

# PHI Detection and Removal

1 이름	2 주소	3 날짜	4 전화번호	5 팩스번호	6 이메일
7 사회보장번호	8 의료기록번호	9 건강보험번호	10 계좌번호	11 면허번호	12 차량번호
13 기기 ID	14 웹 URL	15 IP 주소	16 생체인식정보	17 사진	18 기타 고유식별자

## 하이브리드 접근법: 규칙 기반 + 머신러닝

### 규칙 기반 (Rule-based)

- 정규표현식으로 명확한 패턴 탐지
- 날짜: MM/DD/YYYY
  - 전화: (123) 456-7890
  - 높은 정밀도, 낮은 재현율

### 머신러닝 (ML-based)

- NER 모델로 문맥 기반 탐지
- BiLSTM-CRF, BERT
  - spaCy, MedCAT
  - 높은 재현율, 문맥 이해

