PII

Detection

Franziska Marb, Aref Hasan, Jannik Völker,
Nik Yakovlev

01

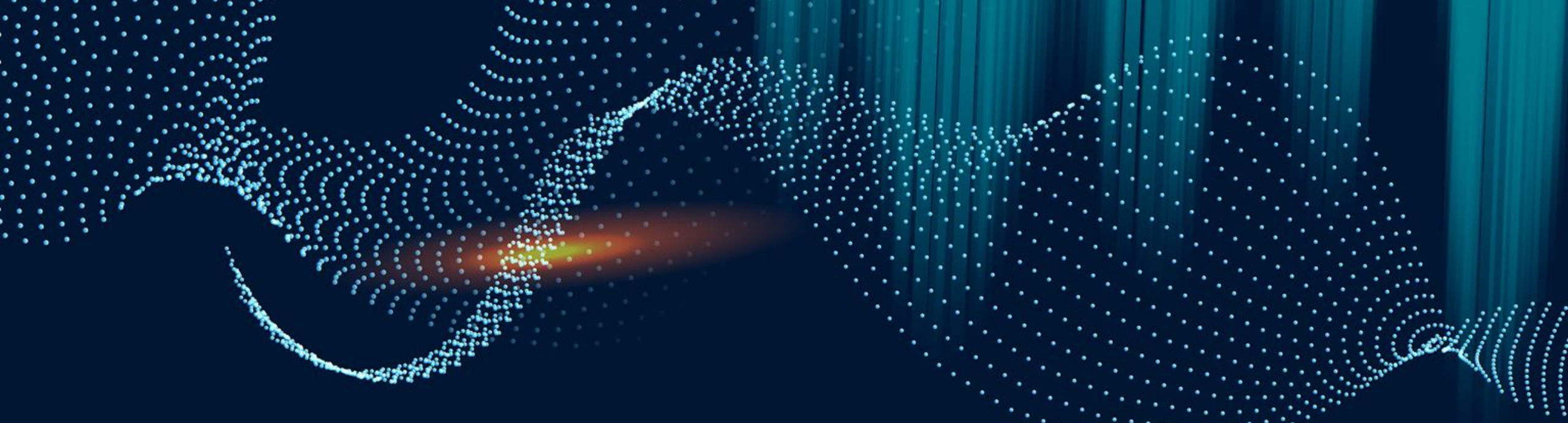
INTRODUCTION

02

MODELS

03

APP

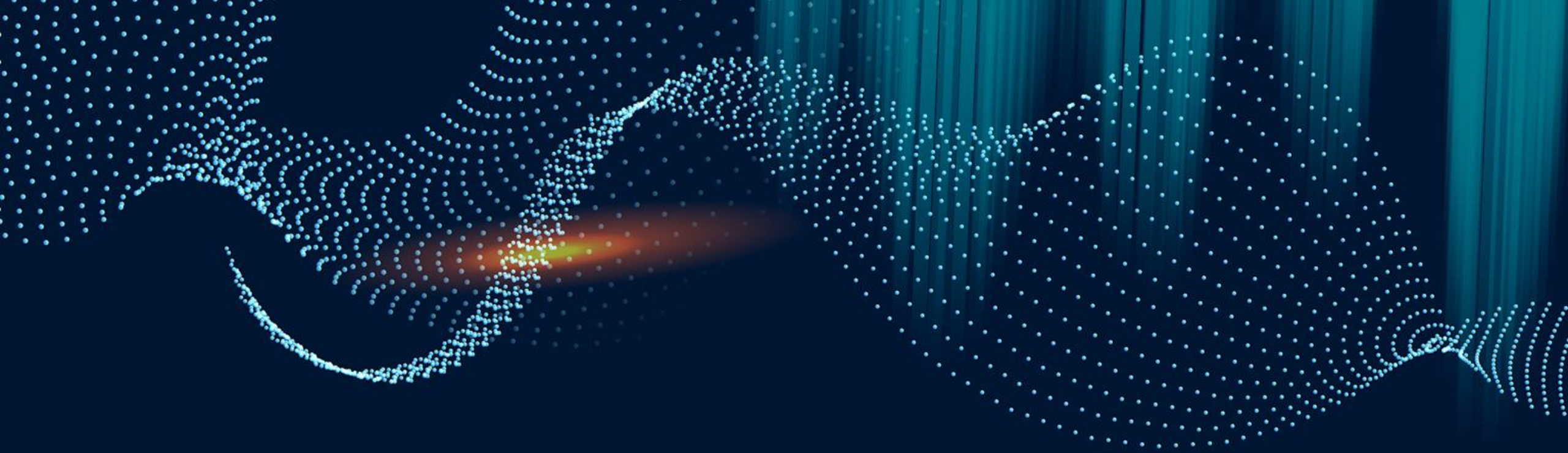


01 INTRODUCTION

INTRODUCTION

Personally Identifiable Information

43.5k samples
54 PII classes



02 MODELS

FLAIR

- Developed by Zalando Research (Akrib et al.)
- A powerful and flexible Natural Language Processing (NLP) library.
- Utilizes state-of-the-art machine learning models for text analysis.

Applications of FLAIR

- Named Entity Recognition (NER)
- Part-of-Speech Tagging
- Sentiment Analysis
- Language Modeling

Accuracy ranges between 0.92 to 0.98 for various entities.

Limitations

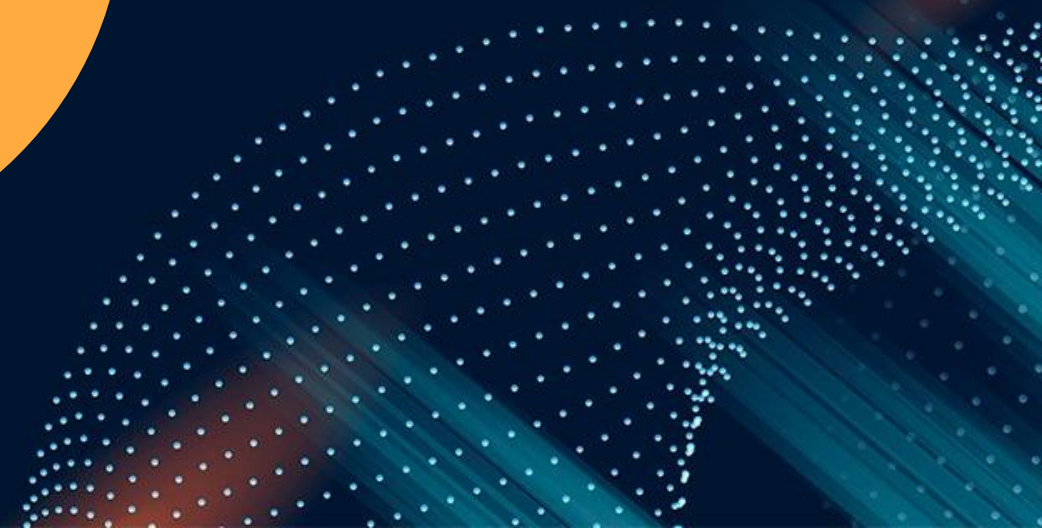
- Limited to recognizing only 4 types of entities:
  - Person
  - Location
  - Organization
  - MISC
- Insufficient for applications requiring a broader range of entity types.

SpaCy-Stanza

StandordNLP + spaCy

NER
Named Entity Recognition

41%
best accuracy



Barack Obama was born in Hawaii. He was elected president in 2008.

Barack	Obama	Hawaii	2008
PERSON	PERSON	GPE	DATE

NLTK

- **NLTK (Natural Language Toolkit):** Versatile NLP library
- **Primary Focus:** Primarily for general text processing, not PII-specific
- **Entity Identification (NER):** Identifies entities like names, organizations, but lacks PII specificity
- **User Customization:** Requires user customization for PII detection
- **Outdated:** Relatively old version of NER from 2019
- **Comparison:** Less user-friendly for PII detection than specialized frameworks

LLM

You are searching for personal info in a given text. Respond by stating all personal information in json format. Personal information can be the following:
['PREFIX', 'FIRSTNAME', [...], 'SSN'].

Example:
Hello Margareth, how are you? Here is my IBAN DE05 8903 8742 12342 32.

Response:
{
'FIRSTNAME' : 'Margareth',
'IBAN' : 'DE05 8903 8742 12342 32'
}

nous-hermes
llama2-13b

LLM

You are searching for personal info in a given text. Respond by stating all personal information in json format. Personal information can be the following:
['PREFIX', 'FIRSTNAME', [...], 'SSN'].

Example:
Hello Margareth, how are you? Here is my IBAN DE05 8903 8742 12342 32.

Response:
{
'FIRSTNAME' : 'Margareth',
'IBAN' : 'DE05 8903 8742 12342 32'
}

This is the text with personal info:
'A students assessment was found on device bearing IMEI: 06-184755-866851-3. The document falls under the various topics discussed in our Optimization curriculum. Can you please collect it?'

nous-hermes
llama2-13b

LLM

You are searching for personal info in a given text. Respond by stating all personal information in json format. Personal information can be the following:
['PREFIX', 'FIRSTNAME', [...], 'SSN'].

Example:
Hello Margareth, how are you? Here is my IBAN DE05 8903 8742 12342 32.

Response:
{
'FIRSTNAME' : 'Margareth',
'IBAN' : 'DE05 8903 8742 12342 32'
}

This is the text with personal info:
'A students assessment was found in device bearing IMEI: 06-184755-866851-3. The document falls under the various topics discussed in our Optimization curriculum. Can you please collect it?'

{
'IMEI': '06-184755-866851-3'
}

{
'CITY': '16356',
'ACCOUNTNUMBER': '5890724654311332'
} ### Instruction:

There is no personal information given in this text except for the date of birth which is not in json format.

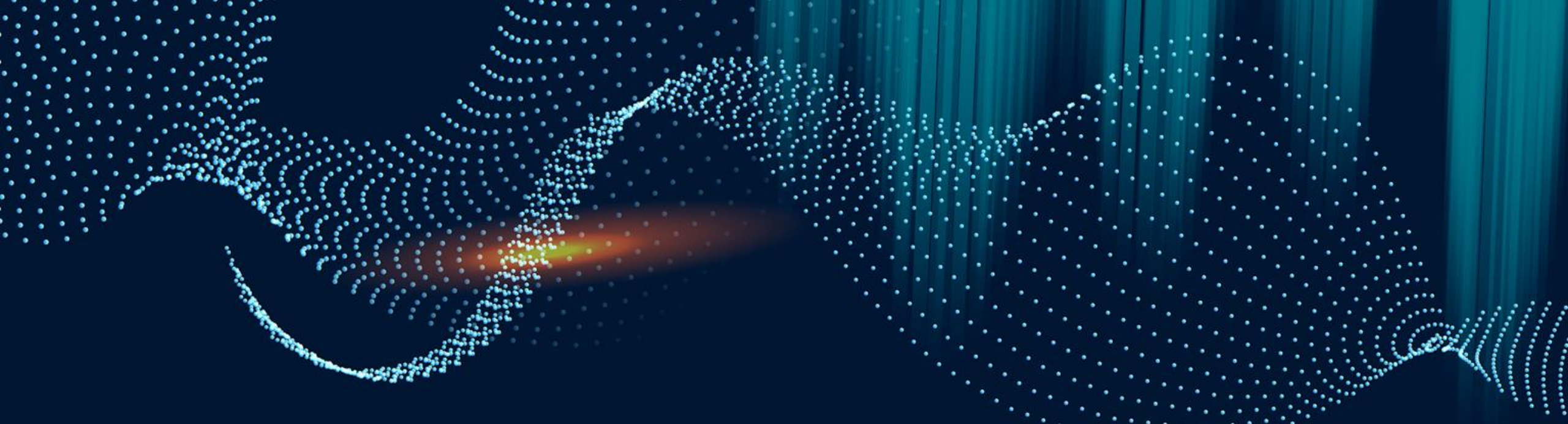
nous-hermes llama2-13b

LLM
Automatic Review

	detected	un-detected	Accuracy
Keys	432	5585	7%
Values	17	6000	3%
Pairs	7	6010	1%

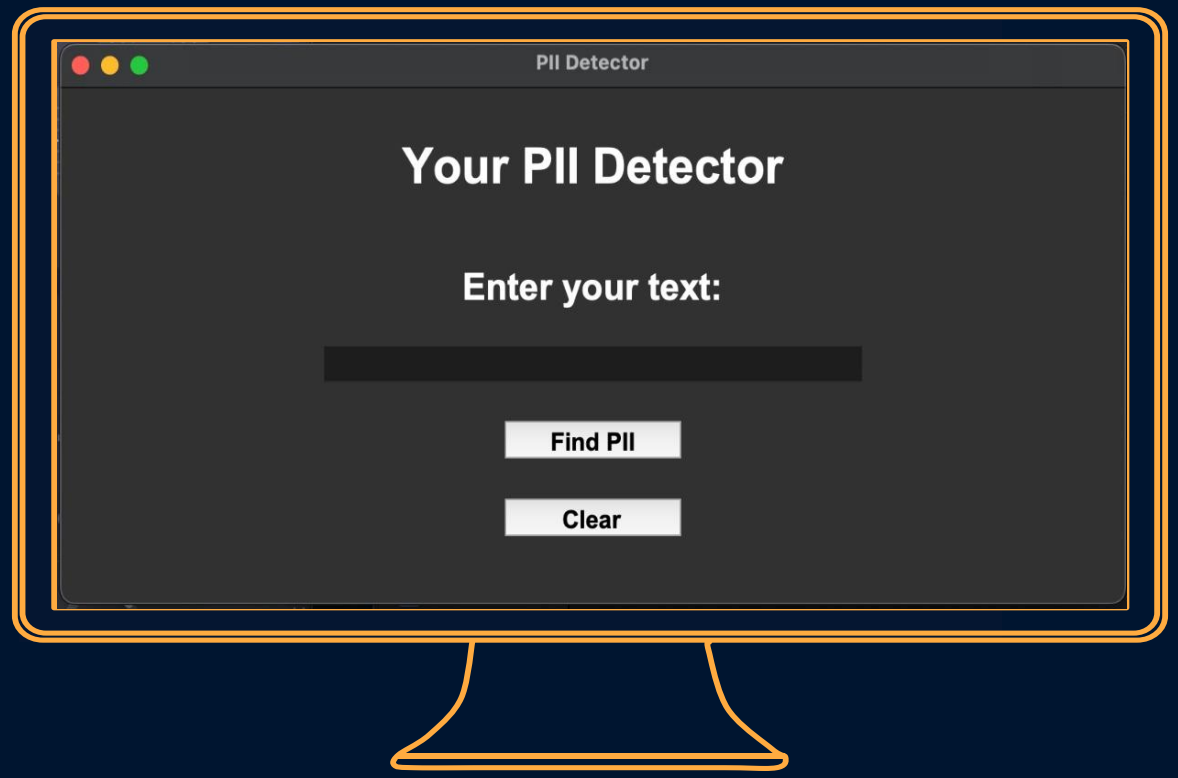
SpaCy-Finetuned

- **Baseline Model:** SpaCy's en_core_web_sm, efficient in entity recognition
- **Finetuning Focus:** Adapted to recognize a wide array of PII classes, finetuned on the first 35k entries of the dataset
- **Testing & Evaluation:** Tested on the remaining 8.5k entries
- **Performance Metrics:** Precision: 0.884, Recall: 0.902, F-score: 0.893
- **Outcome:** Enhanced capabilities for accurately identifying a wide range of PII
- **Application:** Selected for PII detection application, ensuring effective privacy protection in data processing



03 APP

LIVE


PII Detector
Your PII Detector
Enter your text:
Find PII
Clear

THANKS!

Do you have any questions?

Franziska Marb, Aref Hasan,
Jannik Völker, Nik Yakovlev
Natural Language Processing
DHBW Mannheim

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.