

Increase Accuracy with AI-Driven Approach



Personally Identifiable information (PII), is any data that can be used to identify individuals such as name, driver's license number, bank account number, passport number, email address and more.

INTRODUCTION

NEC's offers an AI-tailored solution that can extract personally identifiable information from an input text or the provided documents using the Bidirectional Encoder Representation from Transformers based Named Entity Recognition (BERT – NER) Model.

PII Redaction Data Flow Architecture



- Data collected from different resources in CSV Format

Data Sources



- Cleaning Data
- Convert the raw data into sentences
- Convert the sentences into tokens

Data Cleaning and Tokenization



- Accuracy and performance of model is measured using F1-score, Precision and Recall

Evaluation



- Load the pre-trained BERT model
- Change the last layers of the network for Text labelling

Training BERT

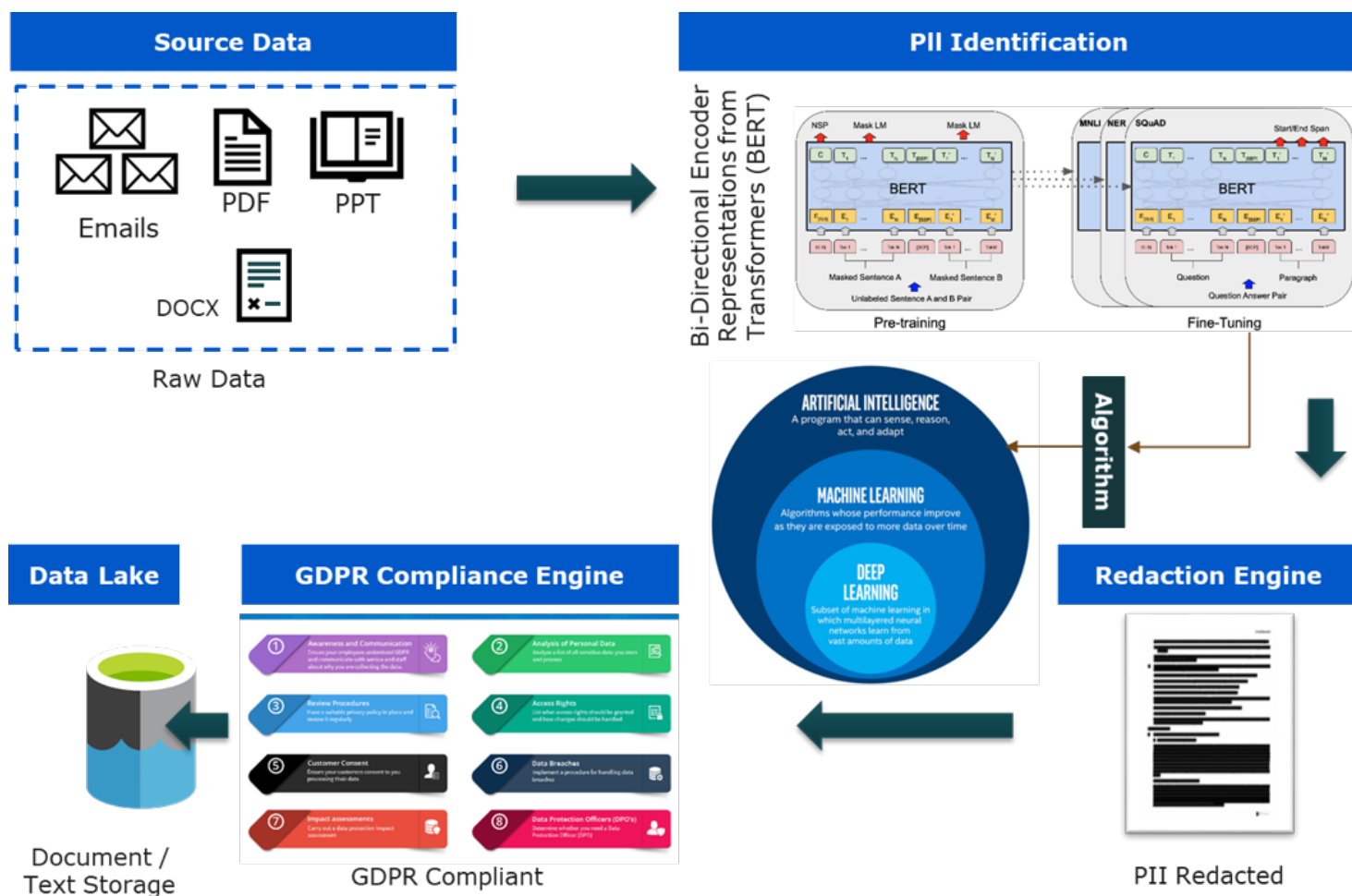




HIGH-LEVEL ARCHITECTURE

NEC replaces the traditional rule-based redaction approach with AI and ML driven approach. The solution works on unstructured and structured data sources and is customizable to different domains. The solution includes:

- Data from multiple sources collected & processed to a structured format (csv / pdf / excel / txt)
- BERT-NER model based training on the above data
- Identifying PII on the documents by using the earlier trained BERT-NER model



\Orchestrating a brighter world



Contact us for more information:

✉ marketing@india.nec.com

🌐 in.nec.com

Copyright © NEC Corporation India Private Limited 2020.
All rights reserved.