

### 1. AWS ML Service

AWS에서는 크게 3 가지의 ML 서비스를 제공한다.

#### 1. Comprehend:

- a. ML 기반 자연어 처리 서비스. 통계 ML 관련으로 보이며, 임베딩, 객체 추출 등 사용가능
- b. 장점: 서울 리전 존재 함, 가벼움
- c. 단점: 통계 ML이라 성능 낮음

#### 2. Bedrock:

- a. LLM 제공 서비스. 각종 LLM 모델들을 API 마냥 제공 함.
- b. 장점: 각종 task에 특화 된 LLM을 사용 및 증강검색기법 등 LLM 최적화 기능 제공
- c. 단점: 서울 리전 없음. 대부분의 모델이 입력-출력 토큰에 대한 비용 존재

#### 3. SageMaker:

- a. Fine-tuning 등 커스텀 LLM 제작 및 모델 학습 서비스
- b. 장점: Serverless 단위에서 모델 학습 및 분석 가능
- c. 단점: 서울 리전 없음. 비쌈

LLM을 사용하여 텍스트 데이터 내에서 특정 키워드 및 의미 추출을 위해 Bedrock 서비스를 사용하여, 제공되는 LLM를 활용하는 방법이 적절할 것으로 보인다.

## 2. LLM 고려 목록

본 프로젝트에서 LLM을 정제 단계에서 비정제 텍스트 데이터 대상으로 다음의 Task를 수행할 목적으로 채용한다.

- 채용공고 글에서 특정 기술 스택 용어를 추출 (키워드 추출)
- 채용공고 글을 기반으로 해당 공고의 직무 분류 추출 (텍스트 요약)

따라서 키워드 추출 및 텍스트 요약을 목적으로 LLM을 사용하여 선정한 LLM들은 다음과 같다.

### 1. Claude

- Anthropic
- 자연어 이해 및 정보 추출에 특화
- 장점: AWS Bedrock에서 지원
- 단점: 입출력 토큰에 대한 비용 존재

### 2. GPT

- OpenAI
- 텍스트 분석 추출 요약 특화
- 장점: 성능이 제일 좋을 것으로 추측
- 단점: API로 호출하기 위해 구독 계정과 토큰별 요금 지불 필요

### 3. Gemini

- Google
- 장점: 일정 용량까지 무료, Google 생태계
- 단점: 성능 확인 않됨