PHI Detection and Removal

1 Name	**2** Address	**3** Date
4 Phone Number	**5** Fax Number	**6** Email
7 SSN	**8** Medical Record Number	**9** Health Plan Number
10 Account Number	**11** License Number	**12** Vehicle Number
13 Device ID	**14** Web URL	**15** IP Address
16 Biometric ID	**17** Photo	**18** Other Unique Identifiers

Hybrid Approach: Rule-based + Machine Learning

Rule-based

Detect clear patterns with regular expressions
- Date: MM/DD/YYYY
- Phone: (123) 456-7890
- High precision, low recall

Machine Learning (ML-based)

Context-based detection with NER models
- BiLSTM-CRF, BERT
- spaCy, MedCAT
- High recall, contextual understanding

STEP 1
Original Text
- Input Data

➤

STEP 2
PHI Detection
- NER + Regex

➤

STEP 3
Masking/Removal
- Anonymize

➤

STEP 4
Validation
- Validate