2025

Societal Impact Report
NPC AI Integration

PUBLICATIEDATUM: 05/12/2025

RUBEN VAN DE RANDE - AI FOR SOCIETY MINOR FONTYS EINDHOVEN

Index

1. Introduction

This document provides a societal impact assessment of the AI-NPC system developed for my personal project, where a locally running language model enhanced with a custom LoRA layer is integrated into a game character. The goal is to create an NPC that behaves naturally within the game world while ensuring compliance with the GDPR and the EU AI Act. The report outlines the system's purpose, data practices, ethical considerations, and potential risks, and reflects on how this technology can be used responsibly in interactive environments.

2. System Overview & Intended Use

2.1 Primary Overview

The AI-NPC system is a prototype conversational character designed for use inside a game world. It is powered by a locally running large language model (meta-llama-3-8b-instruct [3]) enhanced with a custom LoRA training layer. This LoRA layer was trained on synthetic, game-specific text to ensure the NPC naturally behaves as if it exists within the game's lore, without requiring an external personality prompt at runtime.

After merging the LoRA layer with the base model, the AI was integrated into a Unity NPC framework. Players can freely interact with the character through text-based dialogue, and the NPC responds based on its trained in-world knowledge. A lightweight memory system allows the NPC to recall parts of the ongoing conversation to improve continuity and immersion.

The primary purpose of the system is to explore how narrative-rich, contextual AI can enhance player experience by creating characters that behave consistently with the game's setting. This project also serves as an investigation into the technical and ethical implications of embedding AI models directly into interactive environments.

3. Data, Privacy & GDPR Compliance

3.1 Data Sources

The AI-NPC is trained using a LoRA layer created from synthetic datasets and game-lore text produced specifically for this project. All training material is fictional and does not contain real-world personal data. This approach ensures that the model's behaviour is fully derived from controlled, in-game context rather than external or user-derived information, reducing privacy risks and preventing unintended real-world associations.

3.1 Avoidance of Personal Data

During normal gameplay, the NPC does not process or store any personal data about players. Interactions are limited to in-game dialogue, and the system is intentionally designed so that players cannot provide identifying information to the model in a meaningful or persistent way. This aligns with the GDPR's goal of preventing unnecessary data collection, and reduces the risk of the AI generating outputs based on sensitive or private details. [1] [2]

3.1 Relevant GDPR Principles

Several GDPR principles guide the design of the AI-NPC system:

- **Article 5 - Data minimisation & purpose limitation:**
  The system collects no personal data and only processes the immediate dialogue input necessary to generate a response. [1] [2]
- **Article 6 - Lawful basis:**
  Since the system does not store or use personal data, a separate lawful processing basis is not required. All data handled by the model is non-personal and transient. [1] [2]
- **Article 25 - Data protection by design and by default:**
  Privacy is ensured structurally by relying on synthetic data, avoiding external datasets, and preventing long-term storage of player inputs. These measures minimise the possibility of identifying or linking information back to individuals. [1] [2]

3.1 Memory System Limitations and Handling

The NPC includes a short-term memory module to improve conversational continuity. This memory only holds temporary dialogue context and automatically clears when the interaction ends. No logs are stored, exported, or reused across sessions. This prevents profiling, behavioural tracking, or indirect identification of players.
By design, the memory function adheres to GDPR expectations for data minimisation and storage limitation (Art. 5), reducing any long-term privacy exposure.

4. AI Act Risk Level, Safety & Ethical Considerations

4.1 EU AI Act Classification

Under the EU AI Act, the AI-NPC system falls into the category of a Limited-Risk AI System, as it is used for entertainment and narrative purposes within a controlled game environment. According to Article 52, such systems must meet basic transparency obligations and players should be aware that they are interacting with an AI. In this project, the NPC is clearly presented as an AI character, and its function is limited to dialogue within the fictional game world. Since the system is not used for decision-making, profiling, or safety-critical tasks, no high-risk requirements apply.

4.1 Player Impact, Over Trust, and Manipulation Risks

Although the system operates in a fictional setting, AI-driven dialogue can influence how players perceive and emotionally respond to characters. Risks include over-trust, where players may attribute more intelligence or authority to the NPC than intended or undesired persuasive behaviour.

To reduce these risks, the NPC's capabilities are intentionally constrained, its role is narrative rather than advisory, and its interactions stay strictly within gamelore boundaries. The design avoids realworld instructions, sensitive topics, or behaviour that could manipulate players.

4.1 Bias Risks in Training Data

Because the NPC's understanding of the world is derived entirely from synthetic and game-specific training data, it avoids inheriting biases from large external datasets. However, there is still a risk that subtle biases or unintended behaviours emerge from poorly balanced or repetitive synthetic text.

To mitigate this, training data is manually reviewed, and future iterations will use improved and more diverse synthetic datasets to ensure consistent, neutral, and lore accurate responses.

4.1 Safety Mitigations and Design Constraints

Several safeguards are built into the system to prevent harmful behaviour:

- The model operates with restricted context, preventing long-term influence or personalisation.
- The memory module is strictly temporary, reducing chances of profiling or emotional dependency.
- Dialogue generation is constrained by in-world rules, limiting the AI's ability to provide real-world advice or sensitive content.
- Developers maintain the ability to review, test, and adjust model outputs as needed.

These measures ensure the NPC remains aligned with both the AI Act's safety expectations and ethical considerations for responsible AI in interactive environments.

5. Societal Impact & Future Improvements

5.1 Positive Impact

The AI-NPC system demonstrates how locally running language models can enhance immersion, narrative depth, and player engagement in games. By allowing NPCs to respond in a world aware and consistent manner, the system offers a more dynamic and believable storytelling experience.

Beyond entertainment, the project also has educational value, illustrating how synthetic data, LoRA fine-tuning, and responsible AI practices can be applied in interactive digital environments. This contributes to a broader understanding of how AI can be integrated in creative and technically innovative ways.

5.1 Environmental and Technical Considerations

Running the model locally reduces dependence on cloud services, which can help lower the energy footprint associated with large-scale inference. However, local inference still requires significant computational resources, especially when relying on CPU-based processing.

Future GPU optimisation will improve energy efficiency by reducing inference time and overall hardware strain. This reflects a balance between performance, accessibility, and sustainability.

5.1 Future Improvements

Several enhancements are planned to improve both functionality and responsibility:

- **GPU optimisation:** enabling faster inference and smoother real-time interaction.
- **Dataset refinement:** generating higher-quality synthetic training data to improve consistency, reduce bias, and reinforce world-aware behaviour.
- **Safer behaviour policies:** adding clearer constraints to prevent inappropriate or unintended content.
- **Transparency improvements:** ensuring players always understand they are interacting with an AI, complying with the EU AI Act's expectations for limited-risk systems.

These developments aim to strengthen the system's reliability, user experience, and ethical alignment as the project continues.

6. Conclusion

The AI-NPC system demonstrates how a locally running, LoRA-enhanced language model can create immersive and context-aware interactions inside a game world while remaining aligned with GDPR and the EU AI Act. By relying on synthetic data, limiting memory, and operating transparently as an AI character, the system avoids unnecessary personal data processing and maintains a low-risk profile. Although still in development, the project highlights both the creative potential and ethical responsibility involved in bringing AI-driven characters into interactive environments, and provides a foundation for further technical, narrative, and safety improvements.

7. Sources

[1] European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2016. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

[2] European Union, Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (EU AI Act), 2024. Available: https://eur-lex.europa.eu/

[3] Meta, Llama 3 Model Card and Documentation, Meta AI, 2024. Available: https://ai.meta.com/llama

[4] E. Hu et al., “LoRA: Low-Rank Adaptation of Large Language Models,” 2021. Available: https://arxiv.org/abs/2106.09685

[5] Unity Technologies, Unity Documentation: Runtime AI Integration Guidelines, Unity, 2024. Available: https://docs.unity.com/