


AI in research — cheat sheet (Bielik, local)

WPiA UAM workshop · 28.06.2026 · Bielik 4.5B in LM Studio · **everything stays on your laptop, nothing goes to the cloud**

GETTING STARTED (ONCE)

1. You have the model from home: **Bielik-4.5B-v3.0-Instruct** (Discover → `bielik`). Weak laptop ≤8 GB RAM → **1.5B**.
2.  **Chat** → "Select a model to load" → **Bielik 4.5B** → wait for the green "Loaded ✓".
3. How you work: **paste the prompt** → **paste the text under it** → **send as one message**.

3 CORE PROMPTS — COPY AND PASTE

1 • Fix a broken scan (OCR) — file `wycinki-prasowe/wycinek-01-ocr.txt`

Fix the OCR errors in the Polish text below. Restore correct Polish spelling with its diacritics (ą, ę, ł...) and fix the swapped characters (l→ł, 0→o). Keep the meaning. Add nothing of your own.

2 • Transcript → **clean text + summary** — file `relacje-swiadkow/relacja-01-surowa.txt`

This is a raw transcript of a witness recording in Polish. Add punctuation, split it into turns (Witness / Interviewer), and remove speech stumbles ("yyy", "no więc"). At the end, add a 5-sentence summary.

3 • Anonymization (GDPR) — the same account, in the same chat

Anonymize the text above per GDPR. Replace first names, surnames, addresses, and other personal data with placeholders: [WITNESS_1], [STREET], [CITY]. Keep the meaning. At the end, list the changes you made.

Rule: a specific, short task + the text underneath. For "cleaning" tasks set Temperature ≈ 0.3 (right panel). **Always check the result** — the model tidies the form, it isn't a source of truth.

Paid tool → free equivalent

Paid (≈ cost/month)	Free equivalent (local)	Main trade-off
ChatGPT / Claude chat (20 USD)	Bielik 1.5/4.5/11B in LM Studio	weaker on EN, slower
NotebookLM — questions over PDFs (0–20 USD)	LM Studio + Bielik	no podcast generator
DeepL Pro — PL/EN (9 EUR)	Bielik 11B	on Polish, Bielik wins on terminology
Grammarly (12–30 USD)	LanguageTool + Bielik	weaker "tone detector" on EN
Otter.ai — transcription (17 USD)	Whisper.cpp + Buzz GUI	slower, weaker diarization
Adobe Acrobat AI — OCR (15–30 USD)	Marker / olmOCR / Tesseract	handwriting still hard
Notion AI — notes (10–20 USD)	LM Studio (chat-per-project) / Obsidian	no sync out of the box
Gamma — slides (8–20 USD)	Marp + Bielik	needs comfort with Markdown

When is the cloud OK? **Public** data (brainstorming, a public article, an already-published abstract) → ChatGPT/Claude, no problem. **Confidential** data (respondent interviews, an unpublished chapter, case files) → **local only**.

MATERIALS AND CONTACT

- Starter pack (install, prompts, corpus): <https://workshops.mmx3.pl/konferencja-krd-2026/>
- Plan B — Bielik in the browser: <https://bielikchat.pl/>
- Post-workshop survey Sli.do (2 min, thanks!): <https://app.sli.do/event/tf3xpZ9VrJtyam8zmGGBML>