

Ready-made prompts — workshop stations

The full set of prompts for the “Bielik cleans your data” hands-on. Each file is one **station**: ready text to copy + a prompt to paste. No programming — it’s a normal conversation with the model.

How it works (the interaction model)

1. **Open the source file** from the corpus (`przyklad-korpus/poznan-1956/...`) in plain Notepad (Windows) or TextEdit (Mac). Select all (`Ctrl+A` / `A`), copy (`Ctrl+C` / `C`).
2. **In LM Studio** (with Bielik loaded) paste **the prompt first** into the chat, and **under it** the text from the file. Send it all as one message.
3. **Read the result.** Don’t like it? Add “redo it so that...” — the model remembers its previous answer in this chat.

Everything runs **locally, on your laptop**. Nothing leaves for the internet.

Stations

#	File	What it teaches
Core (everyone does these)		
1	01-ocr-en.md	fixing a broken scan / OCR
2	02-transkrypcja-streszczenie-en.md	transcript → clean text + summary
3	03-anonimizacja-en.md	removing personal data (GDPR)
Bonus (for the fast ones)		
4	04-csv-tabela-en.md	tidying an inconsistent table/CSV
5	05-tlumaczenie-en.md	PL EN translation with consistent terminology
6	06-porownanie-zrodel-en.md	comparing conflicting sources (57 vs 58 victims)

The corpus is deliberately “dirty”

The files in `przyklad-korpus/poznan-1956/` are **didactic reconstructions** (apart from the photos and audio, which are real) — deliberately full of OCR errors, missing punctuation, and duplicates. It’s raw material for exercises, not a historical source. Details and verified data → `przyklad-korpus/poznan-1956/README.md` and `ZRODLA.md`.

A note on images and audio

- **Bielik reads text, not images.** The scan `wycinki-prasowe/wycinek-skan.png` needs a model with vision (a VLM) — I show that one myself during the workshop. You work on the ready `wycinek-01-ocr.txt`.
- **Audio (nagrania/)** we turn into text with Whisper — I also show that on screen. Then the analysis goes like Station 2.

For the advanced: `oddzialy-wojskowe/sily-wojskowe.md` (the same unit written 3 ways) — ask Bielik to unify the names and pull out a list: unit, commander, equipment.