

AI Engineering TFT

Roadmap

Cplatform SW사업부



Contents

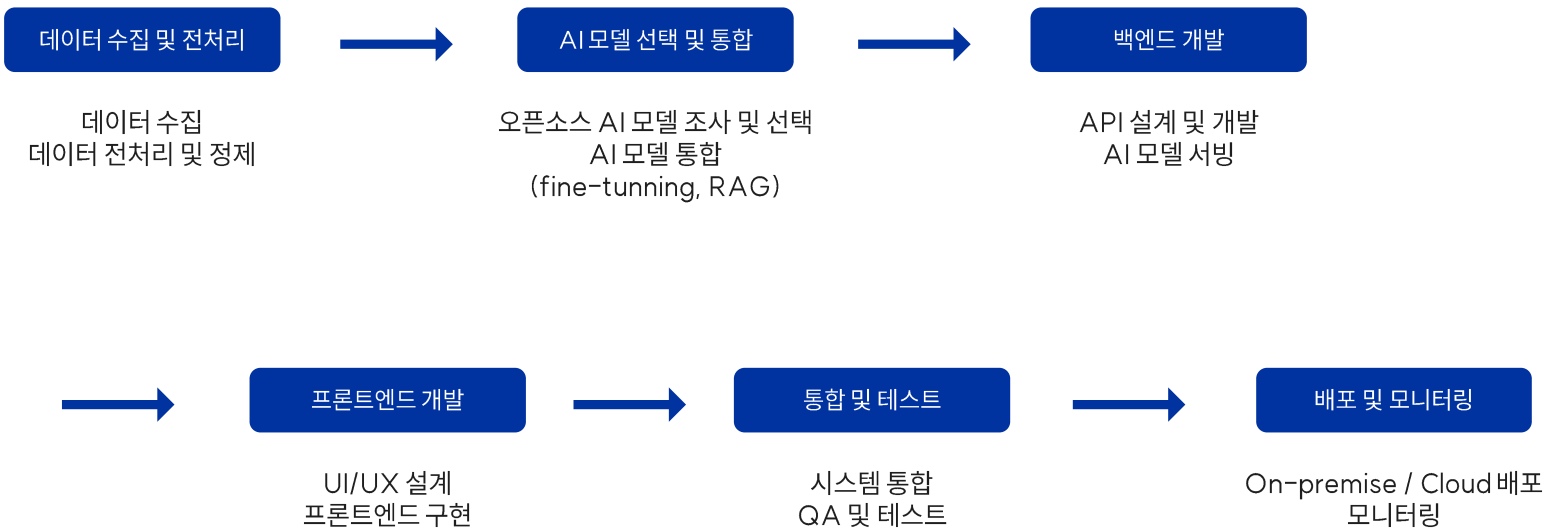
Part 1. Roadmap

Part 2. 기술 설명

Part 3. 사내 교육 ref.

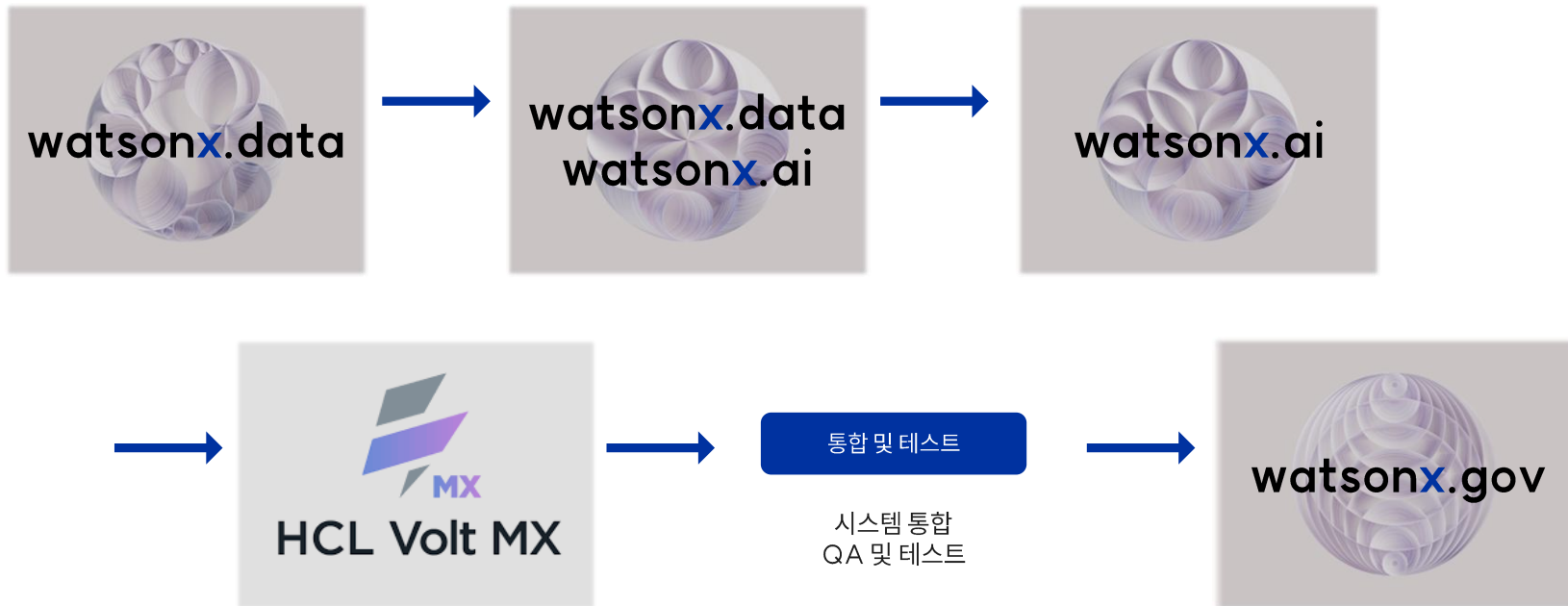
1. Roadmap

At a Glance



1. Roadmap – 플랫폼 대체 방안

At a Glance



2. 기술 설명 : 데이터 수집 및 전처리

데이터 수집

- 웹 스크래핑

기술 스택	설명	난이도
BeautifulSoup	HTML 및 XML 파일에서 데이터를 추출하기 위해 사용되는 파이썬 라이브러리	하
Scrapy	웹 사이트 크롤링을 위한 오픈소스 프레임워크	하
Selenium	웹 브라우저를 이용한 자동화 프레임워크	하

- API 활용

기술 스택	설명	난이도
RESTful API	REST 기반 서비스 API를 통한 데이터 수집	하
GraphQL	클라이언트가 서버로부터 데이터를 요청할 때 쿼리 형식으로 요청	하

- 데이터 저장

기술 스택	설명	난이도
AWS S3	Amazon Simple Storage Service, Object Storage	하
MongoDB 아틀라스	MongoDB의 클라우드 데이터 관리 플랫폼, 클라우드 데이터베이스	중

- 데이터베이스 관리 (SQL, NoSQL)

데이터베이스 관리를 위한 질의 언어, 사내 교육 활용 가능, 난이도 중~하

2. 기술 설명 : 데이터 수집 및 전처리

데이터 전처리 및 정제

- 데이터 클리닝

기술 스택	설명	난이도
pandas	데이터 처리와 분석을 위한 테이블 데이터 형식의 파이썬 라이브러리	하
numpy	Numerical Python, 수치, 과학 연산을 위한 파이썬 라이브러리	하

- 텍스트 전처리

기술 스택	설명	난이도
NLTK	Natural Language Toolkit, 자연어 처리 및 문서 분석용 파이썬 라이브러리	중
SpaCy	자연어 처리를 위한 파이썬 라이브러리	중

- 데이터 정규화 및 표준화

데이터를 공통 척도 내에 위치 시키며, 표준 정규분포표의 속성을 가지도록 조정. 관련 수학 지식 필요, 난이도 중

- 결측치 및 이상치 처리

이상치 (Outlier) : 관측된 데이터의 범위에서 크게 벗어난 값

결측치 (Missing Value) : 데이터 수집 과정에서 측정되지 않거나 누락된 값(Null Value)

이상치와 결측치 처리 방식에 대한 수학적 알고리즘 학습 필요. 난이도 중

2. 기술 설명 : AI 모델 선택 및 통합

오픈 소스 AI 모델 조사 및 선택

- 최신 AI 모델 동향 파악 및 활용 방식

기술 스택	설명	난이도
GPT	Generative Pre-trained Transformer, open ai사의 API를 일정 금액 지불 후 사용 가능	중
open source	Hugging Face에서 구할 수 있는 Granite, Llama 등 오픈 소스로 등록되어 있는 모델 사용	중

- 모델 평가 지표

기술 스택	설명	난이도
BLEU	Bilingual Evaluation Understudy, 기계 번역의 품질을 평가하는데 사용하는 지표	-
ROUGE	Recall-Oriented Understudy for Gisting Evaluation, 문서 요약의 품질을 평가하는데 사용하는 지표	-

2. 기술 설명 : AI 모델 선택 및 통합

AI 모델 통합

- 딥러닝 프레임워크

기술 스택	설명	난이도
PyTorch	facebook 연구팀에서 만든 딥러닝 구현을 위한 파이썬 기반 오픈소스 라이브러리, 간결하고 빠른 구현이 특징, 한국어 튜토리얼 보유 https://tutorials.pytorch.kr/recipes/recipes_index.html	중
TensorFlow	구글에서 개발한 딥러닝 라이브러리 (기본 C++ 구현, python, java, go 등 지원)	상

- 모델 fine-tuning 기법

기술 스택	설명	난이도
SFT	Supervised Fine Tuning, 주어진 작업에 대해서 특화된 성능을 발휘할 수 있도록 하는 fine-tuning 기법 유사한 데이터 세트를 가지고 있거나, 데이터 양이 적을 경우 효과적인 성능을 기대할 수 있음	상
PEFT	Pre-training with Extracted Feature-based Transfer, 사전에 학습된 모델의 특징을 추출하여 다른 작업에 적용하는 방식 기존 모델의 학습을 유지한 채로 활용한다는 점에서 재사용성이 좋다는 특징이 있음	상

- GPU 가속 및 분산 학습 (GPU 사용)

모델의 빠른 학습을 위해 분산 딥러닝 등 GPU의 가속화 분산화를 통한 최적화 방식.

Distributed Data Parallel (DDP), GPU 확보, 관련 라이브러리 지식 추가적인 작업 필요. 난이도 상

- RAG (Retrieval-Augmented Generation)

기술 스택	설명	난이도
임베딩	고차원 공간에서 단어 또는 문장 등의 의미를 반영한 벡터를 생성하는 과정, 임베딩 모델 사용 방식 습득 필요 (코딩 등) 단어 기반 벡터 : Word2Vec, Glove, FastText 등 문장 기반 벡터 : BERT, Sentence-transformer(Sentence BERT) 등	상
벡터 DB	임베딩한 모델을 벡터 DB에 저장. ex) Pinecone, Weviate, Chroma, Qdrant, Milvus	하

2. 기술 설명 : 백엔드 개발

- RESTful API 설계 원칙

URI를 통해 자원을 표시하고, HTTP Method를 이용하여 해당 자원의 행위를 규정하여 그 결과를 받을 수 있도록 설계해야 한다. 난이도 하

- 백엔드 프레임워크

기술 스택	설명	난이도
Django	MVC 패턴을 따르는 고수준의 파이썬 프레임워크, DB 관리를 위한 ORM, 내장된 인증 및 권한, 동적 웹페이지 생성을 위한 템플릿 엔진 등 포함	상
Flask	MVT 패턴을 따르는 간결하고 유연한 파이썬 프레임워크. 소~중규모 프로젝트	중
FastAPI	파이썬 3.7+를 기반으로 한 빠른 웹 프레임워크. 사용하기 쉽고 직관적인 설계	중
Node.js	V8 JavaScript 엔진으로 빌드된 런타임 (JavaScript의 구동 환경)	상

- 인증 및 보안

기술 스택	설명	난이도
JWT	Json Web token, 인터넷 표준 인증 방식	중
OAuth	신뢰할 수 있는 외부 어플리케이션의 open api의 id, pw를 입력하여 인증을 처리해주는 접근 위임을 위한 개방형 표준	중

- 모델 서빙

모델 서빙이란 사용자에게 LLM을 서비스로서 제공하는 것을 의미한다. 모델의 결과물을 입력으로 해서 다음 모델 결과를 생성하는 Autoregressive 모델의 특성으로 인해 어떠한 작업도 진행하지 않은 환경은 사용자의 입장에서 매우 느리게 느껴질 수 있다.

이를 개선하기 위해 KV(key-Value) Cache, CUDA 등의 방식을 사용하기도 하며, 추론 속도를 빠르게 해주는 [tensorrt-llm](#), [vllm](#), [lmdeploy](#), [tgi](#) 등의 라이브러리를 사용하는 방법이 있다.

기술 스택	설명	난이도
llm 모델 직접 서빙	직접 자신의 모델을 서빙 하는 방식	상
서빙 SW 사용	LLM 서빙 오픈소스 소프트웨어들을 통해 서빙 하는 방식	상

- 컨테이너화

효과적인 모델 서빙을 위해 컨테이너화방식 사용

관련 기술 : Docker, Kubernetes, 난이도 상

- 사용자 중심 디자인 원칙

사용자의 경험, 사용을 목적으로 전체 디자인을 개발하는 디자인 원칙
디자인 원칙, 작업 프로세스, 사용성 검증 툴 등 활용 방안 모색 필요, 난이도 중

- 와이어프레임, 목업 제작 도구

기술 스택	설명	난이도
Figma	웹 기반의 UI/UX 디자인 및 프로토타이핑 협업툴, 무료 가능	하
Adobe XD	웹 기반의 UI/UX 디자인 및 프로토타이핑 협업툴, 유료	하

2. 기술 설명 : 프론트엔드 개발

프론트엔드 구현

- 모던 JavaScript

기술 스택	설명	난이도
ES6+	ECMAScript, 자바스크립트의 표준 스펙, ES6+는 2015년에 개정된 ES6를 포함한 상위 스펙을 통칭하는 말	상

- 프론트엔드 프레임워크

기술 스택	설명	난이도
React	메타에서 개발한 오픈 소스 자바스크립트 라이브러리	상
Vue.js	자바스크립트 프레임워크, 재사용이 가능한 컴포넌트 활용	상
파이썬 프레임워크	파이썬 기반의 웹 프론트엔드 프레임워크 활용. ex) Anvil, Pynecone, JustPy	중

- 상태관리

기술 스택	설명	난이도
Redux	state(상태)를 공유할 수 있도록 하는 라이브러리	상
Vuex	store(저장소)를 중심으로 모든 컴포넌트가 공유할 수 있는 vue.js의 상태 관리 라이브러리	상

2. 기술 설명 : 통합 및 테스트

- CI / CD 파이프라인 구축

Continuous Integration / Continuous Deployment, 작업한 코드의 변경 사항을 지속적으로 통합하고 실제 환경에 지속적으로 배포하는 프로세스

기술 스택	설명	난이도
Jenkins	CI/CD를 위한 오픈 소스 도구. 소프트웨어 개발 생명주기를 자동화한다. 코드 변경 사항을 감지하여 빌드, 테스트, 배포를 자동 수행한다.	중

- 버전 관리

기술 스택	설명	난이도
Git	버전 관리, 형상 관리 시스템. 소프트웨어. 분산 버전 관리 시스템	하

- 마이크로서비스 아키텍처 (MSA)

높은 응집도와 낮은 결합력을 통해 어플리케이션을 최소 단위 서비스만 수정하여 확장 및 재배포 가능하게 하는 아키텍처
복잡한 아키텍처와 기술, 비용들로 인해 서비스에 적용 필요성에 대해 고민해 보아야 함. 난이도 상

2. 기술 설명 : 통합 및 테스트

QA 및 테스트

- 테스트 방식

기술 스택	설명	난이도
단위 테스트	하나의 모듈을 기준으로 독립적으로 의도와 맞게 동작하는지 검증하는 절차	-
통합 테스트	두 개 이상의 단위 모듈이 함께 잘 동작하는지 확인하는 절차	-
E2E 테스트	어플리케이션의 흐름이 처음부터 끝까지 종합적으로 테스트하는 것	-
A/B 테스트	실사용자를 대상으로 특정 콘텐츠를 A, B 안으로 나누어 더 많은 관심을 가지는 것을 확인하는 종합 대조 실험	-

- 테스트 자동화 도구

기술 스택	설명	난이도
Selenium	웹 브라우저의 자동화 테스트를 위한 오픈 소스 도구, 크로스 브라우저 테스트에 강하다.	중
Cypress	모던 웹 어플리케이션을 위한 E2E 테스트 자동화 도구, JS에 특화	중
Junit	자바 기반 어플리케이션의 유닛 테스트 프레임워크, 테스트 주도 개발 지원	중

- 성능 테스트 도구

기술 스택	설명	난이도
Lighthouse	Chrome에서 제공하는 웹 어플리케이션의 품질 개선 자동화 도구	중
Locust	Python으로 작성된 부하테스트 도구. 설치와 사용이 간편하고 파이썬 스크립트로 작성하여 부하테스트를 용이하게 가능하다.	하

2. 기술 설명 : 배포 및 모니터링

- 배포 방식

기술 스택	설명	난이도
Deployment	어플리케이션의 성격, 특징 등을 통해 적합한 배포 환경을 선택해야 한다. On-Premise, Cloud, Hybrid 등	중
IaC	Infra as a Code, 수동 프로세스 및 설정 대신 코드를 사용한 컴퓨팅 인프라를 프로비저닝 하는 것. 대표적으로 Terraform이 있다.	상

- 모니터링

기술 스택	설명	난이도
Prometheus	시스템 및 서비스의 상태를 모니터링 하는 오픈소스 모니터링 도구	중
Grafana	시계열 메트릭 데이터를 시각화 하는 오픈소스 도구	중

3. 사내 교육 ref.

사내 교육 (인프런) 소개

관련 기술	과정명	교육비용	교육시간
웹스크래핑	파이썬입문과 크롤링기초 부트캠프 [파이썬, 웹, 데이터 이해 기본까지]	69,300	18H
데이터베이스 관리	MongoDB - Javascript로 배우는 NoSQL DB 처음하는 MongoDB(몽고DB) 와 NoSQL (빅데이터) 데이터베이스 부트캠프 [입문부터 활용까지]	33,000 69,300	6H 10H
데이터 클리닝	처음하는 파이썬 데이터 분석(쉽게! 전처리, pandas, 시각화까지 전과정 기본 익히기)	77,000	12H
텍스트 전처리	[PyTorch] 쉽고 빠르게 배우는 NLP	55,000	6H
AI 모델 활용	실리콘밸리 엔지니어와 함께하는 OpenAI API (ChatGPT)	44,000	6H
딥러닝 프레임워크	[개정판] 파이썬 머신러닝 완벽 가이드 파이썬을 활용한 머신러닝 딥러닝 입문	99,000 66,000	38H 22H
fine-tuning	LLM 101: 2시간에 끝내는 이론-to-실습 코스! 만만의 Llama 채팅데모 프로젝트!	94,600	2H
RAG	RAG 마스터: 기초부터 고급기법까지 (feat. LangChain)	110,000	9H
컨테이너화	Docker를 이용한 MSA 애플리케이션 만들기	49,500	2H
CI/CD	Jenkins를 이용한 CI/CD Pipeline 구축	88,000	16H
버전 관리	지옥에서 온 Git	-	6H
IaC	처음 시작하는 Infrastructure as Code: AWS & 테라폼	38,500	5H
모니터링	실무에서 꼭 필요한 서버 모니터링(Zabbix)	27,500	2H

: 사내 교육에 포함되어 있지 않음.

Thank you

