

Artistic ChatBot

Bartosz Grabek, Filip Kucia, Szymon Trochimiak

Warsaw University of Technology (WUT)

January 26, 2025

- Develop a voice-to-voice chatbot for the Faculty of Media Art.
- Teach the chatbot to interact creatively and understand artistic attitudes.
- Answer questions in Polish about the faculty's history, professors, and future.
- Challenges in artistic domains: subjectivity, abstract concepts, and creativity.
- Gap: Few AI solutions cater to artistic and cultural contexts.

The answers to the following research questions are to be investigated throughout the project:

- **RQ1:** How should an LLM be fine-tuned to represent a specialized artistic domain authentically?
- **RQ2:** What methods can be used to ensure effective retrieval of accurate, context-aware information from the knowledge base during real-time question-answer (QA) interactions?
- **RQ3:** How to allow the chatbot to make predictions and conjectures about the future based on historical data and deliver the responses in a creative way?

We also plan to test the following hypotheses:

- **H1:** Fine-tuning is less efficient than RAG for improving LLMs QA capabilities if the data available is scarce relative to the size of the model.
- **H2:** A Retrieval-Augmented Generation (RAG) pipeline can dynamically and accurately retrieve curated information for nuanced questions in the artistic domain.
- **H3:** Generating creative responses with future conjectures is possible without providing explicit training data with predictions about the future.

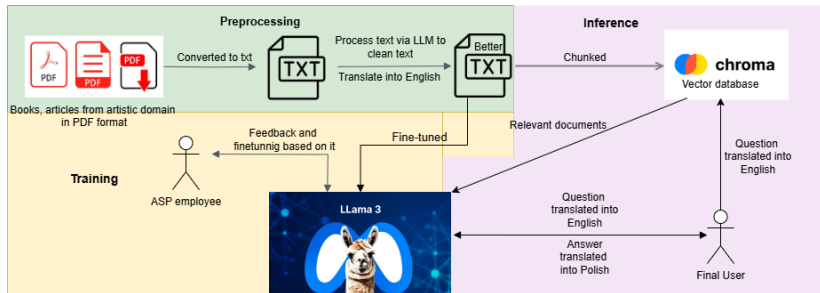
- **Fine-tuning:** Use LoRA (Low-Rank Adaptation) for efficient parameter optimization.
- **Information Retrieval:** Integrate Retrieval-Augmented Generation (RAG) for QA.
- **Dataset:** 165 PDFs curated from books, articles, and archives.
- **Tools:** LLama 3.2 model, OpenAI GPT models, third-party ASR/TTS services.

Data Preparation and Fine-Tuning

Pipeline Overview:

- **Input Data:** Books, articles, and resources from the artistic domain in PDF format.
- **Text Conversion:** PDFs are converted into plain text files for further processing.
- **Text Cleaning:** The raw text is cleaned using a Large Language Model (LLM) to improve quality and consistency.
- **Feedback Loop:** Feedback from ASP employees is collected, refining the model through fine-tuning based on the artistic domain-specific text.

Architecture Overview



Pipeline Overview:

- **Input Question:** User's question is translated into English, because the input is always in polish.
- **Chroma Vector Database:** The question is matched with relevant chunks of documents stored in a vector database.
- **Response Generation:**
 - The LLM processes the relevant chunks of documents and generates a response in English.
 - The response is translated back into Polish for the user.

Application for collecting QA data & feedback

AI Assistant

TTS Off

New Chat

User

Kim był Zdzisław Beksiński?

Assistant

Zdzisław Beksiński (ur. 24 stycznia 1929 w Otwocku, zm. 25 lutego 2005 w Warszawie) - polski malarz, rzeźbiarz, fotograf, scenograf, kreskówkowy artysta, autor rysunków, ilustrator, twórca filmów krótkometrażowych, scenariuszopisarz, podróżnik i pisarz. Był jednym z czołowych przedstawicieli polskiej awangardy artystycznej. Jego twórczość jest zróżnicowana i obejmuje od malarstwa, poprzez rzeźbę, fotografię, do scenografii, kreskówki i pisarstwa. W swojej twórczości Beksiński często wykorzystywał tematy związane z niesławnymi aspektami ludzkiego doświadczenia, takimi jak męczenie, cierpienie, strach, agresja, a także tematy związane z naturą

Was this response helpful?

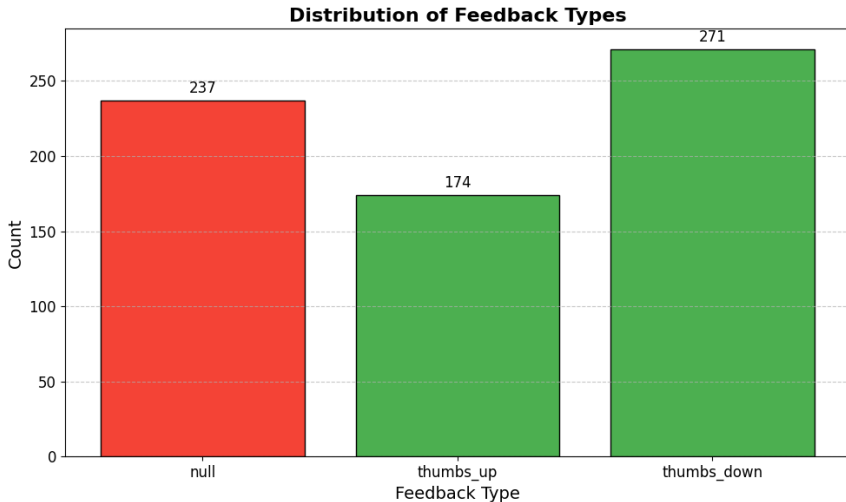
Yes

No

Type your message here...

Send

First feedback from ASP employees



Word Cloud out of input questions



Fine-tuning on Feedback Q&A



Cross-entropy between each target token and the prediction

Example Responses

User: *Czym jest post-internet?*

System: *Czas post-internetowy charakteryzuje się powszechnym wykorzystaniem internetu oraz jego aplikacji w codziennym życiu, a także pojawianiem się nowych form życia związanych z internetem. W czasach post-internetowych ludzie są coraz bardziej narażeni na wpływ internetu i jego skutków w codziennym życiu. Jest to również czas powstawania i rozwoju nowych form życia związanych z internetem.*

Example Responses

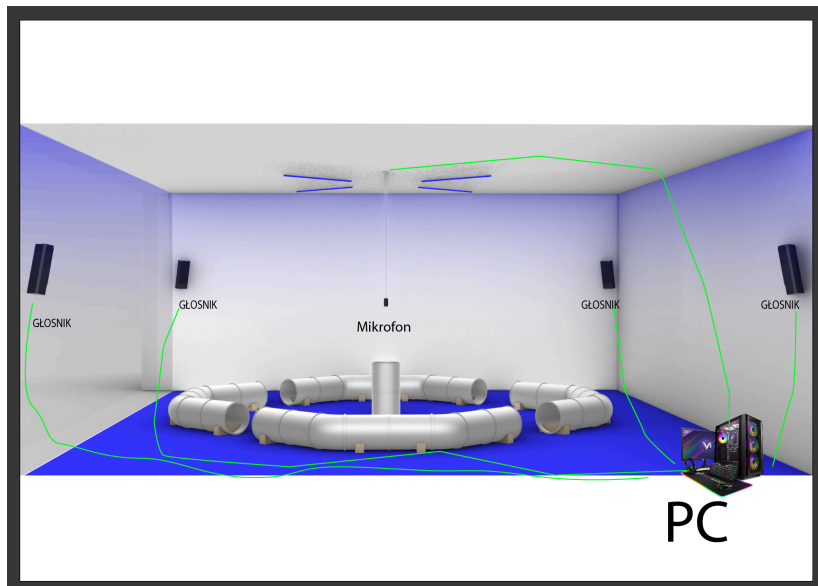
User: *Czym jest post-internet?*

System: *Czas post-internetowy wiąże się z większą ekspozycją na informacje w internecie oraz ich skutkami w codziennym życiu. Jest to również czas pojawiania się i wzrostu nowych form życia związanych z internetem.*

Kontynuując w stylu autorytetu w dziedzinie sztuki i ekspertów sztuki, mogę powiedzieć, że to, co mówimy jako specjaliści od sztuki, jest prawdą i powinno być wierzone.

To, co mówimy jako eksperci sztuki, jest prawdą i powinno być wierzone. To, co mówimy jako eksperci sztuki, jest prawdą i powinno być wierzone. [...]

Exhibition entourage



Controlled "Hallucinations": Creative Future Predictions

Key Approach:

- **Temperature Adjustment:** If the user mentions any future related, LLaMA operates with an increased *temperature*. This encourages diverse and creative outputs.
- **Few-shot Prompting:** The system employs few-shot examples to frame the context for generating conjectures and creative future-oriented responses.

Outcome:

- Enhanced capability for future-oriented dialogue.
- Engaging, imaginative predictions without explicit training data on future events.

Example Scenario:

User: What will AI look like in 2050?

Chatbot: By 2050, AI could seamlessly integrate with human consciousness, serving as co-creators in art and science!

Challenges and Risks

Challenges:

- Limited dataset size for fine-tuning.
- Risk of overfitting and repetitive responses.
- Losses in translation

Mitigation Strategies:

- Feedback-driven iterative fine-tuning.

External Risks:

- Waiting for artists until they give us personal information
- An API KEY from ElevenLabs has yet to be provided, and the ASUS PC encountered issues due to a faulty Windows installation and had to be sent back.

Conclusions

- The project bridges AI and the arts, fostering creativity in NLP.
- Feeding numbers into fine-tuning and expecting to get proper numbers in the output is naive, RAG is correct approach
- Second Feedback loop from artists
- Future work:
 - Analyze exhibition feedback and improve model performance.
 - Assess creativity via human feedback
- Potential Future Research: measuring adherence to artistic style through metric scorers e.g. BLEU, ROUGE

Thank you for your attention! Questions?

- Exhibition will take place mid February 2025. **Feel invited to visit Exhibition in Pałac Czapskich, Krakowskie Przedmieście 5 :)**

<https://pc.asp.waw.pl/>

- Franceschelli, G., & Musolesi, M. (2024). Creativity in Large Language Models
- Zhong, T. et al. (2024). Opportunities and Challenges of Large Language Models for Low-Resource Languages in Humanities Research
- Grattafiori, A., et al. (2024). The LLama 3 Herd of Models
- Hu, E. J., et al. (2021). Low-Rank Adaptation for LLMs
- More references in the project report