

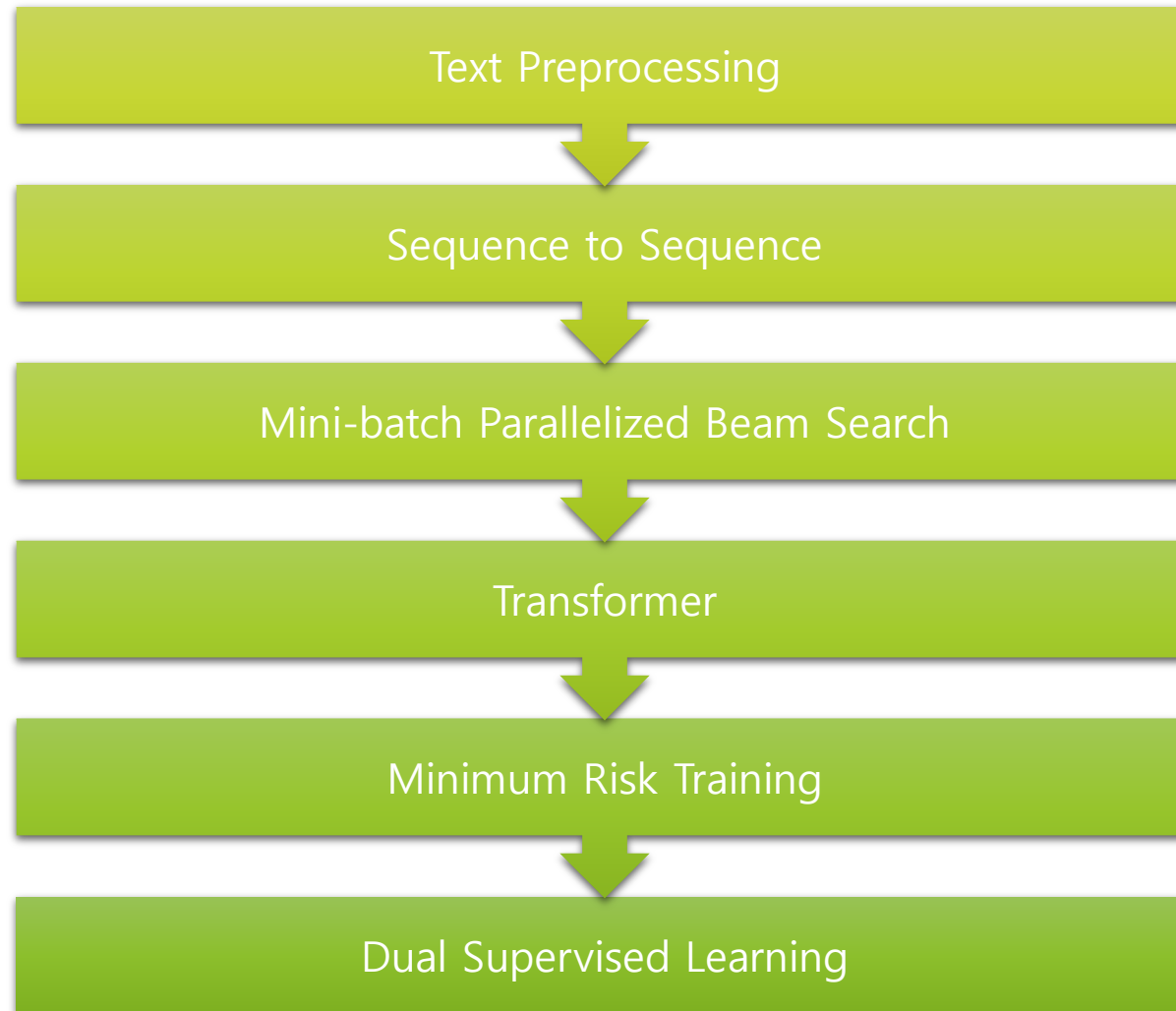
# Data Preparation

Ki Hyun Kim

[nlp.with.deep.learning@gmail.com](mailto:nlp.with.deep.learning@gmail.com)

# 앞으로 우리는...

- Machine Learning Exercise



AI Hub의 다양한 사용자가 AI를 개발 및 활용하기 위한 인프라 서비스 4종 (AI데이터, AI 소프트웨어, AI 컴퓨팅, AI 이지빌더)

AI 활성화를 위한 서비스 4종 (AI 혁신체험, AI 리더보드, AI 커뮤니티, AI 경진대회) 총 8가지 서비스를 제공하고 있습니다.



# 한국어-영어 번역(병렬) 말뭉치 AI데이터

- 한국어-영어 160만 문장의 번역 말뭉치
  - 문어체 한영 번역 110만 문장
    - 뉴스 80만, 정부 웹사이트 콘텐츠 10만, 조례 10만, 한국문화 10만
  - 구어체 한영 번역 50만 문장
    - 구어체 40만, 대화체 10만
- 이외에도 법률, 특허, 일반상식, 한국어대화, 기계독해 등의 말뭉치 제공