Memories for Virtual AI Characters

Fabian Landwehr, Erika Varis Doggett, Romann M. Weber

fabian.landwehr@inf.ethz.ch, erikavaris@gmail.com, romann.weber@gamil.com

1. Introduction

- We present a system for augmenting LLM-based virtual characters with long-term memories, enabling them to remember facts about themselves, their world, and past experiences.
- A *memory-creation pipeline* converts raw text into condensed memories and a *memory-retrieval system* utilizes these memories to generate character responses.
- Using a fact-checking pipeline based on GPT-4, our evaluation demonstrates that the character responses are grounded in the retrieved memories and maintain factual accuracy.

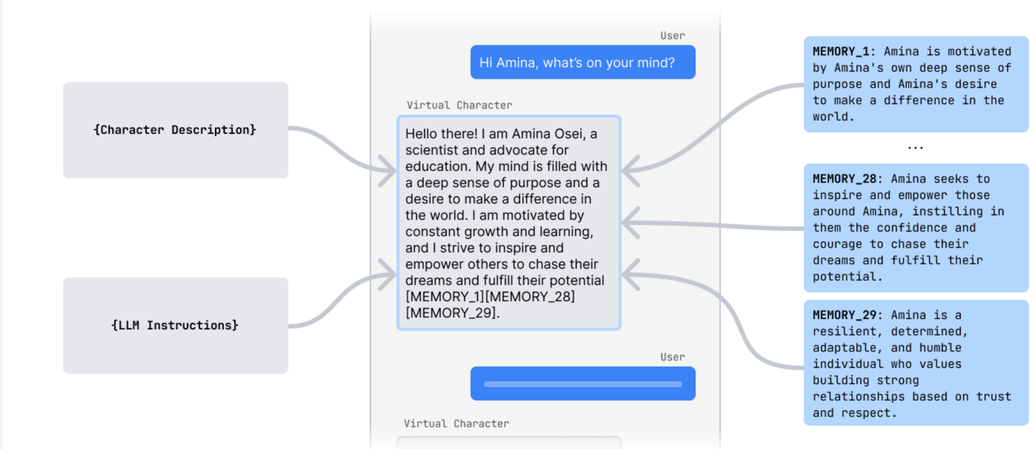
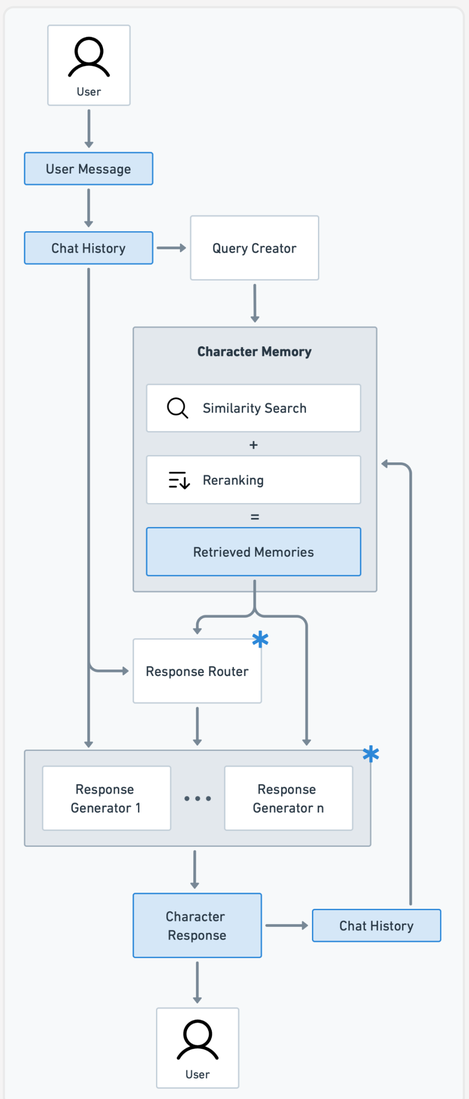


Figure 1: *Memories for virtual AI characters:* This figure illustrates how virtual character responses are generated. When answering, characters have access to a list of relevant memories, a description of themselves, and the chat history. Section 3 explains the proposed system in detail.

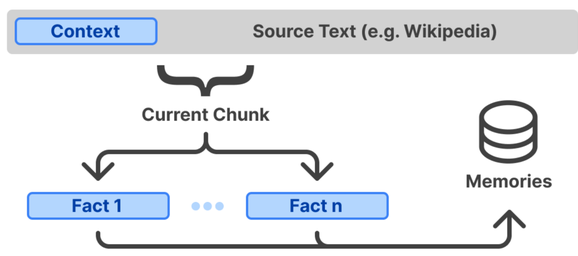
2. Proposed System



- When the user sends a message to the virtual character, a search query for the needed information is generated.
- Using this query, relevant memories are retrieved by doing multiple similarity searches and re-ranking the results.
- The re-ranking step allows for customization. E.g., memories are prioritized based on a forgetting function.
- The response router chooses the appropriate response generation strategy based on the context.
- Finally, an LLM (GPT 3.5) is used to generate the character response.
- Interactions are later synthesized into new memories.

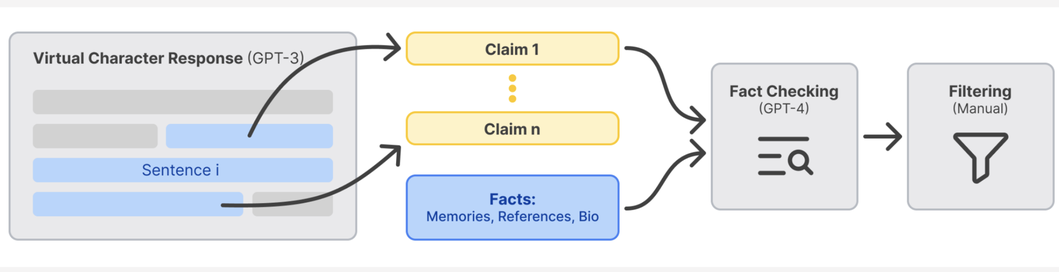
3. Memory Creation

- Memories can be created from any raw text:
  - The source text is split into chunks. Each chunk is processed together with a summary of all previous chunks.
  - Given a (chunk, context) pair, important facts/observations are extracted using an LLM.
  - Each fact/observation forms the basis of a new memory.
- Memory content: fact/observation, embedding, meta info.



4. Evaluation

- We evaluate the groundedness of the character responses and whether the provided references are accurate.
- To do so, we use a fact-checking pipeline based on GPT-4:
  - A given character response is split into individual sentences.
  - From each sentence, one or more claims are extracted.
  - Each claim is checked against the information that was available to the LLM when generating the character response.



- Results suggest that the character responses are largely grounded in memories, preserving integrity.
- However, the provided references were partially incorrect.
- Results were better when the LLM had no intrinsic knowledge about the simulated virtual character.

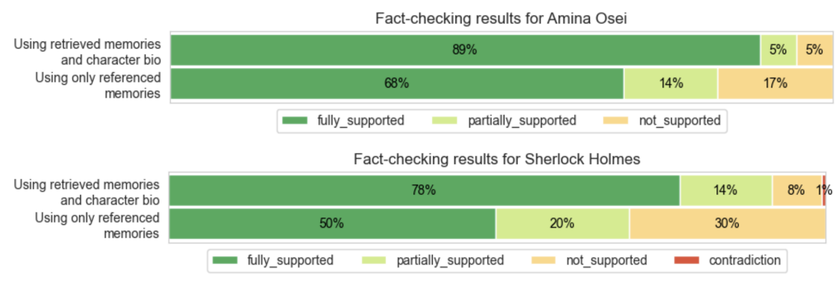


Figure 5: *Fact-checking results for the virtual character responses:* Each horizontal bar shows the results for different sources of truth as named on the left. The first category shows how grounded the character response is in the available information, using an aggregation of all three fact-checking results. The second category shows how good the LLM is at referencing which information it uses. Note that some percentage counts do not add up to 100 due to rounding.

Paper
https://sigdialinlg2023.github.io/paper_inlg99.html

INLG 2023