# HotSwapPII Technical User Manual

# Table of Contents

# 1. Introduction

HotSwapPII is a comprehensive system for detecting and anonymizing personally identifiable information (PII) in text documents. The application provides an interactive interface with multiple detection engines, anonymization methods, and evaluation capabilities.

# 2. Installation

Recommended to setup up a fresh virtual environment before installation below.

## 2.1 Method 1: Using Poetry (Recommended)

i. Install Poetry if not already installed: pip install poetry
ii. Install dependencies: poetry install
iii. Run the application: poetry run streamlit run app.py

## 2.2 Method 2: Using pip

i. Install dependencies: pip install -r requirements.txt
ii. Run the application: streamlit run app.py

# 3. Main Features

- Multiple NER engines support (Presidio, SpaCy, HuggingFace Transformers, GLiNER)
- Various anonymization methods
- Synthetic data generation
- Model evaluation capabilities
- Entity-type performance metrics
- Customizable detection settings

# 4. Sidebar Guide

The sidebar has global settings that can be applied to all the features across the tabs.

This includes at a model level:

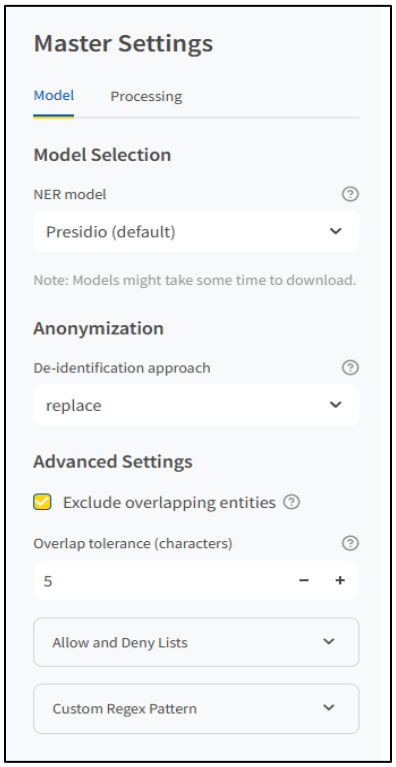



Figure 1: Model Settings

a. A dropdown for model selection
b. Anonymization format
c. Handling of overlapping entities (if overlapping entities should be excluded or not)
d. Overlap tolerance (how many characters of overlap should be considered as "overlapping")
e. Allow lists and deny lists (this is specifically for Presdio models)
f. Custom Regex Patterns (this is specifically for Presdio models, only persists during session)

At the entity selection level we have:

a. Minimum confidence threshold for predictions to qualify
b. Entity selection (which entities to detect for)
    i. This list contains the entities defined by the Presidio framework. The config files contain an exhaustive mapping between Presidio entities and other model entities and the expected ground truth entities. This list can be changed to match the ground truth entities instead

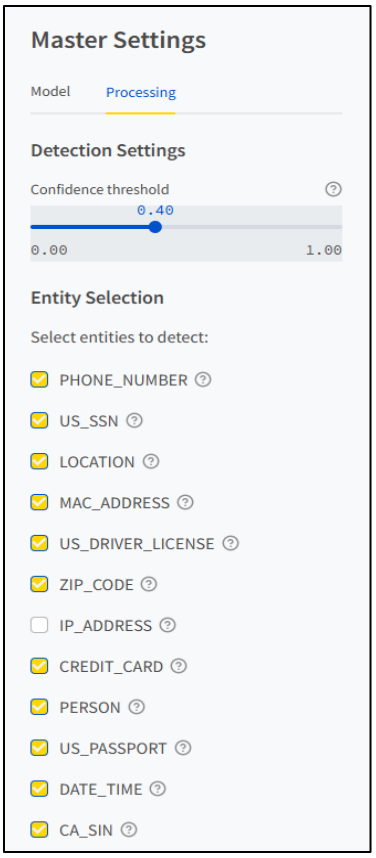



Figure 2: Session settings

# 5. Tab-by-Tab Guide

## 5.1. PII Detection Tab

This is the main tab for processing text and detecting PII.

How to Use:

1. Input text either by:
   a. Typing/pasting directly into the text area
   b. Uploading a text file

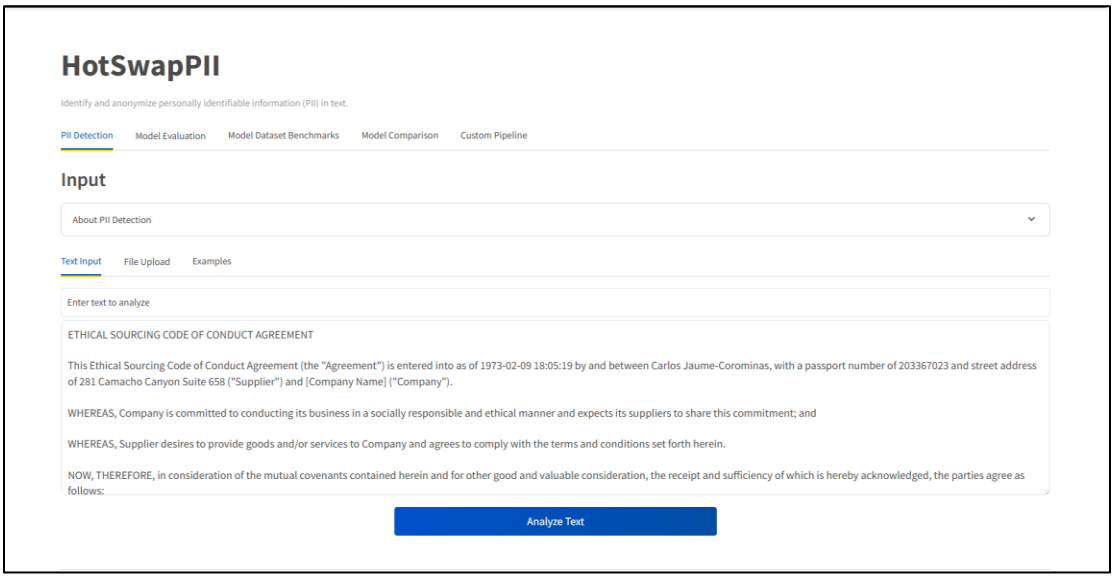



Figure 3: Typing/pasting directly into the text area

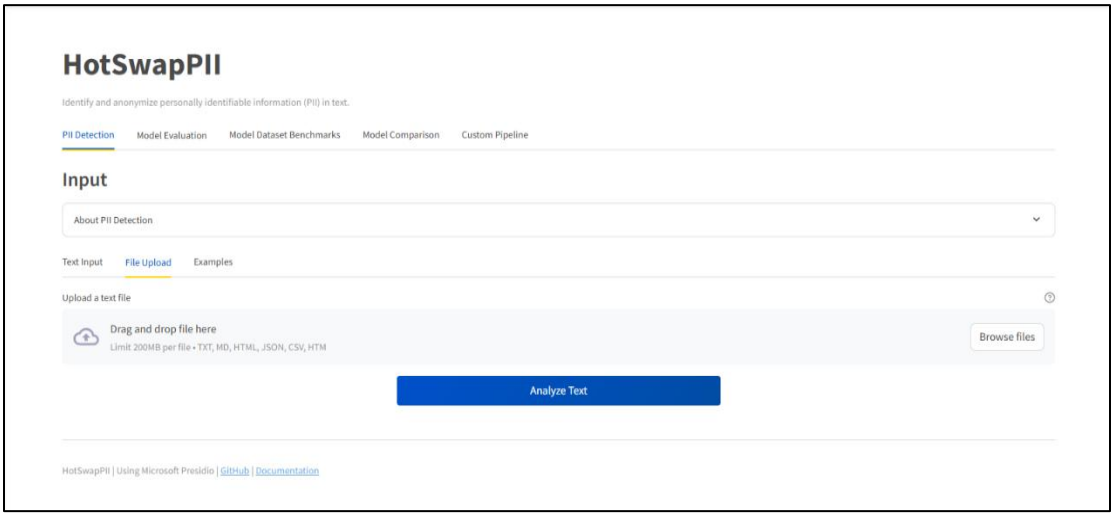
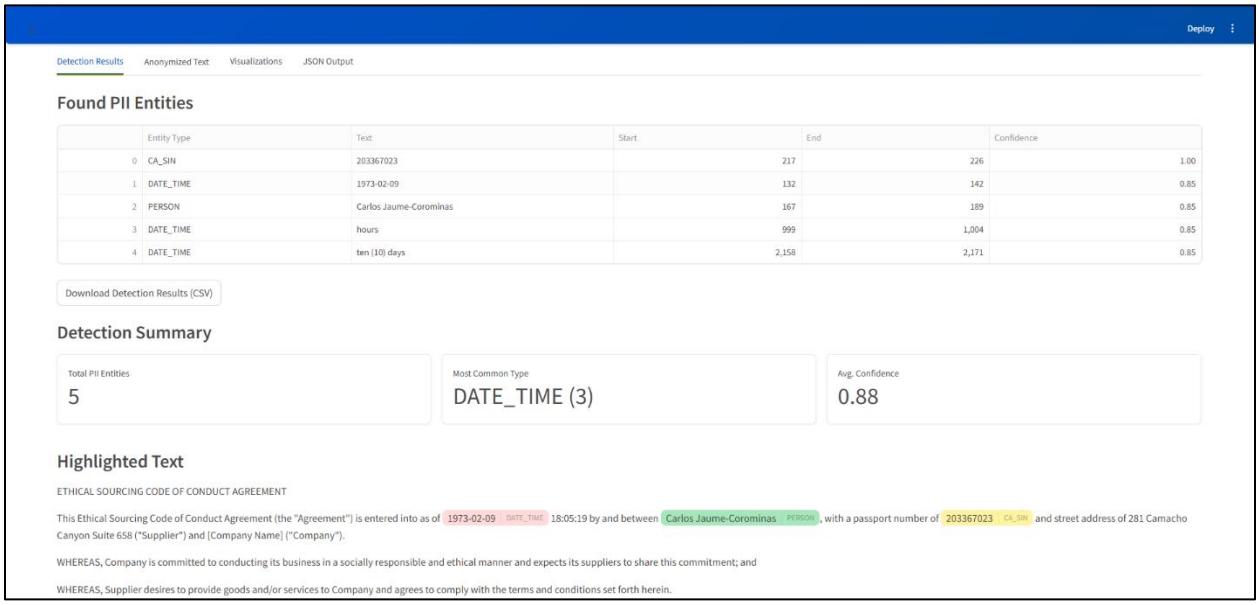



Figure 4: Uploading a text file

2. Settings for processing will be picked from the Sidebar configuration
3. Click "Analyze Text" to process
4. The results have multiple features split over different tabs
   a. Detection Results tab (View detection results)
      i. Found entities




| | Entity Type | Text | Start | End | Confidence |
|---|---|---|---|---|---|
| 0 | CA_SIN | 203367023 | 217 | 226 | 1.00 |
| 1 | DATE_TIME | 1973-02-09 | 132 | 142 | 0.85 |
| 2 | PERSON | Carlos Jaume-Corominas | 167 | 189 | 0.85 |
| 3 | DATE_TIME | hours | 999 | 1,004 | 0.85 |
| 4 | DATE_TIME | ten (10) days | 2,158 | 2,171 | 0.85 |

Figure 5: Found entities

      ii. Highlighted text

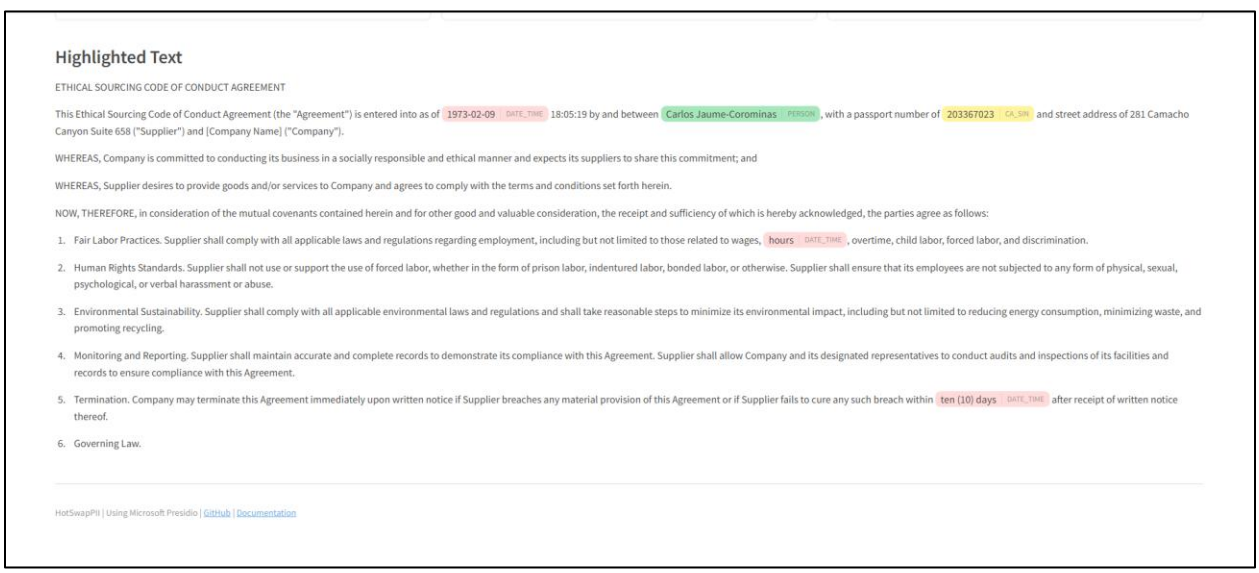



Figure 6: Highlighted text

b. Anonymized Text tab (View the anonymized text according to the global settings for anonymization)


Figure 7: Anonymized Text tab

c. Visualizations tab (View analytics related to the result)

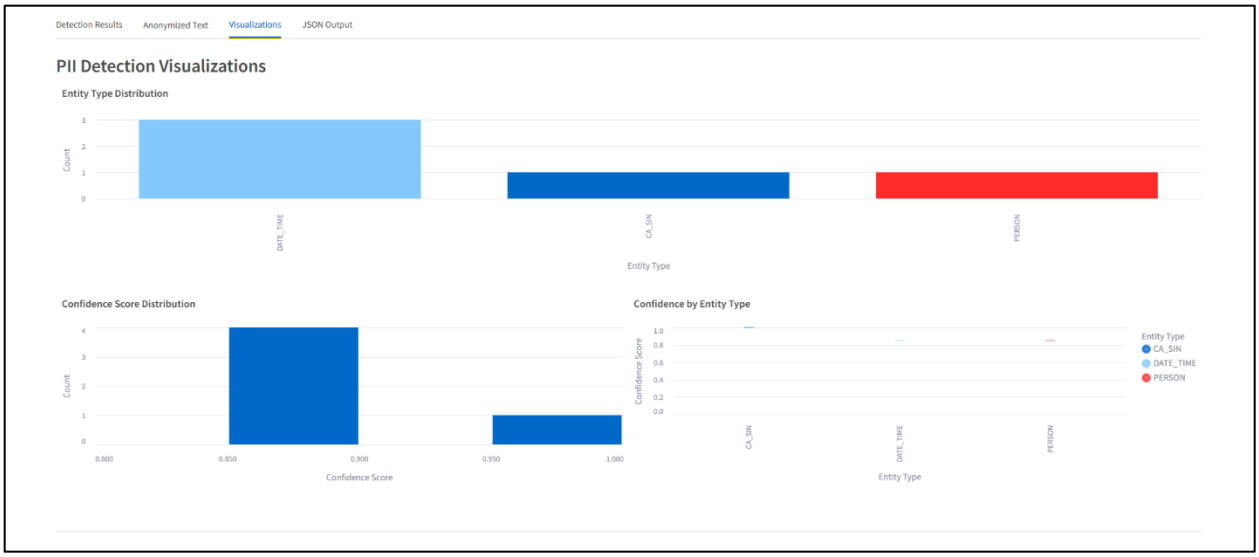



Figure 8: PII Detection Visualization

d. JSON Output tab (View the results as a JSON output)


Figure 9: JSON Output tab

## 5.2. Model Evaluation Tab

Used for evaluating model performance against labeled datasets.

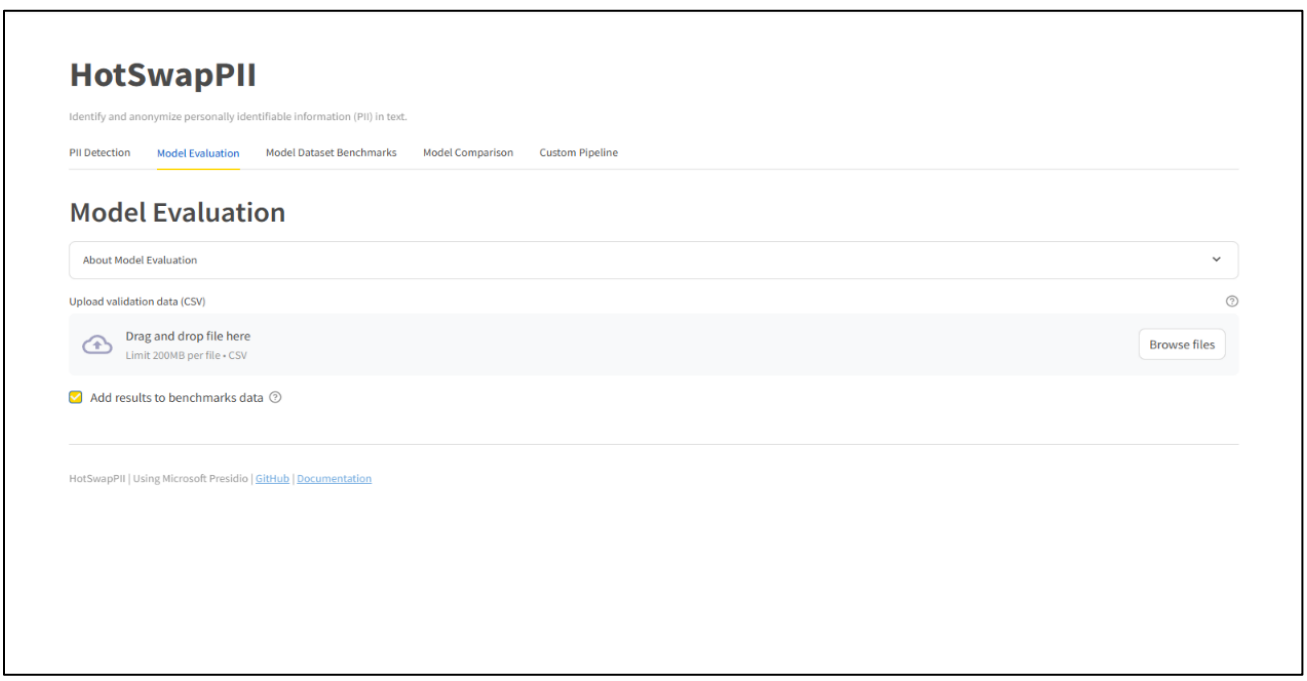



Figure 10: Model Evaluation

**How to Use:**

1. Upload a CSV file with data in the format as outline in the tooltip. The data should contain:
    a. "text" column with input text
    b. "label" column with JSON annotations
2. Choose model (from sidebar)
3. Configure evaluation settings
4. Run evaluation
5. View results (further explained below)


Figure 11: Evaluation Setting

Optionally, one can also choose to update the benchmark dataset metrics if using one of the benchmark datasets for the evaluation. This is ideal when wanting to test out changes to models and see how the changes affect the results on benchmark datasets. This will only work however if the name of the dataset being evaluated is the same as one of the datasets in the benchmark data folder/config. If you want to add a new dataset to the benchmark collection, the config list must be updated with the relevant dataset's file name. The dataset csv must also be added to the benchmark datasets folder to be able to view the sample data.

```python
# File paths for benchmark dataset samples
DATASET_SAMPLE_FILE_PATH: List[str] = [
    "./data/benchmark_datasets/1_original_gretel_ai_conformance_data_500.csv",
    "./data/benchmark_datasets/2_another_gretel_ai_data_500.csv",
    "./data/benchmark_datasets/3_simple_generated_data_500.csv",
    "./data/benchmark_datasets/4_fake_data_with_variance_final.csv",
    "./data/benchmark_datasets/5_enron_data.csv"
]

# File paths for benchmark results
DATASET_BENCHMARK_RESULTS_FILE_PATH: List[str] = [
    "./data/benchmark_results/1_original_gretel_ai_conformance_data_500_results.json",
    "./data/benchmark_results/2_second_gretel_ai_data_500_results.json",
    "./data/benchmark_results/3_simple_generated_data_500_results.json",
    "./data/benchmark_results/4_fake_data_with_variance_final_results.json",
    "./data/benchmark_results/5_enron_data_results.json"
]
```

Figure 12: config/benchmark_config.py

The evaluation results include several different metrics that can be quite useful.

Firstly, there are two methods of metric calculations you can view results by:

a. NERvaluate (https://github.com/MantisAI/nervaluate)
    i. This library provides further metric schemas to use when calculating model performance i.e. matching by Type Partial etc. Viewing the results for each schema can be chosen from its specific tab

b. In House (our own metrics calculation algorithm)

Once you choose your calculation method there are multiple result tabs. For each type of calculation, you can view results at different granularities:

a. Summary tab: This tab contains overall metrics for the model. Here you can see how it performed on the overall dataset.

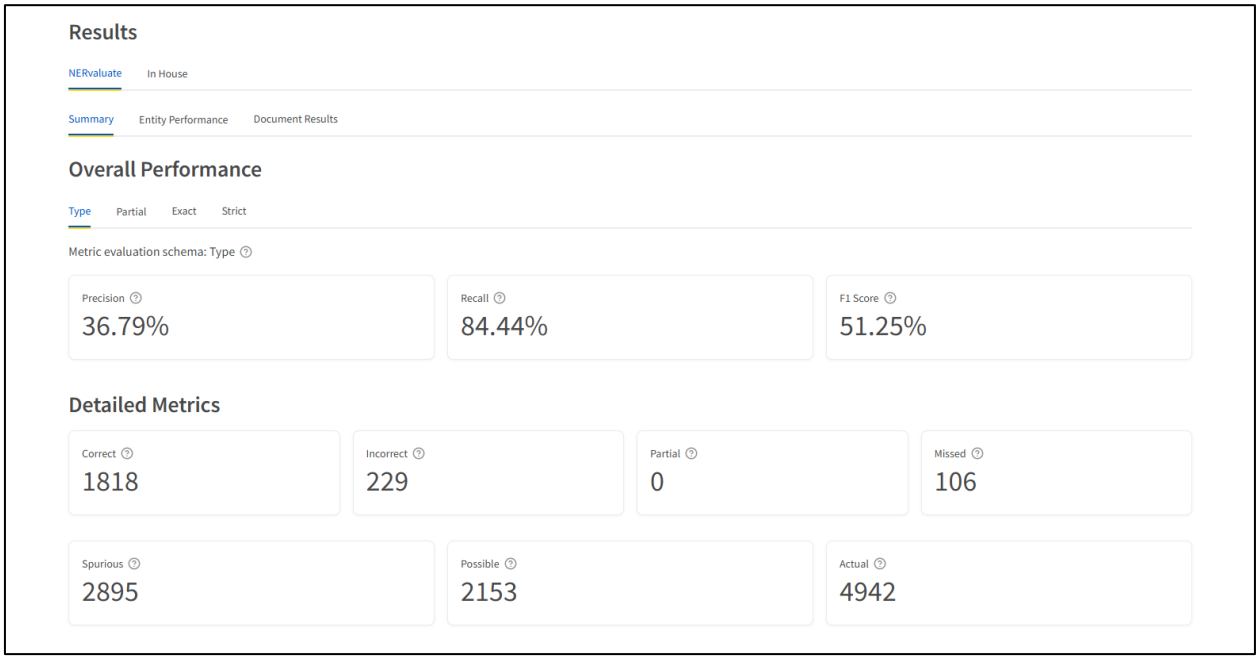
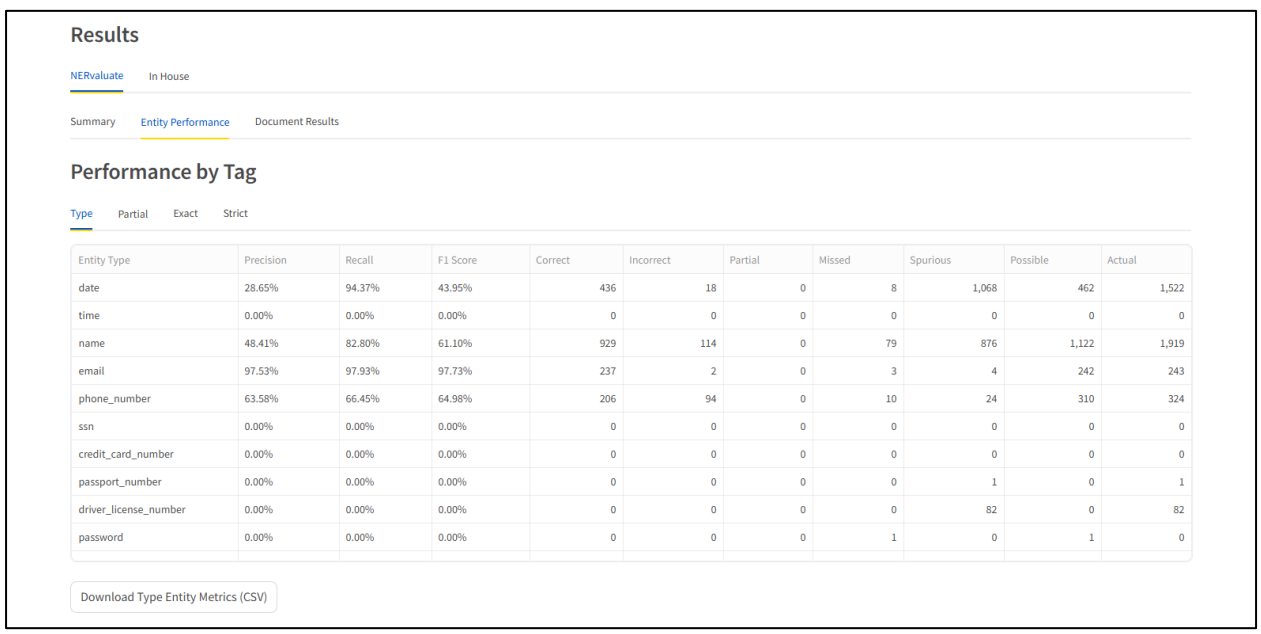



Figure 13: Results - Overall Performance

b. Entity Performance tab: This tab contains metrics broken down by PII type. Here you can see how the model performed per tag along with raw values for number of predictions.

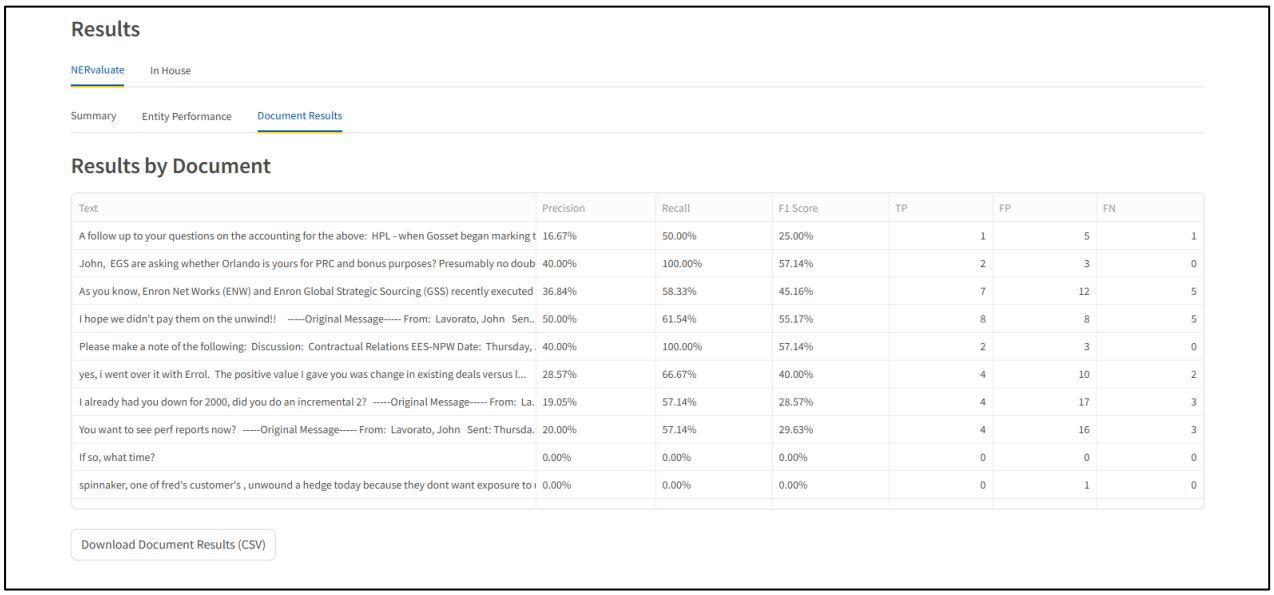



| Entity Type | Precision | Recall | F1 Score | Correct | Incorrect | Partial | Missed | Spurious | Possible | Actual |
|---|---|---|---|---|---|---|---|---|---|---|
| date | 28.65% | 94.37% | 43.95% | 436 | 18 | 0 | 8 | 1,068 | 462 | 1,522 |
| time | 0.00% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| name | 48.41% | 82.80% | 61.10% | 929 | 114 | 0 | 79 | 876 | 1,122 | 1,919 |
| email | 97.53% | 97.93% | 97.73% | 237 | 2 | 0 | 3 | 4 | 242 | 243 |
| phone_number | 63.58% | 66.45% | 64.98% | 206 | 94 | 0 | 10 | 24 | 310 | 324 |
| ssn | 0.00% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| credit_card_number | 0.00% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| passport_number | 0.00% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| driver_license_number | 0.00% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | 82 | 0 | 82 |
| password | 0.00% | 0.00% | 0.00% | 0 | 0 | 0 | 1 | 0 | 1 | 0 |

Figure 14: Results - Performance by Tag

c. Document Results tab:

   i. This is a document wise breakdown of metrics and raw counts of predictions



Results

NERvaluate  In House

Summary  Entity Performance  Document Results

Results by Document

| Text | Precision | Recall | F1 Score | TP | FP | FN |
|---|---|---|---|---|---|---|
| A follow up to your questions on the accounting for the above: HPL - when Gosset began marking t | 16.67% | 50.00% | 25.00% | 1 | 5 | 1 |
| John, EGS are asking whether Orlando is yours for PRC and bonus purposes? Presumably no doub | 40.00% | 100.00% | 57.14% | 2 | 3 | 0 |
| As you know, Enron Net Works (ENW) and Enron Global Strategic Sourcing (GSS) recently executed | 36.84% | 58.33% | 45.16% | 7 | 12 | 5 |
| I hope we didn't pay them on the unwind!! -----Original Message----- From: Lavorato, John Sen.. | 50.00% | 61.54% | 55.17% | 8 | 8 | 5 |
| Please make a note of the following: Discussion: Contractual Relations EES-NPW Date: Thursday, | 40.00% | 100.00% | 57.14% | 2 | 3 | 0 |
| yes, i went over it with Errol. The positive value I gave you was change in existing deals versus l... | 28.57% | 66.67% | 40.00% | 4 | 10 | 2 |
| I already had you down for 2000, did you do an incremental 2? -----Original Message----- From: La. | 19.05% | 57.14% | 28.57% | 4 | 17 | 3 |
| You want to see perf reports now? -----Original Message----- From: Lavorato, John Sent: Thursda. | 20.00% | 57.14% | 29.63% | 4 | 16 | 3 |
| If so, what time? | 0.00% | 0.00% | 0.00% | 0 | 0 | 0 |
| spinnaker, one of fred's customer's , unwound a hedge today because they dont want exposure to | 0.00% | 0.00% | 0.00% | 0 | 1 | 0 |

Download Document Results (CSV)

Figure 15: Result by Document

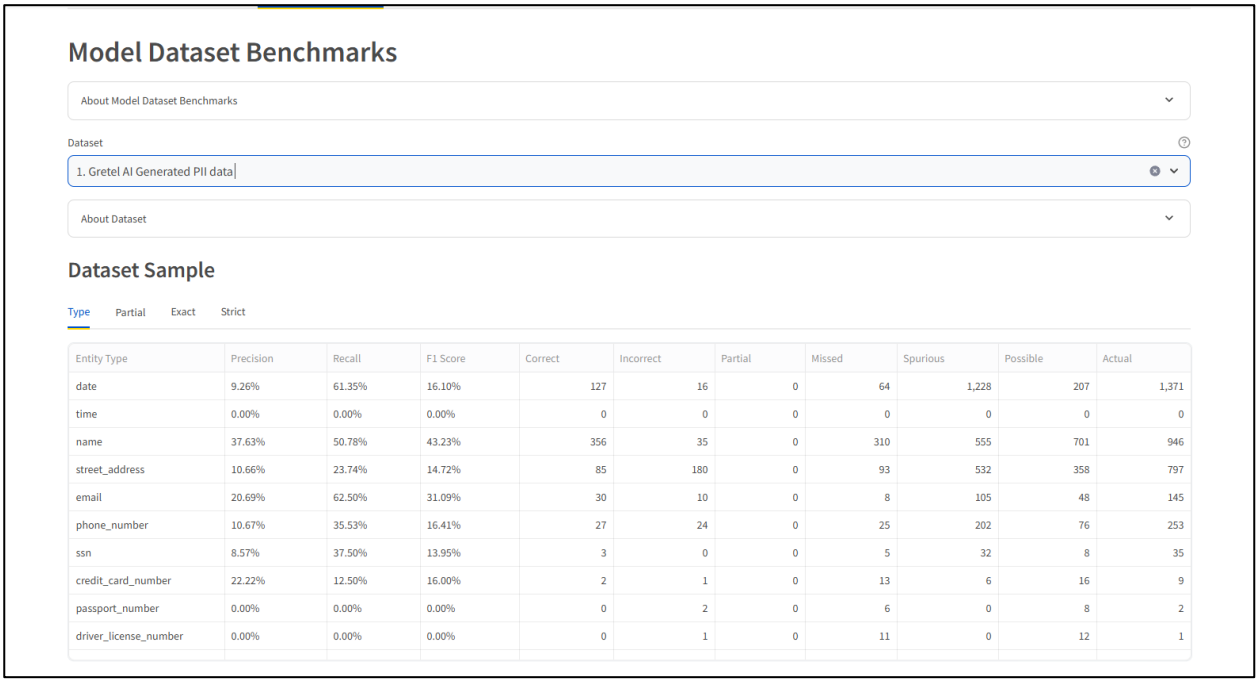
## 5.3. Model Dataset Benchmarks Tab

Compare model performance across different datasets. The datasets along with their file paths are located in the benchmark_config.py file as mentioned above.

How to Use:

1. Select benchmark dataset from the dropdown
2. Choose model to benchmark (from the sidebar)
3. View dataset results
4. Choose which metric schema to view results by (Type, Partial etc)

5. Download as csv button below the results



## Model Dataset Benchmarks

About Model Dataset Benchmarks

Dataset

1. Gretel AI Generated PII data

About Dataset

### Dataset Sample

Type Partial Exact Strict

| Entity Type | Precision | Recall | F1 Score | Correct | Incorrect | Partial | Missed | Spurious | Possible | Actual |
|---|---|---|---|---|---|---|---|---|---|---|
| date | 9.26% | 61.35% | 16.10% | 127 | 16 | 0 | 64 | 1,228 | 207 | 1,371 |
| time | 0.00% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| name | 37.63% | 50.78% | 43.23% | 356 | 35 | 0 | 310 | 555 | 701 | 946 |
| street_address | 10.66% | 23.74% | 14.72% | 85 | 180 | 0 | 93 | 532 | 358 | 797 |
| email | 20.69% | 62.50% | 31.09% | 30 | 10 | 0 | 8 | 105 | 48 | 145 |
| phone_number | 10.67% | 35.53% | 16.41% | 27 | 24 | 0 | 25 | 202 | 76 | 253 |
| ssn | 8.57% | 37.50% | 13.95% | 3 | 0 | 0 | 5 | 32 | 8 | 35 |
| credit_card_number | 22.22% | 12.50% | 16.00% | 2 | 1 | 0 | 13 | 6 | 16 | 9 |
| passport_number | 0.00% | 0.00% | 0.00% | 0 | 2 | 0 | 6 | 0 | 8 | 2 |
| driver_license_number | 0.00% | 0.00% | 0.00% | 0 | 1 | 0 | 11 | 0 | 12 | 1 |

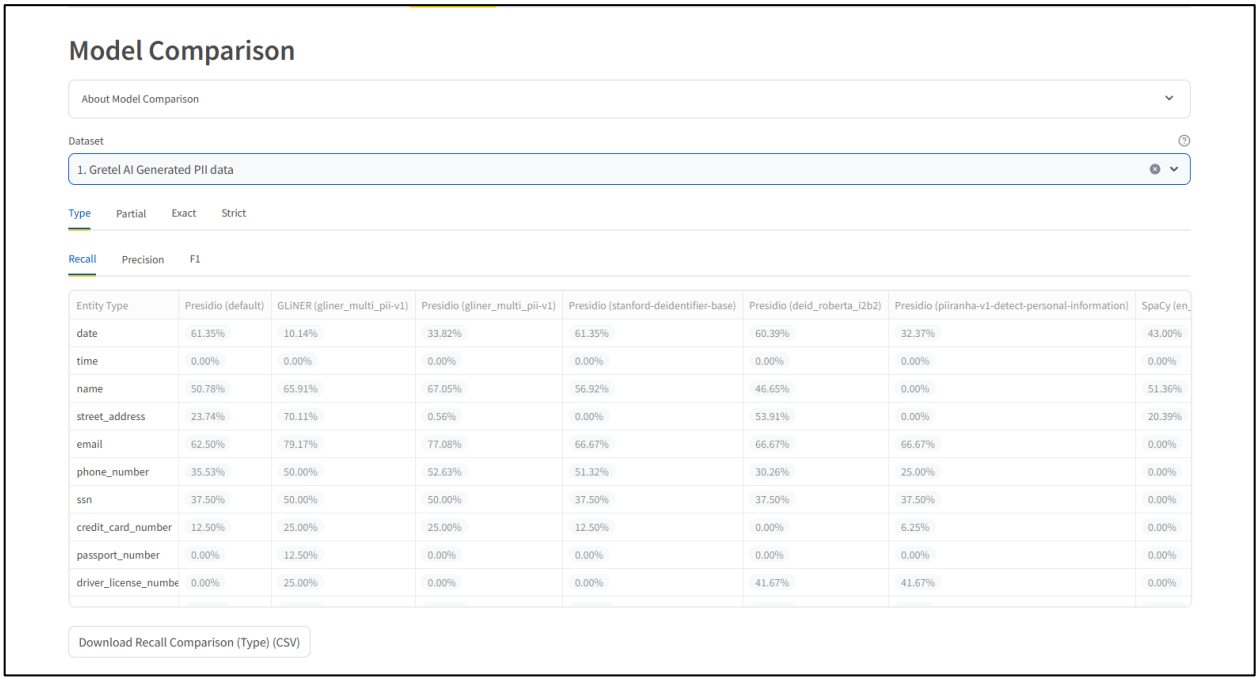
Figure 16: Model Dataset Benchmarks

## 5.4. Model Comparison Tab

Direct comparison between different models. The comparison happens off datasets in the benchmarks tab. Since it uses the results from constructed benchmarks and compares model results for models that have results for that dataset, the model selection in the sidebar is not applied here.

**How to Use:**

1. Select benchmark dataset from the dropdown
2. View side-by-side comparison results
3. Choose which metric schema to view results by (Type, Partial etc)
4. Choose which metric to compare by (F1, Recall, Precision)
5. Download as csv button below the results




| Entity Type | Presidio (default) | GLINER (gliner_multi_pii-v1) | Presidio (gliner_multi_pii-v1) | Presidio (stanford-deidentifier-base) | Presidio (deid_roberta_i2b2) | Presidio (piiranha-v1-detect-personal-information) | SpaCy (en_ |
|---|---|---|---|---|---|---|---|
| date | 61.35% | 10.14% | 33.82% | 61.35% | 60.39% | 32.37% | 43.00% |
| time | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| name | 50.78% | 65.91% | 67.05% | 56.92% | 46.65% | 0.00% | 51.36% |
| street_address | 23.74% | 70.11% | 0.56% | 0.00% | 53.91% | 0.00% | 20.39% |
| email | 62.50% | 79.17% | 77.08% | 66.67% | 66.67% | 66.67% | 0.00% |
| phone_number | 35.53% | 50.00% | 52.63% | 51.32% | 30.26% | 25.00% | 0.00% |
| ssn | 37.50% | 50.00% | 50.00% | 37.50% | 37.50% | 37.50% | 0.00% |
| credit_card_number | 12.50% | 25.00% | 25.00% | 12.50% | 0.00% | 6.25% | 0.00% |
| passport_number | 0.00% | 12.50% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| driver_license_numbe | 0.00% | 25.00% | 0.00% | 0.00% | 41.67% | 41.67% | 0.00% |

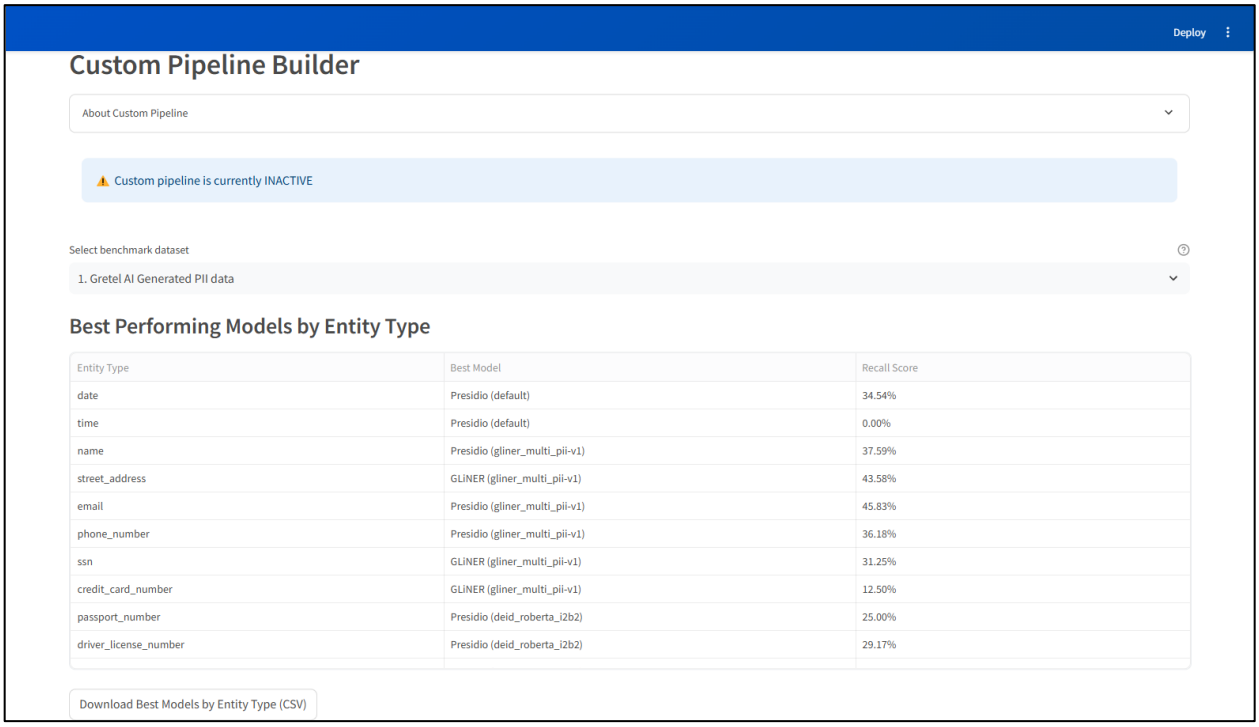
Figure 17: Model Metrics Comparison

## 5.5. Custom Pipeline Tab

Create, explore and manage custom detection pipelines. You can mix and match a custom pipeline here, giving different models precedence of predictions when it comes to specific entities based on previous results or if you want to test out different configurations.

How to Use:

1. (Optional) Select benchmark dataset from the dropdown:
    a. You can select a benchmark dataset that you think performed according to your expectations or in alignment with a particular format of data. The custom pipeline tab will then load the configuration that is best for each entity type given that benchmark dataset.
    b. The metric used for comparison is Recall using the Partial schema




| Entity Type | Best Model | Recall Score |
|---|---|---|
| date | Presidio (default) | 34.54% |
| time | Presidio (default) | 0.00% |
| name | Presidio (gliner_multi_pii-v1) | 37.59% |
| street_address | GLiNER (gliner_multi_pii-v1) | 43.58% |
| email | Presidio (gliner_multi_pii-v1) | 45.83% |
| phone_number | Presidio (gliner_multi_pii-v1) | 36.18% |
| ssn | GLiNER (gliner_multi_pii-v1) | 31.25% |
| credit_card_number | GLiNER (gliner_multi_pii-v1) | 12.50% |
| passport_number | Presidio (deid_roberta_i2b2) | 25.00% |
| driver_license_number | Presidio (deid_roberta_i2b2) | 29.17% |

Figure 18: Custom Pipeline Builder

2. Adjust pipeline configuration accordingly

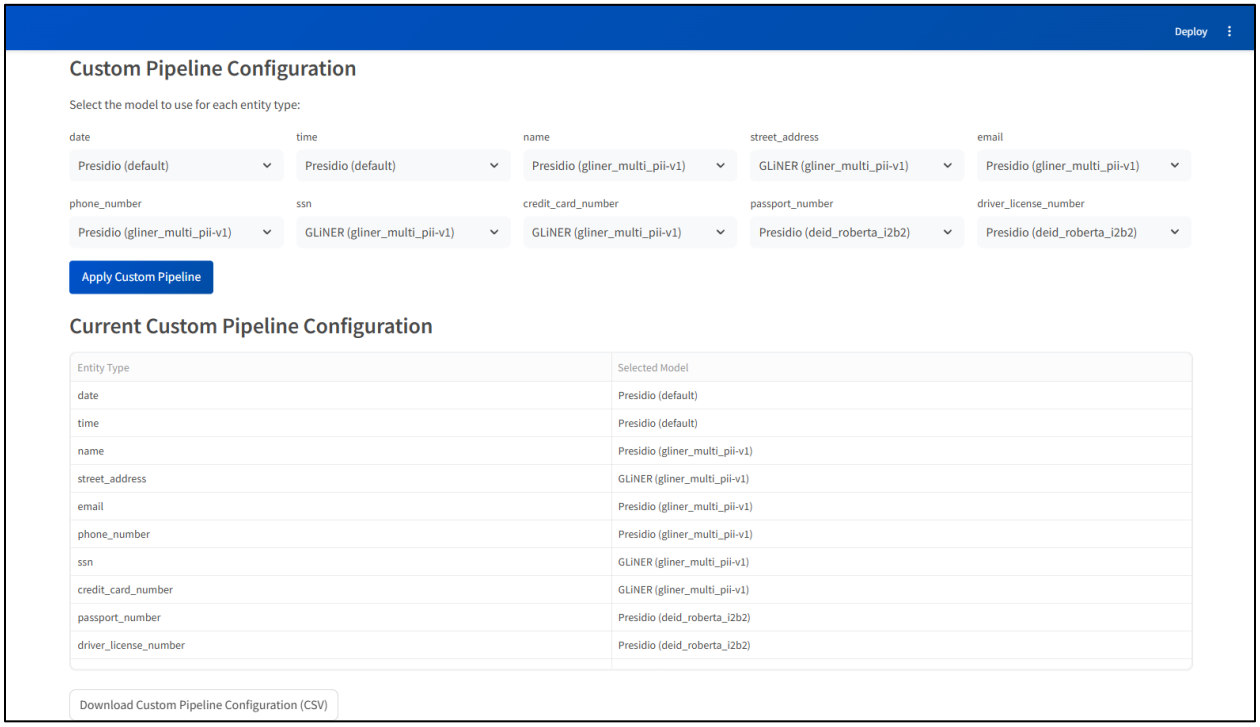



Figure 19: Custom Pipeline Configuration

3. Apply custom pipeline (Activate)
4. Use in the PII detection tab or Evaluation tab as needed (an alert will be present if the custom pipeline is active)
5. Download configuration as csv if needed

# 6. Configuration Guide

## 6.1. Model Configuration

Models can be added permanently by adding them in the models_config.py. Currently only two types of base models and three types of model families are available to set up as follows:

1. Base Model Selection:
    a. Presidio: Microsoft's PII detection engine
    b. Independent: Standalone models
2. Model Family Options:
    a. SpaCy
    b. HuggingFace (any Huggingface Transformer model)
    c. GLiNER

More types of base models and model families can be added but there needs to be a robust set up in accordance with the implementation of the rest of the models to have it work correctly.

## 6.2. Detection Settings

A short summary of the detection settings is as follows:

1. Threshold: Configure confidence threshold (0.0 - 1.0)
2. Entity Types: Select which PII types to detect
3. Overlap Handling: Configure how overlapping entities are managed
4. Custom Lists:
    a. Allow List: Always detect these terms
    b. Deny List: Never detect these terms

## 6.3. Anonymization Options

A short summary of the redaction settings is as follows:

1. Redaction: Complete removal of PII
2. Replacement: Entity type placeholders
3. Masking: Character masking
4. Highlighting: Visual marking
5. Synthesis: AI-generated replacements

## 6.4. Entity Mapping

The entity list in the UI is a list of Presidio default entity types and any entity types loaded in the recognizers.yaml. The entity_config.py file has the settings for the default selected entities on startup (DEFAULT_ENTITY_SELECTION), the entities available for selection in the UI (CORE_ENTITIES), entity mapping for the GLiNER recognizer used with Presidio (GLINER_ENTITY_MAPPING), the list of entities GLiNER searches for (GLINER_LABELS) and a general mapping of ALL model entity types to the ground truth entity types i.e. the entity types expected in the ground truth labelling (MODEL_ENTITIES_TO_STANDARDIZED_ENTITY_MAPPING).

This setup can be changed as required e.g. use Presidio as the entity type for the unified mapping as well as the expected tags in the ground truth labelling.

This manual provides a comprehensive overview of the HotSwapPII system. For specific technical details or advanced configurations, please refer to the source code documentation or reach out to the development team.