DEVELOPMENT OF AN LLM-INTEGRATED GAME ARCHITECTURE

FENS
Faculty of Engineering and Natural Sciences
International University of Sarajevo

Amina Hrustić, Berina Mesihović, Hamza Selimović,
Hamza Adilović, Lamija Kožo
Mentor: Dr. Kanita Karađuzović-Hadžiabdić,
Co-mentor: Prof. Dr. Migdat Hodžić, Co-mentor: Roman Sulejmanpašić, MFA

IUS
International University of Sarajevo

1 Problem & Objectives

The project addresses a combined design and engineering challenge: how to build a coherent narrative game loop that blends traditional gameplay systems (movement, traversal, interaction, UI) with AI-supported dialogue that remains contextually relevant, character-consistent, and measurable in quality.

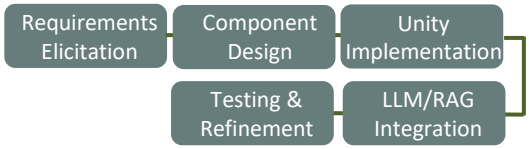
From a software engineering perspective, the challenge includes maintaining system responsiveness, ensuring stable scene and state transitions, and producing a documentation set that satisfies formal graduation reporting requirements.

Objectives:

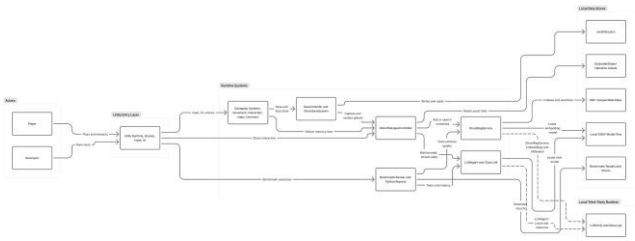
- Deliver a self-contained playable demo with a complete short narrative arc.
- Implement stable movement, sprint-stamina behavior, pause flow, and map transitions.
- Design and integrate ghost dialogue architecture using a shared LLM and NPC-specific RAG context.
- Define measurable performance and AI quality indicators for validation.
- Produce full graduation documentation aligned with the FENS handbook and thesis guidelines.

2 Methodology & Design

This project follows a prototype-based and iterative development methodology, where the playable Unity demo acts as both the engineering output and the research instrument, starting from requirements and user stories, then moving into modular component design, Unity implementation, and validation. Gameplay systems are kept deterministic, while LLM/RAG dialogue is integrated separately to support context-aware conversations.



The design combines traditional Unity game logic with a local LLM dialogue system. Deterministic scripts control movement, inventory, quest progression, item delivery, and save/load behavior, while the LLM is only responsible for generating natural NPC responses. Ghost identities, quest stages, and memory items are stored as ScriptableObject assets, and recovered memories are indexed per NPC so the model can only use information unlocked by the player. This design preserves narrative control while allowing more dynamic and immersive conversations.



The diagram above shows the full system architecture, including the game and the benchmarking.

3 Validation & Results

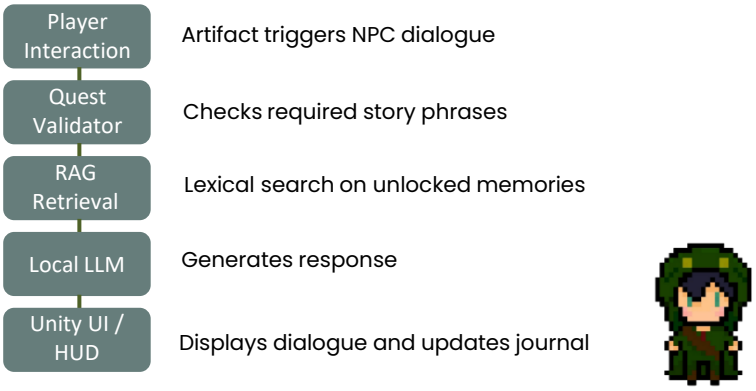
Benchmark Configuration Selected:

- Model: Qwen3.5-4B-Q4_K_M (no-reasoning)
- Retrieval: Lexical RAG(LLMUnity DBSearch)
- Validation: Quest phrase safety layer
- Hardware: Laptops, lower-end PCs

Best balance of response speed, memory usage, retrieval accuracy, and quest-safe dialogue, compared with 14 other models.

Functional tests:

- **TC-01–TC-06:** movement, pause, transitions, save/load, interaction
- **TC-10–TC-13**: footstep SFX, item pickup, hotbar use, LLM ghost dialogue



4 Conclusion & Future Work

Retracted Ocean demonstrates that local AI dialogue can be integrated into a Unity narrative game in a controlled, measurable way. The key contributions are authored memories, retrieval gating, validation and acceptance criteria.

Future work:

- Integrated Unity Play Mode latency testing
- Expanded NPC memory authoring and quest branching
- Complete QA evidence from real play sessions
- Latency optimization

1. Unity Documentation: Input System, Cinemachine, and Universal Render Pipeline.
2. Asai, A., et al. (2023). Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection.
3. ggml-org. "llama.cpp GitHub repository". https://github.com/ggml-org/llama.cpp
4. unum-cloud. "USearch GitHub repository". https://github.com/unum-cloud/usearch
5. Unity Technologies. "ScriptableObject documentation" https://docs.unity3d.com/Manual/class-ScriptableObject.html
6. Unity Technologies. "Application.persistentDataPath scripting documentation" https://docs.unity3d.com/ScriptReference/Application-persistentDataPath.html

12 RESPONSIBLE CONSUMPTION AND PRODUCTION
9 INDUSTRIE, INNOVATION UND INFRASTRUKTUR