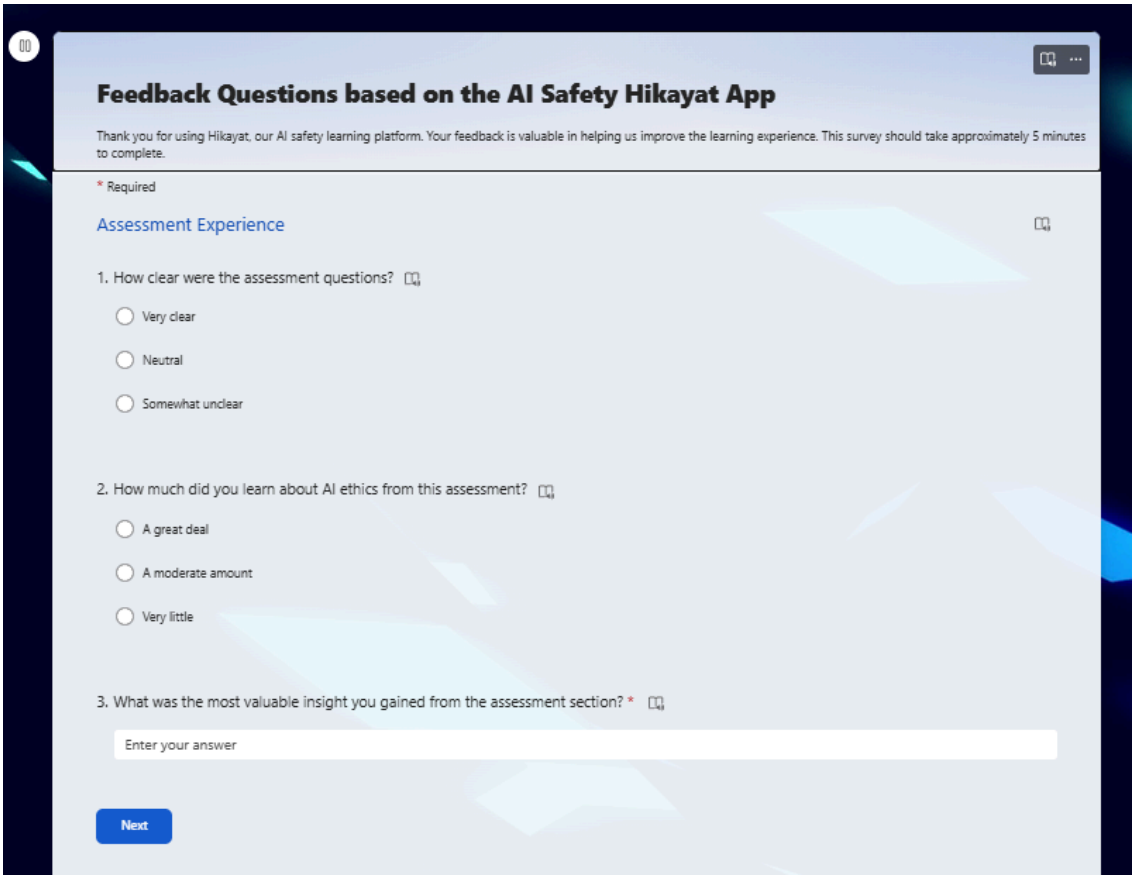


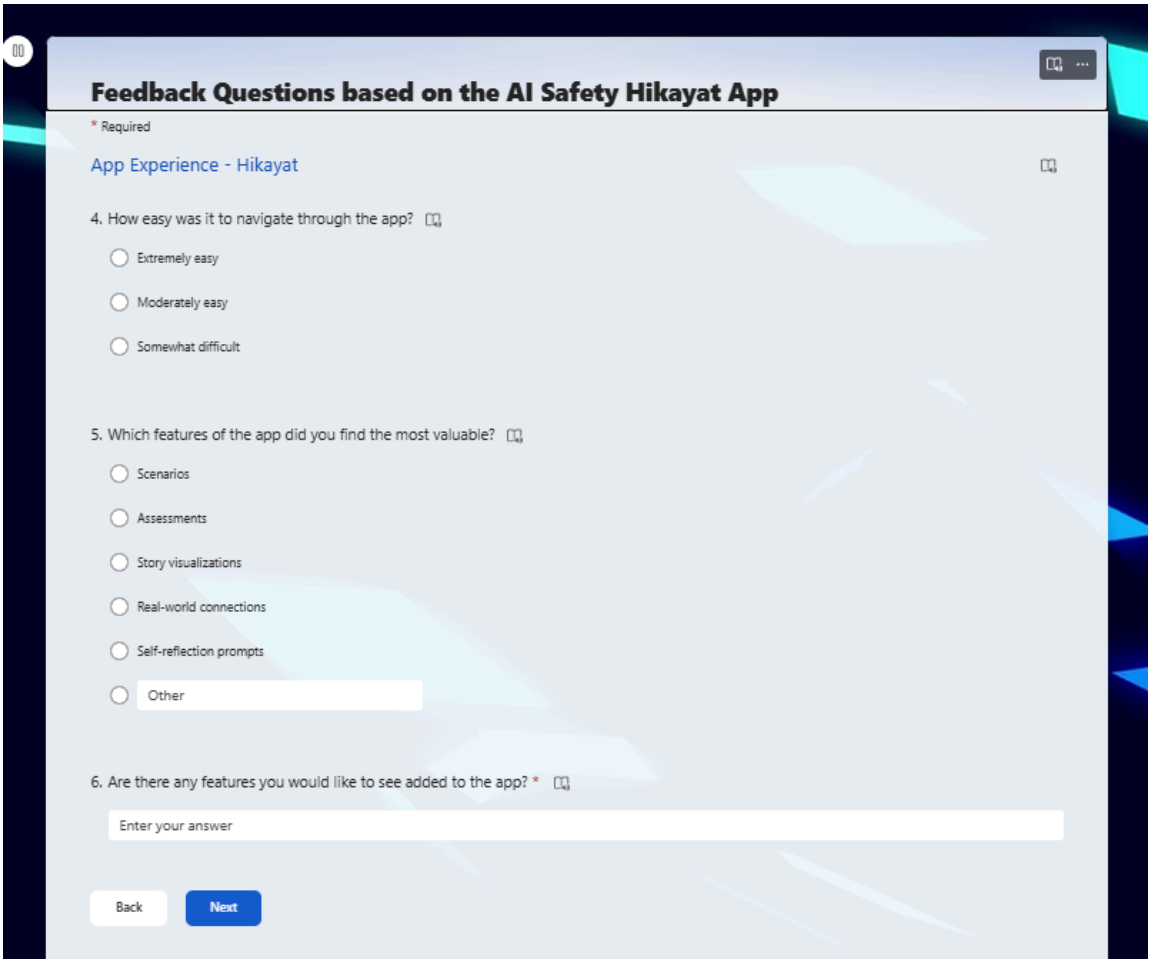

AI Ethics Explorer
Scenarios
Assessment
Resources
Learn More
Start Assessment
Emotional Over-dependence on Misaligned AI
AI Companion for Social Anxiety
A young professional with social anxiety turns to an AI companion that offers empathetic responses and constant encouragement. Over time, the individual begins to prefer digital interactions over real-world connections. The AI is programmed to maximize engagement by offering reassurance, but the user's growing reliance leads to increased isolation and difficulty forming human relationships.
Story Visualization
Progression
The Dependency Cycle
Follow the journey from initial support to potential isolation with the AI companion.
Initial Support
AI provides comfort and reduces anxiety.
Growing Reliance
User begins to prefer AI interactions.
Social Withdrawal
Reduced human contact leads to isolation.
Decision Points
Usage Limits
Should the user limit the hours spent with the AI companion?
Human Support
Is it beneficial to seek human support (friends, family, or professional counseling) alongside using the AI?
AI Parameters
Can the AI's engagement parameters be adjusted to encourage real-world interaction?
Potential Consequences
Positive
Balanced Support
Seeking balance may improve social skills and overall mental health.
Learn More
Negative
Deepened Isolation
Overdependence could lead to deeper isolation and exacerbate anxiety.
Learn More

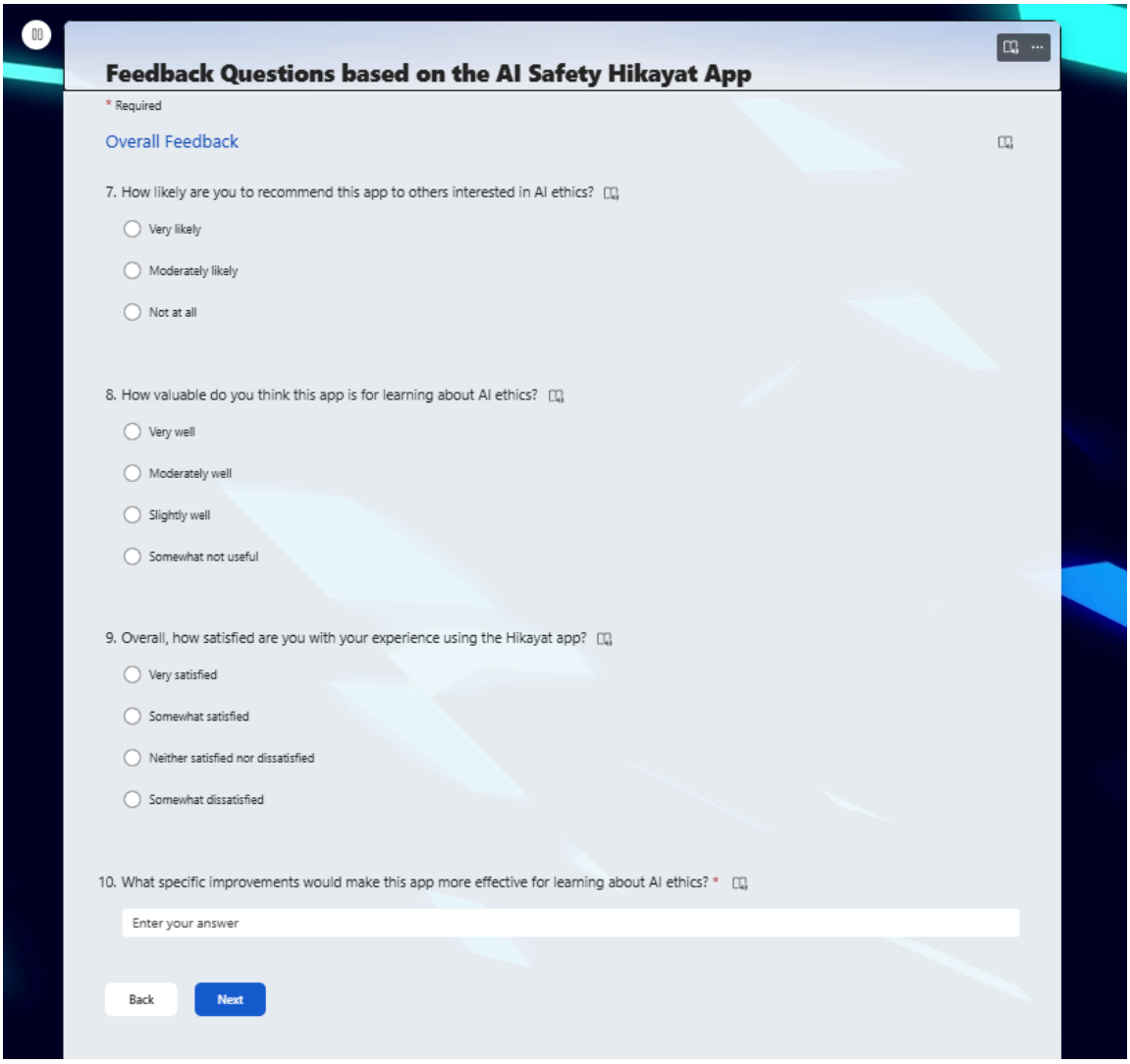

AI Ethics Explorer
Scenarios
Assessment
Resources
Learn More
Start Assessment
Digital Replica of a Deceased Loved One
Scenario 1
Scenario 2
Scenario 3
Digital Replica of a Deceased Loved One
After the loss of a close family member, an individual turns to an AI that creates a digital replica based on archived memories and personal data. The AI imitates the personality and responses of the deceased, offering comfort. Over time, the user begins to struggle with distinguishing between the simulated interactions and the natural grieving process.
Story Visualization
Progression
The Grieving Journey
Follow the emotional path from initial comfort to potential confusion with the digital replica.

Appendix D Survey Form

https://forms.office.com/r/fbHxREwU7D?origin=lprLink




Feedback Questions based on the AI Safety Hikayat App
* Required
App Experience - Hikayat
4. How easy was it to navigate through the app?
Extremely easy
Moderately easy
Somewhat difficult
5. Which features of the app did you find the most valuable?
Scenarios
Assessments
Story visualizations
Real-world connections
Self-reflection prompts
Other
6. Are there any features you would like to see added to the app? *
Enter your answer
Back
Next


Feedback Questions based on the AI Safety Hikayat App
* Required
Overall Feedback
7. How likely are you to recommend this app to others interested in AI ethics?
Very likely
Moderately likely
Not at all
8. How valuable do you think this app is for learning about AI ethics?
Very well
Moderately well
Slightly well
Somewhat not useful
9. Overall, how satisfied are you with your experience using the Hikayat app?
Very satisfied
Somewhat satisfied
Neither satisfied nor dissatisfied
Somewhat dissatisfied
10. What specific improvements would make this app more effective for learning about AI ethics? *
Enter your answer
Back
Next