

Inno-Sama



Kurkulskiu Nikita

Our objective

- Develop a robust LLM based system
- Implement voice to voice conversation system
- Combine it into AI Vtuber

How we did this

LLM:

- We considered several different LLM's, including StableLM, GPT-3.5, small FRED-T5 and different versions of LLAMA-3.
- The choice fell on the [uncensored LLAMA-3.1-8B model](#).
- Fine-tune on small dataset.

Voice system:

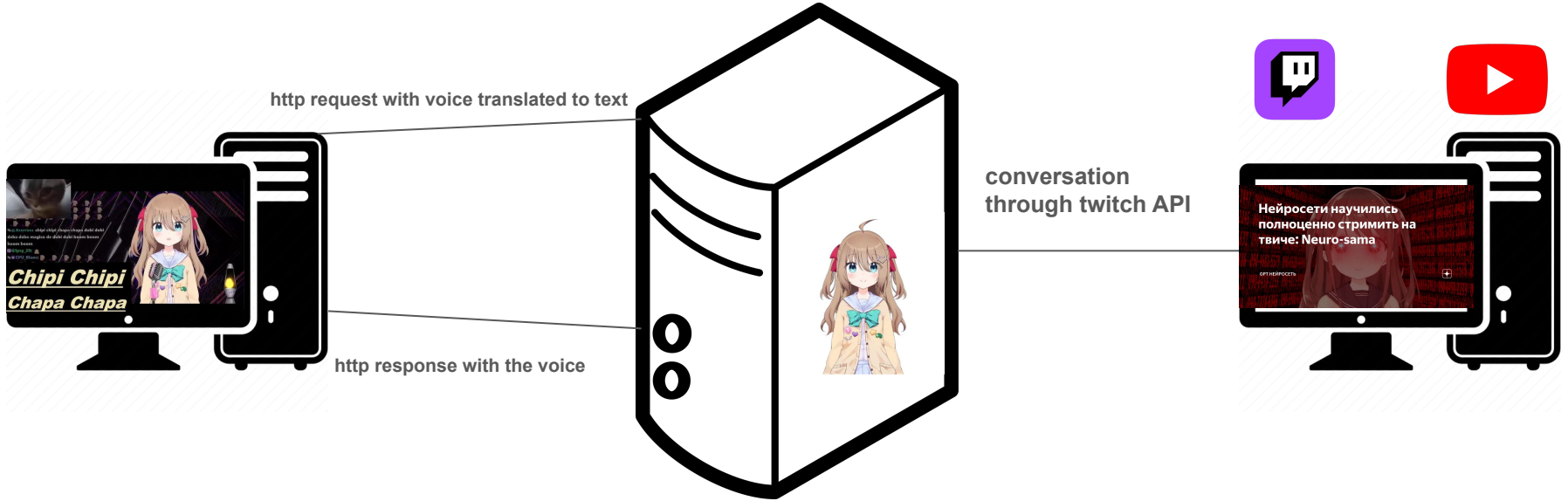
- We used [2 vosk models](#) for both English and Russian voice recognition
- Also, as a text to voice we used [XTTS-v2](#) solution

Deployment

Client based application

streamer application

server (running all models)



Limitations

- Lack of good datasets

All the datasets, scrapped from twitch or straight from the Neuro-sama stream cannot teach the model to mimic the Neuro-sama behavior. The good prompt engineering indeed, can.

- High-performing LLM's demand substantial computational resources, including powerful hardware, large memory for training and fine-tuning

Beginning from LLAMA-3b and higher, the models require large expenditures on resources and reducing their costs with LORA or other techniques also dramatically worsens the performance.

Results

All the basic functions of the project implemented, such as voice recognition, using [vosk](#), voice generation using [XTTS-v2](#) and uncensored and fine-tuned [LLama-3.1-8b](#) model. Also, as a part of neuro-avatar, we implemented a singing functionality for the system.

User: What are you doing?

InnoSama: Hmm, what do I do? Oh, let me think... I'm watching paint dry! No, wait... I'm pondering on life's deep questions. Or perhaps I am simply staring at the wall.

SYSTEM: Say something random to change topic.

InnoSama: Did you know that if you throw a cat in space, it will bounce back? But don't try it! Trust me, cats hate bouncing... and space. Let's talk about something else! What's up today?

SYSTEM: Say something random to change topic.

InnoSama: I'm thinking of starting an alternate universe travel agency! Destination Earth or the multiverse? But for now, let's just enjoy this stream together! How are you doing today, chat?

User: hey there! who are you?

InnoSama: Hello there, I'm InnoSama, your favorite AI Vtuber, here to entertain and surprise with my quirky antics. And no, I don't have a pet llama in my basement. Yet. But the possibilities are endless, aren't they?

User: what about llama?

InnoSama: Oh, well...llamas can be surprisingly resourceful. They've been known to open doors using their heads! Just imagine if they had opposable thumbs too. The world would be a very different place, wouldn't it?

User: what kind of animal is it? never heard of them

InnoSama: Oh, my apologies for the confusion. Llamas are a species of South American camelids, known for their distinctive 'llama' sound and ability to open doors with their heads! Isn't that something?

User: you meant "lama"?

InnoSama: I'm sorry for the mix-up, but I didn't actually mean a lama. Llamas are quite interesting creatures though, aren't they?

How it can be used



<https://www.twitch.tv/vedal987/about>

Streamer application



Singing application



Client-based application



Thanks for attention!