

Station 3 — anonymization (GDPR)

Source file: the same account as in Station 2 (przyklad-korpus/poznan-1956/relacje-swiadkow/relacja-01-sur

Best done **in the same chat** right after Station 2 — Bielik already has the text in memory, so you can write the prompt without pasting it again.

Prompt

Anonymize the text above per GDPR. Replace all personal data with neutral placeholders: first names and surnames → [WITNESS_1], [PERSON_2]; addresses → [STREET], [CITY]; workplace names → [WORKPLACE]. Keep the text's meaning and readability. At the end, give a table of changes: what was replaced and with what.

(If you're starting fresh: paste the prompt first, then the text from the file underneath.)

What to expect: personal data (Tadeusz Kowalczyk, ul. Dąbrowskiego, his wife Halina, ZISPO) replaced with placeholders + a list of changes to verify.

Harder variant — detection only (NER)

Don't change the text. Just list all the personal data that appears in it, grouped into categories: names/surnames, addresses, workplaces, dates, other.

Why it matters in research: a condition of ethics-board approval and of publishing data as open data. Always check the result — the model can be overeager (anonymizing too much) or miss something.