

Experiment with Dutch local LLMs

[MacWhisper](#) (Mac only) is a great local transcription tool that converts audio to text. It also has an option to add LLMs through [Ollama](#) so that you can summarize your transcripts, which would make a great complete suite for recording and summarizing meetings running only local models. For English this seems to work rather well, but local LLMs are known for not being all to reliable for Dutch. So we ran an experiment, summarizing transcripts of the [HKU en AI podcast](#).

“ **Conclusion:** No success yet for summarizing Dutch texts, in Dutch. (May 2025)

Setup

Setting up LLMs in MacWhisper is quite straightforward: install Ollama and install your model. Next, open MacWhisper and go to Global (the settings menu), then AI, Services. If Ollama is running, you should be able to select all installed models in MacWhisper by clicking Ollama under Add another service. Now once you've made a transcript you can interact with ollama under the AI tab (three stars) at the top right.

Testing Dutch in MacWhisper (May '25)

Interacting with any model through MacWhisper in Dutch gives strange results. Replies are often in English, or seem to ignore the prompt completely.

- Gemma3 and Mistral give pretty accurate summaries, but in English only. Interestingly it does seem to understand the Dutch contents of the transcript (although it misses some key points as well).
- Deepseek goes completely off the rails
- Llama3.2 gives a very short summary that misses key points.

Changing prompts or moving to chat mode does not seem to improve anything.

Testing Dutch in LLM directly (May '25)

I thought MacWhisper might be interfering in some way (as I could not get the LLM to react to anything else than 'summarize'), so I moved to [Open WebUI](#). In this way I could still interact with

the LLM and add a text file as input. The textfile was the transcript export from MacWhisper.

While did this improve the interaction as I could talk to the LLM directly, results were similar to above. Some additional models tested here:

- Granite3.2 gave mixed bulletpoints, some accurate, some wildly off. Granite did respond in Dutch.
- Phi3,5 and Phi4 had nonsense results, although in Dutch.
- Two Dutch LLM models Geitje-7b en Fietje-2b completely derailed. They did not answer any questions but went rambling about daycare for young children, paper crafting and Dutch politics. It's clear what these models were trained on...

Revision #11

Created 2025-07-01 09:49:18 UTC by mikal

Updated 2026-03-25 14:07:55 UTC by mikal