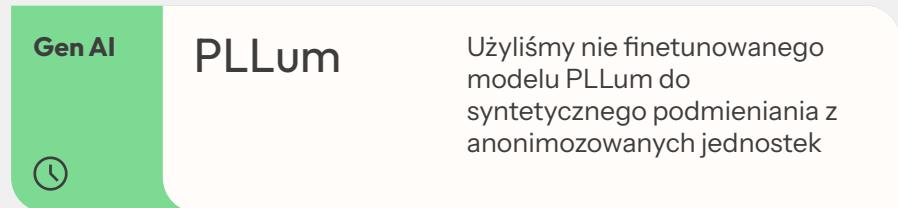
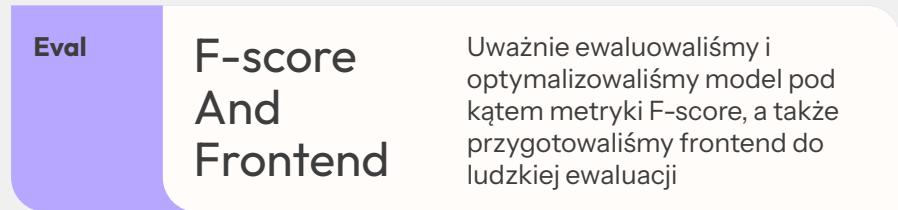
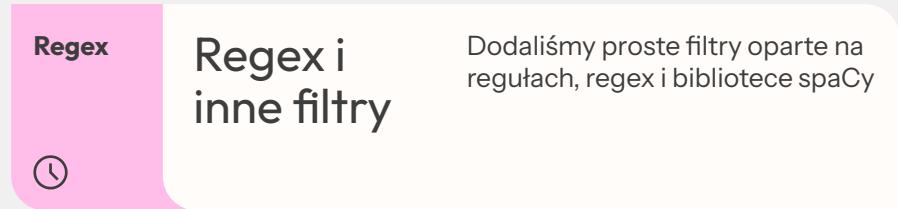
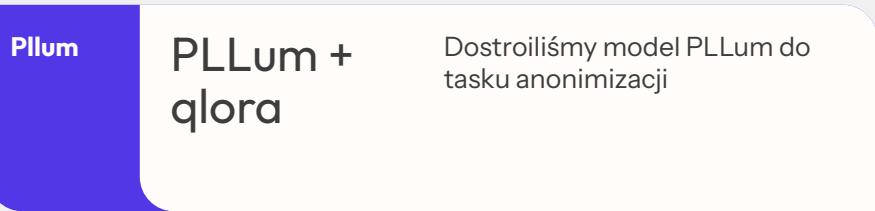


# Dane bez twarzyc

Habanero team



# Pipeline anonimizacji



# Anonimizacja

## Llama-PLLM-8B-instruct + qlora

- Trenowaliśmy na całym dostępnym zbiorze od organizatorów
- Mały model z szybką inferencją dzięki bibliotece VLLM

## Regex, spacy

- Python + regex: skrypt w Pythonie,
- spaCy NER + heurystyki: wykorzystuje model spaCy (pl\_core\_news\_md lub inny) do rozpoznawania encji (osoby, organizacje, lokalizacje, daty) oraz reguły słów kluczowych

**Ewaluacja:** metryka f-score (osięgneliśmy wynik 0.55 na zbiorze testowym!!!)



# Compare results

### Settings

Original file  
example\_data/test.short.in.t

Predicted file  
example\_data/test.short.pre

Reference file  
example\_data/test.short.ref

Show only lines with errors  
 Show per-line statistics  
 Show original text

---

### Legend

- Green: Correct (TP)
- Yellow: Wrong label
- Pink: False positive
- Orange: False negative (missing)

### Original (test.input.txt)

```
Przepompownia ścieków Przyłącze wodociągowe Studnia wodomierzowa Hydrant podziemny Złącze kablowo pomiarowe Szafa sterownicza Stup oświetleniowy H 5m z prefabrykatem betonowym do posadowienia projektowany teren utwardzony pompowni zieleń projektowane ogrodzenie pompowni z bramą projektowana trasa kabli energetycznych i sieci uziemienia w miejscu skrzyżowań i pod terenem utwardzonym kable układane w rurach ochronnych typu Arot 2x110 projektowana trasa kabli sterowniczych kable prowadzone w rurze ochronnej typu Arot 110 JEDNOSTKA PROJEKTUJĄCA FPUH Matulewi(z ulica Wrocławawska l2/14 79-741 Głogów tel +48 782 212 g01 fax +48 609 487 192 KANALIZACJA OBSZARU PARKU KRAJOBRAZOWEGO PUSZCZA ZIELONKA OKOLICE
```

### Predicted (out.txt)

```
Przepompownia ścieków Przyłącze wodociągowe Studnia wodomierzowa Hydrant podziemny Złącze kablowo pomiarowe Szafa sterownicza Stup oświetleniowy H 5m z prefabrykatem betonowym do posadowienia projektowany teren utwardzony pompowni zieleń projektowane ogrodzenie pompowni z bramą projektowana trasa kabli energetycznych i sieci uziemienia w miejscu skrzyżowań i pod terenem utwardzonym kable układane w rurach ochronnych typu Arot 2x110 projektowana trasa kabli sterowniczych kable prowadzone w rurze ochronnej typu Arot 110 JEDNOSTKA PROJEKTUJĄCA [company]/14 79-741 [city] tel +48 782 212 g01 fax [phone] KANALIZACJA OBSZARU PARKU KRAJOBRAZOWEGO PUSZCZA ZIELONKA OKOLICE
```

### Reference (test.ref.txt)

```
Przepompownia ścieków Przyłącze wodociągowe Studnia wodomierzowa Hydrant podziemny Złącze kablowo pomiarowe Szafa sterownicza Stup oświetleniowy H 5m z prefabrykatem betonowym do posadowienia projektowany teren utwardzony pompowni zieleń projektowane ogrodzenie pompowni z bramą projektowana trasa kabli energetycznych i sieci uziemienia w miejscu skrzyżowań i pod terenem utwardzonym kable układane w rurach ochronnych typu Arot 2x110 projektowana trasa kabli sterowniczych kable prowadzone w rurze ochronnej typu Arot 110 JEDNOSTKA PROJEKTUJĄCA [company] [address] tel [phone] fax [phone] KANALIZACJA OBSZARU PARKU KRAJOBRAZOWEGO PUSZCZA ZIELONKA OKOLICE
```

### Differences

False Positives:

- [city] after ...any]/14 79-741

False Negatives (Missing):

- [address] after ...JĄCA [company]

# Generate synthetic data

Llama-PLLM-8B-instruct (bez finetuningu)

## Evaluacja

- metryka **BLEU**  
(utrzymywaliśmy wynik 0.60)
- ilość znalezionych nie podmienionych encji (70% dobrze wstrzykniętych danych!)

```
system_message = (
    "Jesteś asystentem generującym dane syntetyczne po polsku. "
    "Dostajesz szablon z nawiasami klamrowymi (np. [name], [surname], [pesel], [address], itp.). Uzupełnij go fikcyjnymi, ale realistycznymi danymi osobowymi. "
    "Zachowaj strukturę zdania, zastąp tylko nawiasy klamrowe konkretnymi wartościami. "
    "Nie używaj prawdziwych danych żadnej istniejącej osoby – traktuj wszystko jako zmyślone."
)

prompt = f"""<|system|>
{system_message}
<|user|>
Szablon (wejście):
{template_text}

Uzupełnij szablon i zwróć TYLKO gotowe zdanie z podstawionymi danymi (bez nawiasów klamrowych).
<|assistant|>
Wynik:"""

print(prompt)
```