Ohad Cohen

March 2025

1 Regex Solution

A rule-based approach involves using predefined regular expressions (regex) and keyword matching to detect PII patterns like phone numbers, emails, credit card numbers, and personal names.

Pros:

- Simple to implement and interpret.
- Fast and efficient for well-defined PII patterns (e.g., credit card numbers, phone numbers).
- No need for labeled data or model training.

Cons:

- Cannot detect context-dependent PII (e.g., "My salary is 80K").
- Requires manual creation of patterns.
- Needs frequent updates for new PII types.
- High false negatives for variations not covered by regex.

Files: main_sol1.py

2 Fine-tuned Transformer Model (DistilBERT) Solution

A deep learning-based approach where a transformer model is fine-tuned on the provided dataset to classify whether a text contains PII or not.

Pros:

- High accuracy for diverse and complex PII patterns.
- Can detect context-based PII (e.g., "My income is 100K").
- SOTA for NLP.

- Adaptable to different definitions of PII.

Cons:

- Requires training and labeled data.
- Higher inference time compared to regex-based approaches.
- Needs GPU for efficient training and inference.

Files: main_sol2.py, data_pii.py, distilbert.py

I chose a deep learning approach and used DistilBERT, a smaller and faster version of BERT that still performs well. This keeps a good balance between accuracy and speed for real-time PII detection.