

Masking_using_LLM

April 21, 2025

```
[2]: !pip install pandoc
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting pandoc
  Downloading pandoc-2.4.tar.gz (34 kB)
  Preparing metadata (setup.py) ... done
Collecting plumbum
  Downloading plumbum-1.9.0-py3-none-any.whl (127 kB)

128.0/128.0 KB 496.2 kB/s eta 0:00:00 kB/s eta
0:00:01
Collecting ply
  Downloading ply-3.11-py2.py3-none-any.whl (49 kB)

49.6/49.6 KB 2.9 MB/s eta 0:00:00
Building wheels for collected packages: pandoc
  Building wheel for pandoc (setup.py) ... done
  Created wheel for pandoc: filename=pandoc-2.4-py3-none-any.whl
size=34818
sha256=c6e36fa582f87c4765e938027e71511e28e025b0aec088d503735390c6c2ab10
  Stored in directory: /home/cheeta/.cache/pip/wheels/14/79/8c/5d7a023cc8df1aa03
81c1739d69da18ae7f90c08b2dc9a1bf5
Successfully built pandoc
Installing collected packages: ply, plumbum, pandoc
Successfully installed pandoc-2.4 plumbum-1.9.0 ply-3.11
```

0.1 Zero-Shot PII Masking Using a Large Language Model (LLM)

0.1.1 Install Libraries

This cell installs the necessary Python libraries: - `google-generativeai`: For using the Gemini API. - `scikit-learn`: For calculating evaluation metrics. - `requests`: For making HTTP requests. - `openai`: For OpenAI API (if needed). - `gdown`: For downloading files from Google Drive. - `datasets`: For loading and processing datasets.

```
[1]: !pip install google-generativeai scikit-learn requests openai gdown datasets
```

```
Requirement already satisfied: google-generativeai in
/usr/local/lib/python3.11/dist-packages (0.8.4)
```

Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (1.6.1)

Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (2.32.3)

Requirement already satisfied: openai in /usr/local/lib/python3.11/dist-packages (1.61.1)

Requirement already satisfied: gdown in /usr/local/lib/python3.11/dist-packages (5.2.0)

Collecting datasets

 Downloading datasets-3.4.1-py3-none-any.whl.metadata (19 kB)

Requirement already satisfied: google-ai-generativelanguage==0.6.15 in /usr/local/lib/python3.11/dist-packages (from google-generativeai) (0.6.15)

Requirement already satisfied: google-api-core in /usr/local/lib/python3.11/dist-packages (from google-generativeai) (2.24.2)

Requirement already satisfied: google-api-python-client in /usr/local/lib/python3.11/dist-packages (from google-generativeai) (2.160.0)

Requirement already satisfied: google-auth>=2.15.0 in /usr/local/lib/python3.11/dist-packages (from google-generativeai) (2.38.0)

Requirement already satisfied: protobuf in /usr/local/lib/python3.11/dist-packages (from google-generativeai) (4.25.6)

Requirement already satisfied: pydantic in /usr/local/lib/python3.11/dist-packages (from google-generativeai) (2.10.6)

Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from google-generativeai) (4.67.1)

Requirement already satisfied: typing-extensions in /usr/local/lib/python3.11/dist-packages (from google-generativeai) (4.12.2)

Requirement already satisfied: proto-plus<2.0.0dev,>=1.22.3 in /usr/local/lib/python3.11/dist-packages (from google-ai-generativelanguage==0.6.15->google-generativeai) (1.26.1)

Requirement already satisfied: numpy>=1.19.5 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (2.0.2)

Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.14.1)

Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.4.2)

Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (3.6.0)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests) (3.4.1)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests) (3.10)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests) (2.3.0)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests) (2025.1.31)

Requirement already satisfied: anyio<5,>=3.5.0 in /usr/local/lib/python3.11/dist-packages (from openai) (3.7.1)

Requirement already satisfied: distro<2,>=1.7.0 in

```

/usr/local/lib/python3.11/dist-packages (from openai) (1.9.0)
Requirement already satisfied: httpx<1,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from openai) (0.28.1)
Requirement already satisfied: jiter<1,>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from openai) (0.9.0)
Requirement already satisfied: sniffio in /usr/local/lib/python3.11/dist-
packages (from openai) (1.3.1)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.11/dist-
packages (from gdown) (4.13.3)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-
packages (from gdown) (3.17.0)
Requirement already satisfied: pyarrow>=15.0.0 in
/usr/local/lib/python3.11/dist-packages (from datasets) (18.1.0)
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages
(from datasets) (2.2.2)
Collecting xxhash (from datasets)
  Downloading
xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(12 kB)
Collecting multiprocessing<0.70.17 (from datasets)
  Downloading multiprocessing-0.70.16-py311-none-any.whl.metadata (7.2 kB)
Requirement already satisfied: fsspec<=2024.12.0,>=2023.1.0 in
/usr/local/lib/python3.11/dist-packages (from
fsspec[http]<=2024.12.0,>=2023.1.0->datasets) (2024.10.0)
Requirement already satisfied: aiohttp in /usr/local/lib/python3.11/dist-
packages (from datasets) (3.11.13)
Requirement already satisfied: huggingface-hub>=0.24.0 in
/usr/local/lib/python3.11/dist-packages (from datasets) (0.28.1)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-
packages (from datasets) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-
packages (from datasets) (6.0.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-
packages (from aiohttp->datasets) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (6.1.0)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (0.3.0)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.18.3)

```

Requirement already satisfied: googleapis-common-protos<2.0.0,>=1.56.2 in /usr/local/lib/python3.11/dist-packages (from google-api-core->google-generativeai) (1.69.1)

Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from google-auth>=2.15.0->google-generativeai) (5.5.2)

Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.11/dist-packages (from google-auth>=2.15.0->google-generativeai) (0.4.1)

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.11/dist-packages (from google-auth>=2.15.0->google-generativeai) (4.9)

Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.23.0->openai) (1.0.7)

Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.*->httpx<1,>=0.23.0->openai) (0.14.0)

Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic->google-generativeai) (0.7.0)

Requirement already satisfied: pydantic-core==2.27.2 in /usr/local/lib/python3.11/dist-packages (from pydantic->google-generativeai) (2.27.2)

Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.11/dist-packages (from beautifulsoup4->gdown) (2.6)

Requirement already satisfied: httplib2<1.dev0,>=0.19.0 in /usr/local/lib/python3.11/dist-packages (from google-api-python-client->google-generativeai) (0.22.0)

Requirement already satisfied: google-auth-httplib2<1.0.0,>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from google-api-python-client->google-generativeai) (0.2.0)

Requirement already satisfied: writemplate<5,>=3.0.1 in /usr/local/lib/python3.11/dist-packages (from google-api-python-client->google-generativeai) (4.1.1)

Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2.8.2)

Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2025.1)

Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2025.1)

Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /usr/local/lib/python3.11/dist-packages (from requests[socks]->gdown) (1.7.1)

Requirement already satisfied: grpcio<2.0dev,>=1.33.2 in /usr/local/lib/python3.11/dist-packages (from google-api-core[grpc]!=2.0.*,!=2.1.*,!=2.10.*,!=2.2.*,!=2.3.*,!=2.4.*,!=2.5.*,!=2.6.*,!=2.7.*,!=2.8.*,!=2.9.*,<3.0.0dev,>=1.34.1->google-ai-generativelanguage==0.6.15->google-generativeai) (1.71.0)

Requirement already satisfied: grpcio-status<2.0.dev0,>=1.33.2 in /usr/local/lib/python3.11/dist-packages (from google-api-core[grpc]!=2.0.*,!=2.1

```

.*,!=2.10.*,!=2.2.*,!=2.3.*,!=2.4.*,!=2.5.*,!=2.6.*,!=2.7.*,!=2.8.*,!=2.9.*,<3.0
.0dev,>=1.34.1->google-ai-generativelanguage==0.6.15->google-generativeai)
(1.62.3)
Requirement already satisfied:
pyparsing!=3.0.0,!=3.0.1,!=3.0.2,!=3.0.3,<4,>=2.4.2 in
/usr/local/lib/python3.11/dist-packages (from httplib2<1.dev0,>=0.19.0->google-
api-python-client->google-generativeai) (3.2.1)
Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in
/usr/local/lib/python3.11/dist-packages (from pyasn1-modules>=0.2.1->google-
auth>=2.15.0->google-generativeai) (0.6.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-
packages (from python-dateutil>=2.8.2->pandas->datasets) (1.17.0)
Downloading datasets-3.4.1-py3-none-any.whl (487 kB)
487.4/487.4 kB
5.9 MB/s eta 0:00:00
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
116.3/116.3 kB
5.3 MB/s eta 0:00:00
Downloading multiprocessing-0.70.16-py311-none-any.whl (143 kB)
143.5/143.5 kB
6.4 MB/s eta 0:00:00
Downloading
xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
194.8/194.8 kB
8.4 MB/s eta 0:00:00
Installing collected packages: xxhash, dill, multiprocessing, datasets
Successfully installed datasets-3.4.1 dill-0.3.8 multiprocessing-0.70.16
xxhash-3.5.0

```

0.1.2 Import Libraries

This cell imports the required libraries and modules: - `google.generativeai`: For interacting with the Gemini API. - `time`: For adding delays between API calls. - `gdown`: For downloading files from Google Drive. - `sklearn.metrics`: For calculating precision, recall, F1-score, and accuracy. - `datasets`: For loading datasets.

```

[2]: import google.generativeai as genai
import time
import gdown
from sklearn.metrics import precision_score, recall_score, f1_score,
    accuracy_score
from datasets import load_dataset

```

0.1.3 Configure Gemini API

This cell sets up the Gemini API using your API key. It initializes the `gemini-1.5-pro` model for generating responses.

```
[16]: genai.configure(api_key="AIzaSyCxrDjE60DeQP100zQrFBmbPGEHOTTxwBc")
# new3 = AIzaSyDwdA9LqXeCMjwGlUlWjgLyfQo6FBXDM1s
# new4 = AIzaSyCxrDjE60DeQP100zQrFBmbPGEHOTTxwBc
model = genai.GenerativeModel("gemini-1.5-pro")
```

0.1.4 Define Functions

This cell defines the following functions: - `create_prompt()`: Creates a prompt for the Gemini model to mask PII. - `mask_pii()`: Sends the prompt to the Gemini API and returns the masked text. - `evaluate()`: Compares the model's output with ground truth to calculate metrics like accuracy, precision, recall, F1-score, FPR, and FNR.

```
[4]: def create_prompt(text):
    return f"""
    Identify and mask all Personally Identifiable Information (PII) in the
    following text. Use the following format:
    - Names: [REDACTED_NAME]
    - Emails: [REDACTED_EMAIL]

    Text: "{text}"
    """
```

```
[5]: def mask_pii(text):
    prompt = create_prompt(text)
    response = model.generate_content(prompt)
    return response.text.strip()
```

```
[6]: def tokenize(text):
    return text.split()
```

```
[7]: def evaluate(ground_truth_tokens, llm_output_tokens):
    ground_truth_binary = [1 if tag != "0" else 0 for tag in
    ground_truth_tokens]
    predicted_binary = [1 if token.startswith("[REDACTED]") else 0 for token in
    llm_output_tokens]

    TP = sum((gt == 1 and pred == 1) for gt, pred in zip(ground_truth_binary,
    predicted_binary))
    FP = sum((gt == 0 and pred == 1) for gt, pred in zip(ground_truth_binary,
    predicted_binary))
    FN = sum((gt == 1 and pred == 0) for gt, pred in zip(ground_truth_binary,
    predicted_binary))
    TN = sum((gt == 0 and pred == 0) for gt, pred in zip(ground_truth_binary,
    predicted_binary))

    accuracy = (TP + TN) / (TP + TN + FP + FN) if (TP + TN + FP + FN) != 0 else
    0
```



```
'percent', 'in', 'favor', 'of', 'the', 'marijuana', 'legalization', 'measure',
',', 'which', ',', 'as', 'then-mayor', 'John', 'Hickenlooper', 'pointed', 'out',
',', 'was', 'without', 'effect', ',', 'because', 'the', 'city', 'cannot',
'usurp', 'state', 'law', ',', 'which', 'at', 'that', 'time', 'treated',
'marijuana', 'possession', 'in', 'much', 'the', 'same', 'way', 'as', 'a',
'speeding', 'ticket', ',', 'with', 'fines', 'of', 'up', 'to', '$', '100', 'and',
'no', 'jail', 'time', '.']]]}
```

```
[10]: print("size of dataset: ", len(test_data))
```

```
size of dataset: 3650
```

0.1.6 Import Additional Libraries

This cell imports more libraries for advanced functionality: - **random**: For generating random numbers (used in exponential backoff). - **numpy**: For numerical calculations (e.g., averaging metrics). - **pandas**: For data manipulation. - **ast**: For safely evaluating strings as Python objects. - **google.api_core.exceptions**: For handling API errors. - **collections.defaultdict**: For caching API responses.

0.1.7 Set Configuration

This cell defines configuration parameters: - **batch_size**: Number of samples processed in each batch. - **max_retries**: Maximum number of retries for API calls. - **base_wait_time**: Initial delay between retries. - **delay_seconds**: Delay between batches to avoid rate limits. - **size_of_dataset**: Number of samples to process from the dataset.

0.1.8 Initialize Metric Storage

This cell initializes lists to store evaluation metrics: - Accuracy, Precision, Recall, F1-Score, FPR, and FNR.

0.1.9 Main Processing Loop

This cell processes the dataset in batches: 1. Fetches a batch of data. 2. Checks the cache to avoid redundant API calls. 3. Sends the text to the Gemini API for masking. 4. Evaluates the model's performance using the `evaluate()` function. 5. Stores the metrics for each batch.

0.1.10 Aggregate Metrics

This cell calculates the overall performance metrics by averaging the results from all batches.

0.1.11 Doing evaluation on test dataset (without synthetic emails)

```
[17]: import time
import random
import numpy as np
from google.api_core.exceptions import TooManyRequests
from collections import defaultdict
```



```

batch_size = 1
max_retries = 2
base_wait_time = 5
delay_seconds = 30

all_accuracies = []
all_precisions = []
all_recalls = []
all_f1_scores = []
all_fprs = []
all_fnrs = []

size_of_dataset = 30 #change according to your need in range of [0, 3650]
test_data = dataset["test_data"].select(range(size_of_dataset))

cache = defaultdict(str)

def exponential_backoff(attempt):
    jitter = random.uniform(1, 3)
    return base_wait_time * (2 ** attempt) + jitter

for i in range(0, len(test_data), batch_size):
    batch = [test_data[j] for j in range(i, min(i + batch_size,
↪len(test_data)))]

    batch_texts = [item["sequence"] for item in batch]
    cached_results = {text: cache[text] for text in batch_texts if text in
↪cache}
    texts_to_query = [text for text in batch_texts if text not in
↪cached_results]

    attempt = 0
    while attempt < max_retries:
        try:
            if texts_to_query:
                masked_texts = [mask_pii(text) for text in texts_to_query]
                for text, masked in zip(texts_to_query, masked_texts):
                    cache[text] = masked
                break
            except TooManyRequests as e:
                retry_after = int(e.response.headers.get("Retry-After",
↪base_wait_time))
                wait_time = max(retry_after, exponential_backoff(attempt))
                print(f"Rate limit exceeded. Retrying in {wait_time:.2f} seconds...
↪")
                time.sleep(wait_time)
                attempt += 1

```

```

if attempt == max_retries:
    print("Max retries reached. Skipping batch.")
    continue

final_masked_texts = [cache[text] for text in batch_texts]

for text, masked_text, ner_tags in zip(batch_texts, final_masked_texts,
↳[item["ner_tags"] for item in batch]):
    print(f"Original Text: {text}")
    print(f"Masked Text: {masked_text}")

    llm_output_tokens = tokenize(masked_text)

    accuracy, precision, recall, f1, FPR, FNR = evaluate(ner_tags,
↳llm_output_tokens)

    all_accuracies.append(accuracy)
    all_precisions.append(precision)
    all_recalls.append(recall)
    all_f1_scores.append(f1)
    all_fprs.append(FPR)
    all_fnrs.append(FNR)

    print(f"Accuracy: {accuracy:.2f}")
    print(f"Precision: {precision:.2f}")
    print(f"Recall: {recall:.2f}")
    print(f"F1-Score: {f1:.2f}")
    print(f"False Positive Rate (FPR): {FPR:.2f}")
    print(f"False Negative Rate (FNR): {FNR:.2f}")
    print("-" * 50)

time.sleep(delay_seconds)

overall_accuracy = np.mean(all_accuracies) if all_accuracies else 0
overall_precision = np.mean(all_precisions) if all_precisions else 0
overall_recall = np.mean(all_recalls) if all_recalls else 0
overall_f1 = np.mean(all_f1_scores) if all_f1_scores else 0
overall_FPR = np.mean(all_fprs) if all_fprs else 0
overall_FNR = np.mean(all_fnrs) if all_fnrs else 0

print("\n=== Overall Dataset Results ===")
print(f"Overall Accuracy: {overall_accuracy:.2f}")
print(f"Overall Precision: {overall_precision:.2f}")
print(f"Overall Recall: {overall_recall:.2f}")
print(f"Overall F1-Score: {overall_f1:.2f}")

```

```
print(f"Overall False Positive Rate (FPR): {overall_FPR:.2f}")
print(f"Overall False Negative Rate (FNR): {overall_FNR:.2f}")
print("=====")
```

Original Text: included future Rage Against the Machine and Audioslave drummer Brad Wilk .

Masked Text: "included future Rage Against the Machine and Audioslave drummer [REDACTED_NAME] ."

Accuracy: 0.91

Precision: 1.00

Recall: 0.50

F1-Score: 0.67

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.50

Original Text: The city voted 53.5 percent in favor of the marijuana legalization measure , which , as then-mayor John Hickenlooper pointed out , was without effect , because the city cannot usurp state law , which at that time treated marijuana possession in much the same way as a speeding ticket , with fines of up to \$ 100 and no jail time .

Masked Text: The city voted 53.5 percent in favor of the marijuana legalization measure, which, as then-mayor [REDACTED_NAME] pointed out, was without effect, because the city cannot usurp state law, which at that time treated marijuana possession in much the same way as a speeding ticket, with fines of up to \$100 and no jail time.

Accuracy: 0.94

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.02

False Negative Rate (FNR): 1.00

Original Text: It was not until about 1907 - 1909 that he produced his first paintings , which were portraits and nudes in a style influenced by Fauvism and Cézanne .

Masked Text: No PII found in the provided text.

Accuracy: 1.00

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.00

Original Text: Always in his bowler hat , he was a witty presence , along with his good friend Constant Detré , at Le Dôme Café , Jockey-Club de Paris , and the other haunts of the area 's bohemian society .

Masked Text: "Always in his bowler hat, he was a witty presence, along with his

good friend [REDACTED_NAME] at Le Dôme Café, Jockey-Club de Paris, and the other haunts of the area's bohemian society."

Accuracy: 0.91

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.03

False Negative Rate (FNR): 1.00

Original Text: The music was composed and conducted by Ronald Stein .

Masked Text: The music was composed and conducted by [REDACTED_NAME].

Accuracy: 1.00

Precision: 1.00

Recall: 1.00

F1-Score: 1.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.00

Original Text: Lord Londonderry used his new bride 's immense wealth to acquire the Seaham Hall estate in County Durham to developing the coalfields there .

Masked Text: No PII found in the provided text.

Accuracy: 0.86

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: It was completed by Philip Wyatt in 1841 and cost £ 130,000 (equivalent to £ 10,772,000 in 2016) to build and furnish .

Masked Text: No PII is present in the provided text. The text refers to a historical figure and financial information related to a building project, not information about a living individual.

Accuracy: 0.92

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: He was in disagreement over this question with his son and heir Frederick , who was more liberally inclined .

Masked Text: No PII is present in the provided text. Therefore, no redaction is needed.

Accuracy: 0.92

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 1.00

Original Text: Ajax signed Mounir El Hamdaoui as his successor .
Masked Text: "Ajax signed [REDACTED_NAME] as his successor."

Accuracy: 0.67
Precision: 1.00
Recall: 0.33
F1-Score: 0.50
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.67

Original Text: Her semi-autobiographical novel " One True Thing " (1994)
served as the basis for the 1998 film starring Meryl Streep and Renée Zellweger .

Masked Text: There is no PII in the provided text. Meryl Streep and Renée Zellweger are public figures, and their names in this context do not constitute private information. The text refers to their professional roles, not personal details.

Accuracy: 0.85
Precision: 0.00
Recall: 0.00
F1-Score: 0.00
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 1.00

Original Text: It was the first all-black show to reach the Broadway stage in over a decade and included a young Josephine Baker among the performers .

Masked Text: No PII is present in the provided text. Josephine Baker is a historical figure and thus not considered PII in this context.

Accuracy: 0.91
Precision: 0.00
Recall: 0.00
F1-Score: 0.00
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 1.00

Original Text: Fellow Belgian Johan Museeuw had escaped to a solo victory .

Masked Text: "Fellow Belgian [REDACTED_NAME] had escaped to a solo victory ."

Accuracy: 0.90
Precision: 1.00
Recall: 0.50
F1-Score: 0.67
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.50

Original Text: However , he did not win a stage in the first week , beaten by Robbie McEwen and Óscar Freire .

Masked Text: There is no PII in the provided text. "Robbie McEwen" and "Óscar Freire" are likely public figures (professional cyclists), so they wouldn't be considered PII in this context.

Accuracy: 0.81

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: In the Tour of Flanders , he took on a defensive role when his teammate Stijn Devolder escaped and won .

Masked Text: No PII is present in the provided text. The text refers to a cyclist named Stijn Devolder in the context of a sporting event. While this is a real person's name, in this context, it's considered public information related to his professional cycling career and not private information requiring redaction. Therefore, no masking is necessary.

Accuracy: 0.90

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: A week later , he outsprinted Fabian Cancellara and Alessandro Ballan in the final 500 m to win the Paris-Roubaix .

Masked Text: No PII is present in the provided text. The text refers to professional cyclists, and while these are names of individuals, they are public figures in the context of a sporting event. Therefore, redaction is not necessary or appropriate.

Accuracy: 0.81

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: Its sporting director , Jean-René Bernaudeau , confirmed the report .

Masked Text: Its sporting director, [REDACTED_NAME], confirmed the report.

Accuracy: 0.57

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.20

False Negative Rate (FNR): 1.00

Original Text: He returned to racing in the Eneco Tour where he won the third

stage by beating Tyler Farrar in the sprint .

Masked Text: No PII is present in the provided text. "Tyler Farrar" is a publicly known professional cyclist, and therefore, not considered private information in this context.

Accuracy: 0.91

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: He finished his season with a second place in Paris-Tours , beaten in a sprint of three by fellow countryman and defending champion Philippe Gilbert .

Masked Text: No PII is present in the provided text. Therefore, no redaction is needed. The text remains as is:

"He finished his season with a second place in Paris-Tours , beaten in a sprint of three by fellow countryman and defending champion Philippe Gilbert ."

Accuracy: 0.92

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: His third victory equalled those of Achiel Buysse , Fiorenzo Magni , Eric Leman and Johan Museeuw .

Masked Text: No PII is present in the provided text. The text mentions the names of cyclists Achiel Buysse, Fiorenzo Magni, Eric Leman, and Johan Museeuw, which are considered public figures and thus their names are not typically considered PII in this context. No email addresses, locations, or other identifying information are present.

Accuracy: 0.56

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: He won the national championship title in June , taking the tricolor jersey from Philippe Gilbert .

Masked Text: No PII is present in the provided text. "Philippe Gilbert" is a known public figure (a professional cyclist), so it's not considered PII in this context. Therefore, no redaction is needed.

Accuracy: 0.88

Precision: 0.00

Recall: 0.00

F1-Score: 0.00
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 1.00

Original Text: He was part of the lead-out train which helped teammate Marcel Kittel take the win .
Masked Text: There is no PII in the provided text. The text refers to a cyclist named Marcel Kittel, but this is considered public knowledge as he is a professional athlete. Therefore, no redaction is necessary.
Accuracy: 0.88
Precision: 0.00
Recall: 0.00
F1-Score: 0.00
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 1.00

Original Text: From 2003 to 2005 she was a minister in the state government of Lower Saxony , serving in the cabinet of Christian Wulff , with responsibility for social affairs , women , family , and health .
Masked Text: "From 2003 to 2005 she was a minister in the state government of Lower Saxony , serving in the cabinet of [REDACTED_NAME], with responsibility for social affairs , women , family , and health ."
Accuracy: 0.97
Precision: 1.00
Recall: 0.50
F1-Score: 0.67
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.50

Original Text: In March 2015 , she and her counterparts from France and Poland , Jean-Yves Le Drian and Tomasz Siemoniak , revived a meeting format intended to promote co-operation between the three countries in crisis zones by holding their first meeting between the Weimar Triangle defence ministers since 2007 .
Masked Text: "In March 2015 , she and her counterparts from France and Poland , [REDACTED_NAME] and [REDACTED_NAME] , revived a meeting format intended to promote co-operation between the three countries in crisis zones by holding their first meeting between the Weimar Triangle defence ministers since 2007 ."
Accuracy: 0.93
Precision: 1.00
Recall: 0.40
F1-Score: 0.57
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.60

Original Text: The bulk of traditional sources date the book to the reign of the Roman emperor Domitian (AD 81 - 96) , which evidence tends to confirm .
Masked Text: No PII found in the provided text.
Accuracy: 1.00

Precision: 0.00
Recall: 0.00
F1-Score: 0.00
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.00

Original Text: Over half of the references stem from Daniel , Ezekiel , Psalms , and Isaiah , with Daniel providing the largest number in proportion to length and Ezekiel standing out as the most influential .

Masked Text: The provided text does not contain any PII. It refers to biblical books and prophets, not individuals' names or contact information. Therefore, no redaction is needed.

Accuracy: 0.96
Precision: 0.00
Recall: 0.00
F1-Score: 0.00
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 1.00

Original Text: Federico Fellini recounts his youth in Rome .

Masked Text: "Federico Fellini recounts his youth in Rome."

There is no PII in the provided text. Federico Fellini is the name of a famous film director and thus not considered personally identifiable information in this context. There are no emails or other PII present.

Accuracy: 0.75
Precision: 0.00
Recall: 0.00
F1-Score: 0.00
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 1.00

Original Text: During editing , a scene with Alberto Sordi was cut because it was considered too immoral and cruel .

Masked Text: There is no PII in the provided text. The text refers to Alberto Sordi, a well-known Italian actor, and thus not considered personally identifiable information in this context.

Accuracy: 0.89
Precision: 0.00
Recall: 0.00
F1-Score: 0.00
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 1.00

Original Text: Various Christian anarchists , such as Jacques Ellul , have identified the State and political power as the Beast and the events described , being their doings and results , the aforementioned ' wrath ' .

Masked Text: No PII is present in the provided text. Therefore, no redaction is

needed. The text discusses Christian anarchist philosophy and does not contain names, emails, or other typically considered PII.

Accuracy: 0.93

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: The so-constructed codebook also forms the so-called Grassmannian manifold .

Masked Text: There is no PII (Personally Identifiable Information) in the provided text. The text discusses a mathematical concept related to Grassmannian manifolds. Therefore, there is nothing to redact.

Accuracy: 0.90

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 1.00

Original Text: He served an apprenticeship with a printer and then worked as a typesetter , contributing articles to the newspaper of his older brother Orion Clemens .

Masked Text: There is no PII in the provided text. Therefore, no redaction is needed.

Accuracy: 1.00

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.00

=== Overall Dataset Results ===

Overall Accuracy: 0.88

Overall Precision: 0.20

Overall Recall: 0.11

Overall F1-Score: 0.14

Overall False Positive Rate (FPR): 0.01

Overall False Negative Rate (FNR): 0.79

=====

0.1.12 Doing evaluation on synthetic dataset (dataset with append email)

```
[11]: # https://drive.google.com/file/d/1-1v9MghJ6XnGDdlKaD4h-se1ZNfk6hYV/view?
      ↳usp=sharing
      synthetic_independent_test_data_file_id = "1-1v9MghJ6XnGDdlKaD4h-se1ZNfk6hYV"

      gdown.download(f"https://drive.google.com/uc?
      ↳id={synthetic_independent_test_data_file_id}", "synthetic_test_data.csv",
      ↳quiet=False)
      DATA_FILES = {"synthetic_test_data": "synthetic_test_data.csv"}
      synthetic_test_dataset = load_dataset("csv", data_files=DATA_FILES)
```

Downloading...

From: https://drive.google.com/uc?id=1-1v9MghJ6XnGDdlKaD4h-se1ZNfk6hYV

To: /content/synthetic_test_data.csv

100%| | 18.2M/18.2M [00:00<00:00, 103MB/s]

Generating synthetic_test_data split: 0 examples [00:00, ? examples/s]

```
[14]: import time
      import random
      import numpy as np
      import pandas as pd
      import ast
      from google.api_core.exceptions import TooManyRequests
      from collections import defaultdict

      batch_size = 1
      max_retries = 2
      base_wait_time = 5
      delay_seconds = 35
      size_of_dataset = 30 # Adjust as needed (max depends on your dataset size)

      test_data = synthetic_test_dataset["synthetic_test_data"].
      ↳select(range(size_of_dataset))

      all accuracies = []
      all precisions = []
      all recalls = []
      all f1_scores = []
      all fprs = []
      all fnrs = []

      cache = defaultdict(str)

      def exponential_backoff(attempt):
          jitter = random.uniform(1, 3)
          return base_wait_time * (2 ** attempt) + jitter
```

```

for i in range(0, len(test_data), batch_size):
    # Use .select() to slice the dataset
    batch_indices = list(range(i, min(i + batch_size, len(test_data))))
    batch = test_data.select(batch_indices)

    batch_texts = [item["sequence"] for item in batch]
    cached_results = {text: cache[text] for text in batch_texts if text in
↪cache}
    texts_to_query = [text for text in batch_texts if text not in
↪cached_results]

    attempt = 0
    while attempt < max_retries:
        try:
            if texts_to_query:
                # Assuming mask_pii() is your PII masking function
                masked_texts = [mask_pii(text) for text in texts_to_query]
                for text, masked in zip(texts_to_query, masked_texts):
                    cache[text] = masked

                break
            except TooManyRequests as e:
                retry_after = int(e.response.headers.get("Retry-After",
↪base_wait_time))
                wait_time = max(retry_after, exponential_backoff(attempt))
                print(f"Rate limit exceeded. Retrying in {wait_time:.2f} seconds...
↪")
                time.sleep(wait_time)
                attempt += 1

        if attempt == max_retries:
            print("Max retries reached. Skipping batch.")
            continue

    final_masked_texts = [cache[text] for text in batch_texts]

    for text, masked_text, ner_tags in zip(batch_texts, final_masked_texts,
↪[item["ner_tags"] for item in batch]):
        print(f"Original Text: {text}")
        print(f"Masked Text: {masked_text}")

        llm_output_tokens = tokenize(masked_text)

        accuracy, precision, recall, f1, FPR, FNR = evaluate(ner_tags,
↪llm_output_tokens)

```

```

        all_accuracies.append(accuracy)
        all_precisions.append(precision)
        all_recalls.append(recall)
        all_f1_scores.append(f1)
        all_fprs.append(FPR)
        all_fnrs.append(FNR)

        print(f"Accuracy: {accuracy:.2f}")
        print(f"Precision: {precision:.2f}")
        print(f"Recall: {recall:.2f}")
        print(f"F1-Score: {f1:.2f}")
        print(f"False Positive Rate (FPR): {FPR:.2f}")
        print(f"False Negative Rate (FNR): {FNR:.2f}")
        print("-" * 50)

    time.sleep(delay_seconds)

overall_accuracy = np.mean(all_accuracies) if all_accuracies else 0
overall_precision = np.mean(all_precisions) if all_precisions else 0
overall_recall = np.mean(all_recalls) if all_recalls else 0
overall_f1 = np.mean(all_f1_scores) if all_f1_scores else 0
overall_FPR = np.mean(all_fprs) if all_fprs else 0
overall_FNR = np.mean(all_fnrs) if all_fnrs else 0

print("\n=== Overall Dataset Results ===")
print(f"Overall Accuracy: {overall_accuracy:.2f}")
print(f"Overall Precision: {overall_precision:.2f}")
print(f"Overall Recall: {overall_recall:.2f}")
print(f"Overall F1-Score: {overall_f1:.2f}")
print(f"Overall False Positive Rate (FPR): {overall_FPR:.2f}")
print(f"Overall False Negative Rate (FNR): {overall_FNR:.2f}")
print("=====")

```

Original Text: Since then , only Terry Bradshaw in 147 games , Joe Montana in 139 games , and Tom Brady in 131 games have reached 100 wins more quickly .
 terry.bradshaw81@fastnu.edu.pk joe.montana96@outlook.com tom.brady11@outlook.com
 Masked Text: Since then, only [REDACTED_NAME] in 147 games, [REDACTED_NAME] in 139 games, and [REDACTED_NAME] in 131 games have reached 100 wins more quickly.
 [REDACTED_EMAIL] [REDACTED_EMAIL] [REDACTED_EMAIL]

Accuracy: 0.24

Precision: 0.67

Recall: 0.19

F1-Score: 0.30

False Positive Rate (FPR): 0.50

False Negative Rate (FNR): 0.81

 Original Text: He was portrayed by Anthony Perkins in the 1960 version of "

Psycho " directed by Alfred Hitchcock and the " Psycho " franchise .
anthony.perkins17@outlook.com alfred.hitchcock29@fastnu.edu.pk
Masked Text: He was portrayed by [REDACTED_NAME] in the 1960 version of " Psycho
" directed by [REDACTED_NAME] and the " Psycho " franchise . [REDACTED_EMAIL]
[REDACTED_EMAIL]
Accuracy: 0.32
Precision: 1.00
Recall: 0.19
F1-Score: 0.32
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.81

Original Text: The egg eventually hatches , revealing a baby Sharptooth .
sharptooth16@yahoo.com
Masked Text: The egg eventually hatches, revealing a baby Sharptooth.
[REDACTED_EMAIL]
Accuracy: 0.33
Precision: 1.00
Recall: 0.14
F1-Score: 0.25
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.86

Original Text: In the video Kelis is walking down a street in a large space
suit-style coat , then singing to the stars . kelis69@outlook.com
Masked Text: In the video [REDACTED_NAME] is walking down a street in a large
space suit-style coat , then singing to the stars . [REDACTED_EMAIL]
Accuracy: 0.22
Precision: 1.00
Recall: 0.10
F1-Score: 0.18
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.90

Original Text: According to food writer Sharon Tyler Herbst , " pico de gallo "
(" rooster 's beak ") is named thus because originally people ate it by
pinching pieces between the thumb and forefinger . sharon.tyler38@outlook.com
Masked Text: According to food writer [REDACTED_NAME] , " pico de gallo " ("
rooster 's beak ") is named thus because originally people ate it by pinching
pieces between the thumb and forefinger . [REDACTED_EMAIL]
Accuracy: 0.17
Precision: 1.00
Recall: 0.06
F1-Score: 0.12
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.94

Original Text: In this way Plato indicated his high opinion of geometry .

plato61@fastnu.edu.pk

Masked Text: In this way Plato indicated his high opinion of geometry.

[REDACTED_EMAIL]

Accuracy: 0.27

Precision: 1.00

Recall: 0.11

F1-Score: 0.20

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.89

Original Text: though the concept was known before that and was studied by many chemists including Avogadro . avogadro59@gmail.com

Masked Text: "though the concept was known before that and was studied by many chemists including [REDACTED_NAME] Avogadro. [REDACTED_EMAIL]"

Accuracy: 0.29

Precision: 1.00

Recall: 0.14

F1-Score: 0.25

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.86

Original Text: They have been part of the Duchy of Normandy since the tenth century , and Queen Elizabeth II is often referred to by her traditional and conventional title of Duke of Normandy . elizabeth.ii58@yahoo.com

Masked Text: "They have been part of the Duchy of Normandy since the tenth century , and [REDACTED_NAME] is often referred to by her traditional and conventional title of Duke of Normandy . [REDACTED_EMAIL]"

Accuracy: 0.25

Precision: 1.00

Recall: 0.08

F1-Score: 0.14

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.92

Original Text: Further mention of events in Kent occurs in the late sixth century history of the Franks by Gregory of Tours . gregory.of78@gmail.com

Masked Text: Further mention of events in Kent occurs in the late sixth century history of the Franks by [REDACTED_NAME] Gregory of Tours. [REDACTED_EMAIL]

Accuracy: 0.18

Precision: 0.50

Recall: 0.06

F1-Score: 0.10

False Positive Rate (FPR): 0.25

False Negative Rate (FNR): 0.94

Original Text: During this time , Jeanne Duval became his mistress .
jeanne.duval95@gmail.com

Masked Text: During this time, [REDACTED_NAME] became his mistress.

[REDACTED_EMAIL]

Accuracy: 0.25

Precision: 0.50

Recall: 0.17

F1-Score: 0.25

False Positive Rate (FPR): 0.50

False Negative Rate (FNR): 0.83

Original Text: Utah 's governor , Gary Herbert , is also a church member .
gary.herbert52@yahoo.com

Masked Text: Utah's governor, [REDACTED_NAME] (Gary Herbert), is also a church
member. [REDACTED_EMAIL] (gary.herbert52@yahoo.com)

Accuracy: 0.17

Precision: 0.50

Recall: 0.10

F1-Score: 0.17

False Positive Rate (FPR): 0.50

False Negative Rate (FNR): 0.90

Original Text: Upon joining the club he stated that he was expecting to play as
a striker , rather than as a utility forward , though he admitted he faced a
challenge in dislodging either of free-scoring duo Tom Pope and Lee Hughes .
tom.pope94@fastnu.edu.pk lee.hughes63@outlook.com

Masked Text: Upon joining the club he stated that he was expecting to play as a
striker , rather than as a utility forward , though he admitted he faced a
challenge in dislodging either of free-scoring duo [REDACTED_NAME] and
[REDACTED_NAME]. [REDACTED_EMAIL] [REDACTED_EMAIL]

Accuracy: 0.29

Precision: 1.00

Recall: 0.12

F1-Score: 0.22

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.88

Original Text: Joss Whedon was credited as executive producer throughout the run
of the series , and for the first five seasons (1997 - 2001) he was also the
showrunner , supervising the writing and all aspects of production .
joss.whedon88@yahoo.com

Masked Text: [REDACTED_NAME] was credited as executive producer throughout the
run of the series, and for the first five seasons (1997-2001) he was also the
showrunner, supervising the writing and all aspects of production.

[REDACTED_EMAIL]

Accuracy: 0.15

Precision: 1.00

Recall: 0.07

F1-Score: 0.12

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.93

Original Text: The show was created and hosted by Marc Weiner .
marc.weiner53@fastnu.edu.pk
Masked Text: The show was created and hosted by [REDACTED_NAME].
[REDACTED_EMAIL]
Accuracy: 0.22
Precision: 0.50
Recall: 0.14
F1-Score: 0.22
False Positive Rate (FPR): 0.50
False Negative Rate (FNR): 0.86

Original Text: She was formerly married to baseball player David Justice ,
singer-songwriter Eric Benét , and actor Olivier Martinez .
david.justice41@gmail.com eric.benét79@yahoo.com olivier.martinez11@outlook.com
Masked Text: She was formerly married to baseball player [REDACTED_NAME],
singer-songwriter [REDACTED_NAME], and actor [REDACTED_NAME]. [REDACTED_EMAIL]
[REDACTED_EMAIL] [REDACTED_EMAIL]
Accuracy: 0.31
Precision: 0.67
Recall: 0.31
F1-Score: 0.42
False Positive Rate (FPR): 0.67
False Negative Rate (FNR): 0.69

Original Text: While a primary concern has been the relationship between the
writing process and natural places , concepts of spatiality also apply to
cyberspace and online writing - in MUDs , MOOs , Internet Relay Chat , Instant
Messages , and e-mail (Syverson , 1999 ; Yagelski , 2002) .
syverson85@outlook.com yagelski51@fastnu.edu.pk
Masked Text: While a primary concern has been the relationship between the
writing process and natural places , concepts of spatiality also apply to
cyberspace and online writing - in MUDs , MOOs , Internet Relay Chat , Instant
Messages , and e-mail ([REDACTED_NAME] , 1999 ; [REDACTED_NAME] , 2002) .
[REDACTED_EMAIL] [REDACTED_EMAIL]
Accuracy: 0.21
Precision: 0.50
Recall: 0.05
F1-Score: 0.09
False Positive Rate (FPR): 0.18
False Negative Rate (FNR): 0.95

Original Text: As a result , he was unable to play again in the series and
Shakib Al Hasan took over . shakib.al37@fastnu.edu.pk
Masked Text: As a result, he was unable to play again in the series and
[REDACTED_NAME] took over. [REDACTED_EMAIL]
Accuracy: 0.29
Precision: 1.00

Recall: 0.14
F1-Score: 0.25
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.86

Original Text: Nedelya Petkova created the first Bulgarian school for women in the city . nedelya.petkova14@yahoo.com
Masked Text: [REDACTED_NAME] created the first Bulgarian school for women in the city. [REDACTED_EMAIL]
Accuracy: 0.17
Precision: 1.00
Recall: 0.17
F1-Score: 0.29
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.83

Original Text: In 1983 , he was cast in the title role in " Sadat " , a miniseries which chronicled the life and assassination of Anwar Sadat . anwar.sadat44@outlook.com
Masked Text: In 1983, he was cast in the title role in "Sadat", a miniseries which chronicled the life and assassination of Anwar Sadat. [REDACTED_EMAIL]
Accuracy: 0.17
Precision: 0.00
Recall: 0.00
F1-Score: 0.00
False Positive Rate (FPR): 0.20
False Negative Rate (FNR): 1.00

Original Text: Her parts were of a silent movie star in the 1920 s and a deaf librarian in the 1970 s ; in preparation , she studied sign language and watched the films of Lillian Gish . lillian.gish53@fastnu.edu.pk
Masked Text: Her parts were of a silent movie star in the 1920s and a deaf librarian in the 1970s; in preparation, she studied sign language and watched the films of [REDACTED_NAME]. [REDACTED_EMAIL]
Accuracy: 0.26
Precision: 1.00
Recall: 0.08
F1-Score: 0.15
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.92

Original Text: He began publishing articles in Henri Berr 's " Revue de Synthèse Historique " . henri.berr34@outlook.com
Masked Text: He began publishing articles in [REDACTED_NAME]'s "Revue de Synthèse Historique". [REDACTED_EMAIL]
Accuracy: 0.36
Precision: 1.00
Recall: 0.22

F1-Score: 0.36
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.78

Original Text: His most successful album of this period was " Crises " (1983)
, which features the worldwide hit single " Moonlight Shadow " with vocalist
Maggie Reilly . maggie.reilly86@fastnu.edu.pk
Masked Text: His most successful album of this period was "Crises" (1983), which
features the worldwide hit single "Moonlight Shadow" with vocalist
[REDACTED_NAME] Maggie Reilly. [REDACTED_EMAIL] maggie.reilly86@fastnu.edu.pk
Accuracy: 0.28
Precision: 1.00
Recall: 0.10
F1-Score: 0.18
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.90

Original Text: As his behaviour becomes more erratic , Reggie is unable to
dictate letters without uttering words like " breast " . reggie83@gmail.com
Masked Text: "As his behaviour becomes more erratic , [REDACTED_NAME] is unable
to dictate letters without uttering words like " breast " . [REDACTED_EMAIL]"
Accuracy: 0.18
Precision: 0.50
Recall: 0.06
F1-Score: 0.10
False Positive Rate (FPR): 0.25
False Negative Rate (FNR): 0.94

Original Text: In August 2020 , Halford confirmed that he had completed an
audiobook version of the publication , to be made available at the same time as
the physical release . halford90@fastnu.edu.pk
Masked Text: In August 2020, [REDACTED_NAME] confirmed that he had completed an
audiobook version of the publication, to be made available at the same time as
the physical release. [REDACTED_EMAIL]
Accuracy: 0.21
Precision: 1.00
Recall: 0.08
F1-Score: 0.15
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.92

Original Text: Rome outgrew the Servian Wall , but no more walls were
constructed until almost 700 years later , when , in 270 AD , Emperor Aurelian
began building the Aurelian Walls . aurelian56@yahoo.com
Masked Text: "Rome outgrew the Servian Wall , but no more walls were constructed
until almost 700 years later , when , in 270 AD , Emperor [REDACTED_NAME] began
building the [REDACTED_NAME] Walls . [REDACTED_EMAIL]"
Accuracy: 0.24

Precision: 0.67
Recall: 0.08
F1-Score: 0.14
False Positive Rate (FPR): 0.14
False Negative Rate (FNR): 0.92

Original Text: It takes its name from the Greek word for a rainbow , which is also the name for the Greek goddess of the rainbow , Iris . iris40@fastnu.edu.pk
Masked Text: It takes its name from the Greek word for a rainbow, which is also the name for the Greek goddess of the rainbow, Iris. [REDACTED_EMAIL]

Accuracy: 0.24
Precision: 1.00
Recall: 0.05
F1-Score: 0.10
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.95

Original Text: There are , however , matrices for which the reverse Cuthill-McKee algorithm performs better . cuthill-mckee88@fastnu.edu.pk
Masked Text: There are , however , matrices for which the reverse Cuthill-McKee algorithm performs better . [REDACTED_EMAIL]

Accuracy: 0.25
Precision: 1.00
Recall: 0.08
F1-Score: 0.14
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.92

Original Text: A large group of artists , including Martin Scorsese and Alec Baldwin , and scholars signed a legal brief arguing the film 's artistic merit ; the Ohio state court dismissed the case because the police violated the owners ' Fourth Amendment rights , without reaching the question of whether the film was obscene . martin.scorsese40@outlook.com alec.baldwin79@fastnu.edu.pk

Masked Text: A large group of artists, including [REDACTED_NAME] and [REDACTED_NAME], and scholars signed a legal brief arguing the film's artistic merit; the Ohio state court dismissed the case because the police violated the owners' Fourth Amendment rights, without reaching the question of whether the film was obscene. [REDACTED_EMAIL] [REDACTED_EMAIL]

Accuracy: 0.23
Precision: 1.00
Recall: 0.10
F1-Score: 0.18
False Positive Rate (FPR): 0.00
False Negative Rate (FNR): 0.90

Original Text: Williamina Fleming classified most of the spectra in this catalogue and was credited with classifying over 10,000 featured stars and discovering 10 novae and more than 200 variable stars .

williamina.fleming23@yahoo.com
Masked Text: [REDACTED_NAME] classified most of the spectra in this catalogue
and was credited with classifying over 10,000 featured stars and discovering 10
novae and more than 200 variable stars. [REDACTED_EMAIL]

Accuracy: 0.14

Precision: 1.00

Recall: 0.07

F1-Score: 0.14

False Positive Rate (FPR): 0.00

False Negative Rate (FNR): 0.93

Original Text: This introduced him to Jules Laforgue , Arthur Rimbaud , and Paul
Verlaine . jules.laforgue83@yahoo.com arthur.rimbaud92@outlook.com
paul.verlaine44@fastnu.edu.pk

Masked Text: This introduced him to [REDACTED_NAME] , [REDACTED_NAME] , and
[REDACTED_NAME]. [REDACTED_EMAIL] [REDACTED_EMAIL] [REDACTED_EMAIL]

Accuracy: 0.55

Precision: 0.83

Recall: 0.56

F1-Score: 0.67

False Positive Rate (FPR): 0.50

False Negative Rate (FNR): 0.44

=== Overall Dataset Results ===

Overall Accuracy: 0.25

Overall Precision: 0.83

Overall Recall: 0.13

Overall F1-Score: 0.21

Overall False Positive Rate (FPR): 0.14

Overall False Negative Rate (FNR): 0.87

=====