



연합학습기반 신약개발 가속화 프로젝트

(3 세부) 연합학습 플랫폼 활용 활성화 지원

연합학습 기반 ADMET 예측을 위한 언어-그래프 앙상블 모델 개발

주관기관 : 목암생명과학연구소 (신현진)
공동기관 : 서울대학교 (장문영)

|연구 역량| 연구개발기관 및 연구책임자 소개

신약개발 전문지식을 바탕으로 다양한 인공지능 연구 수행 경험 보유

[주관기관] 목암생명과학연구소



- 목암생명과학연구소는 인공지능을 통한 신약개발을 연구하는 민간 비영리재단법인
- 국내·외 바이오 및 인공지능 분야 인력 구성 및 인공지능 신약개발 관련 내/외부 과제 다수 수행 중
- 서울대병원, KAIST 등 국내 우수 연구기관과 긴밀한 협력 추진을 통해 자체 연구 파이프라인 확대 중

[공동기관] 서울대학교 공과대학 인공지능 연구실



- 서울대학교 공과대학 인공지능 연구실 (Data Science and AI Lab)
- 바이오 메디컬 응용 분야 최상위 저널에 다수의 논문 발표
- NeurIPS, ICLR, ICML, CVPR 등 세계 최정상급 학술 대회에 지속적으로 우수한 성과 발표

연구책임자 : 신현진 소장

- 글로벌 제약사 등 우수 연구기관 인공지능 연구
 - 15 years in

참여연구자

- 자담바 에르햄바야르 책임연구원
- 이현수 연구원

연구책임자 : 장문영 BK부교수

- 다년간 대학 병원에서 환자 진료 및 연구
 - 14 years in 서울대학교병원 중앙대학교병원

참여연구자

- 이종현 박사과정
- 이동혁 박사과정

연구 수행 분야 및 역량

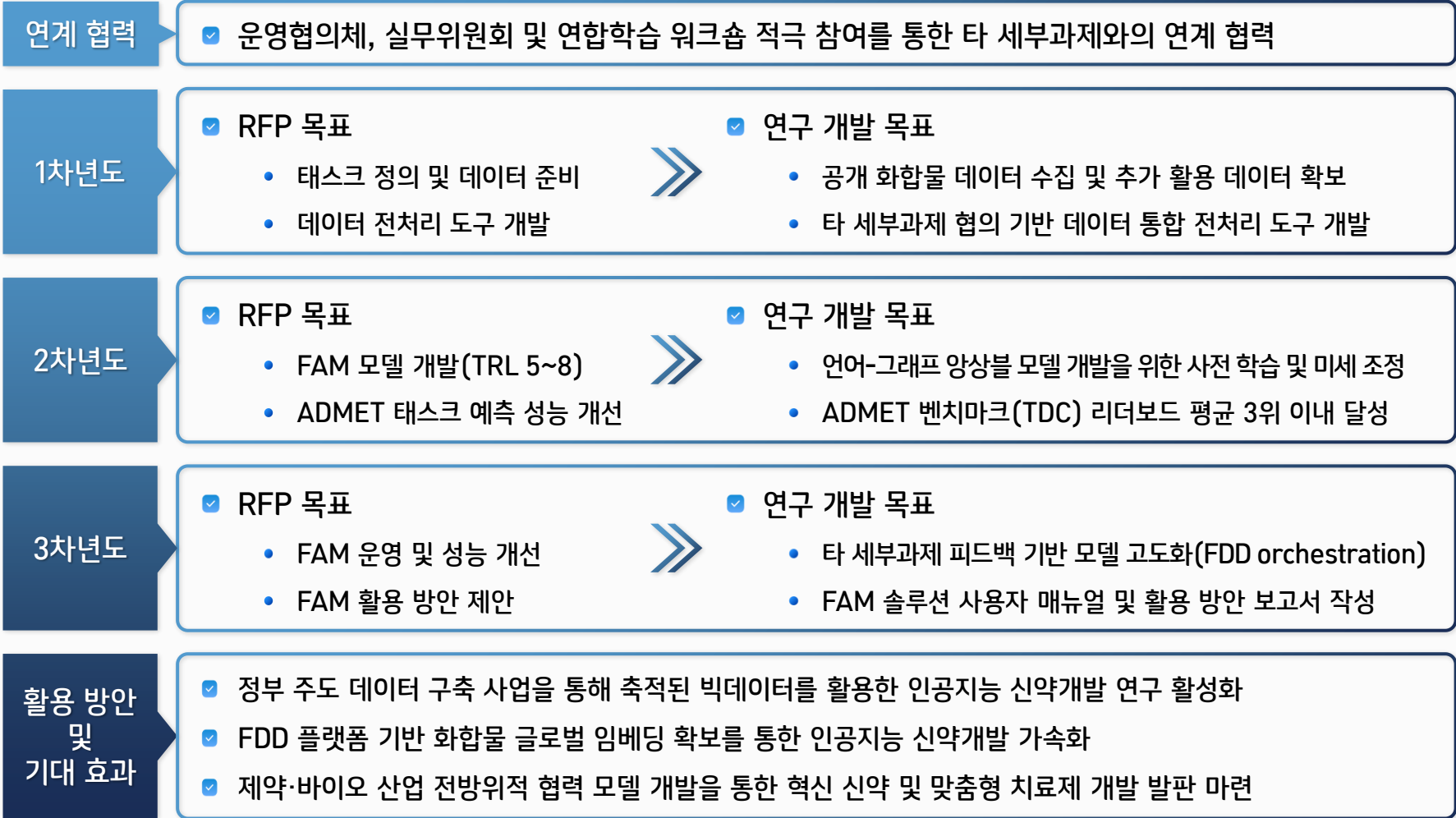
- 현재 LLM을 적용한 신약개발 내부 연구 수행 중
 - Lipid nanoparticle(LNP) 특성 예측 및 생성을 위한 인공지능 연구
 - mRNA codon 최적화를 위한 인공지능 연구

연구 수행 분야 및 역량

- 현재 GFM을 적용한 신약개발 내부 연구 수행 중
 - Biomedical knowledge graph를 이용한 질환 관련 유전자/단백질 예측을 위한 인공지능 연구

|연구 목표 및 전략| 연구 개발 단계별 목표

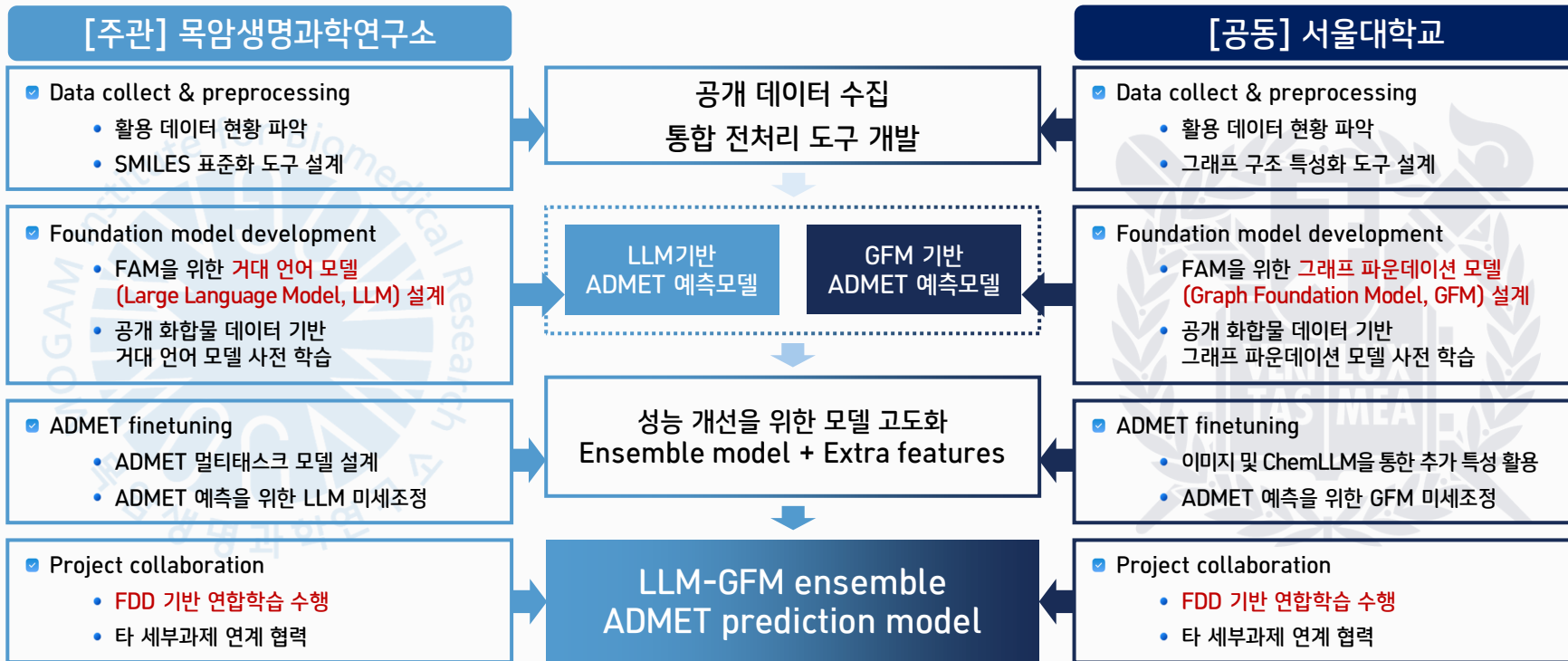
연합학습 기반 ADMET 예측을 위한 언어-그래프 앙상블 모델 개발



|연구 목표 및 전략| 연구 추진 전략 개요

연합학습 협의체 기반 병렬 연구 수행

운영협의체 - 실무위원회 - 연합학습 워크숍

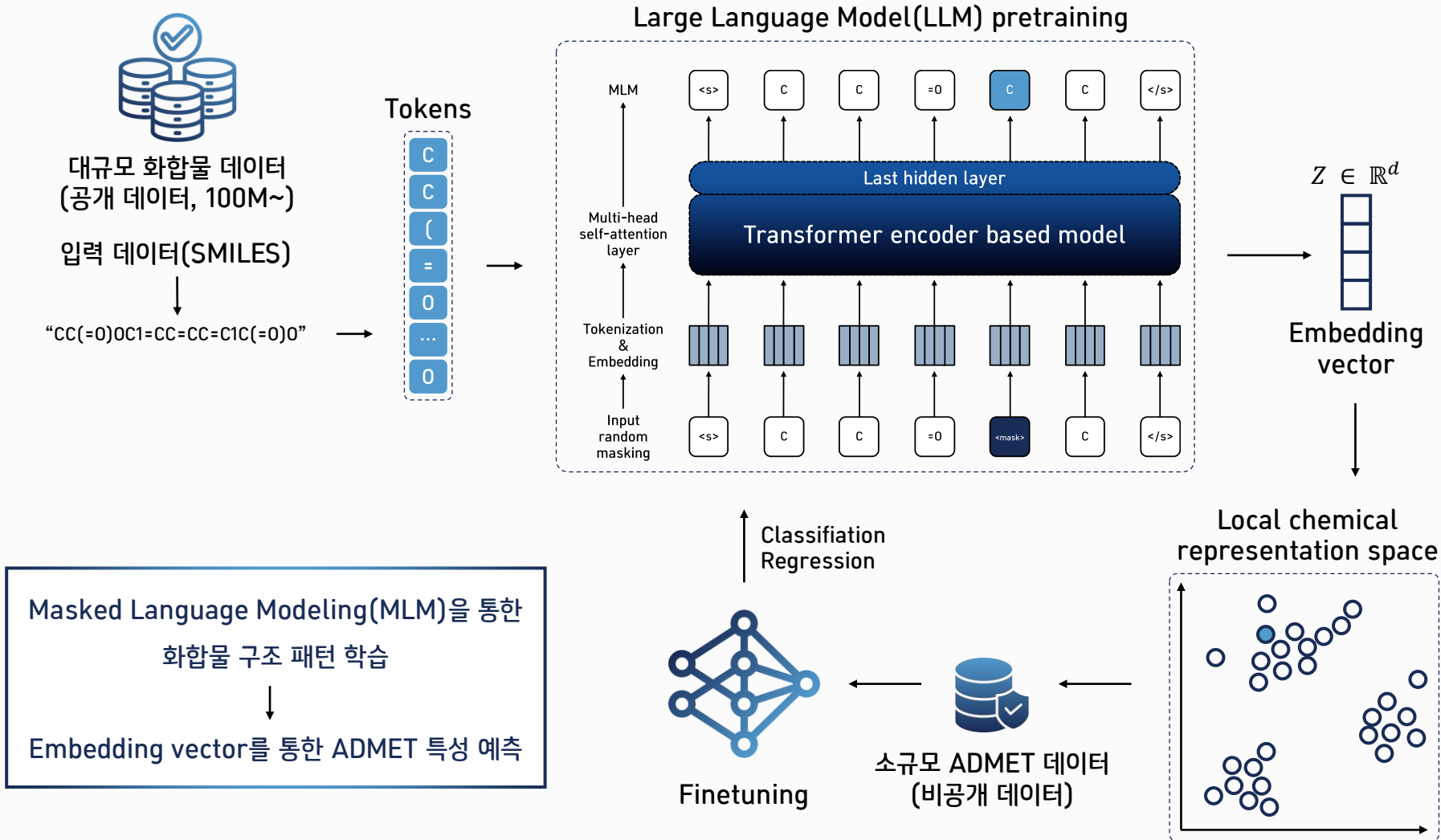


주요 마일스톤

- ✓ (1차년도) LLM/GFM 학습을 위한 공개 화합물 데이터 수집 및 전처리
- ✓ (2차년도) LLM/GFM 사전 학습 및 미세 조정을 통한 FAM 확보
- ✓ (3차년도) FDD 플랫폼 내에서 ADMET 태스크별 연합학습 수행 및 모델 고도화

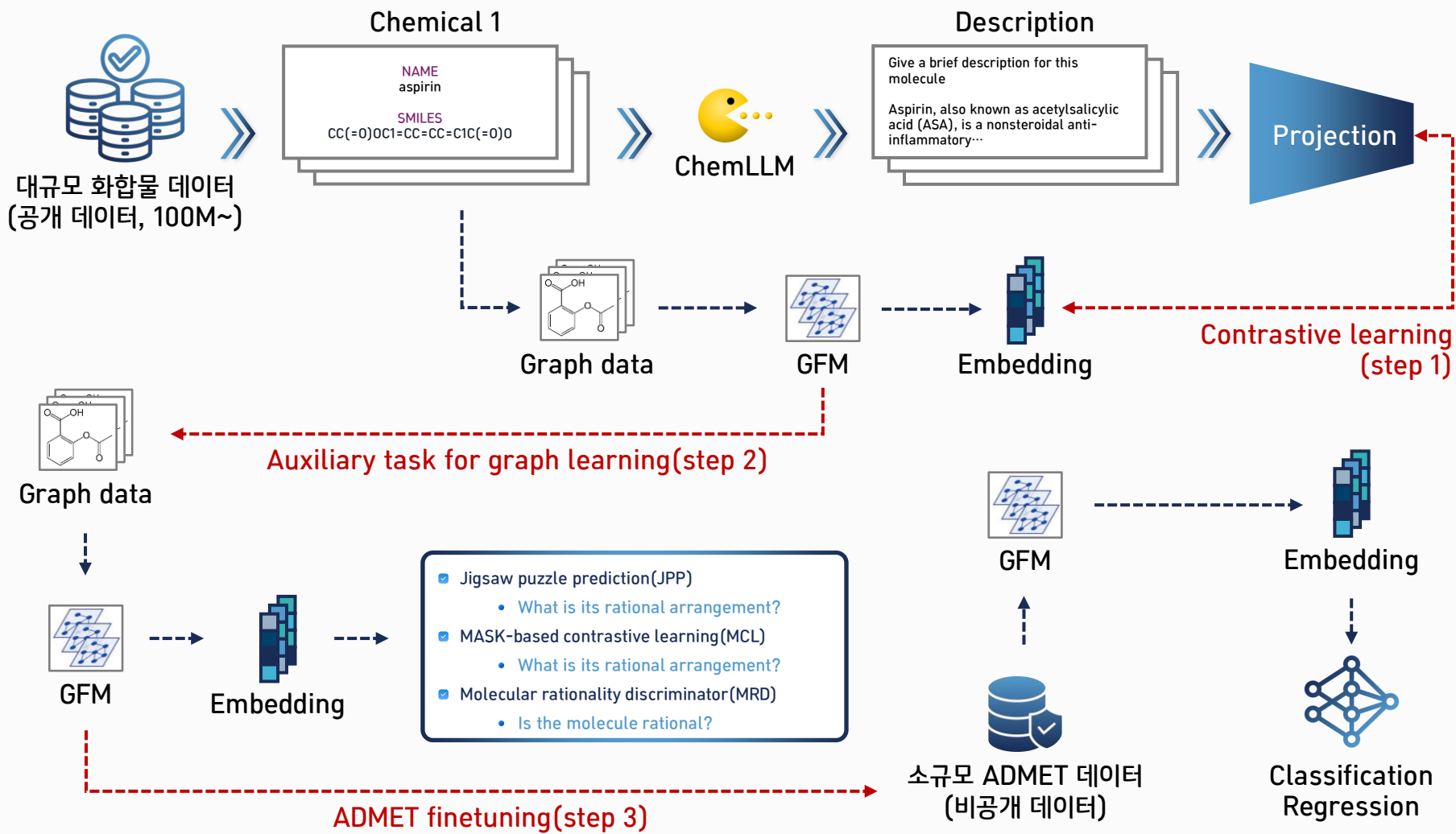
|핵심 연구 내용 및 기술| LLM/GFM 사전 학습 및 미세 조정을 통한 FAM 확보

거대 언어 모델(Large Language Model, LLM) 학습 전략



|핵심 연구 내용 및 기술| LLM/GFM 사전 학습 및 미세 조정을 통한 FAM 확보

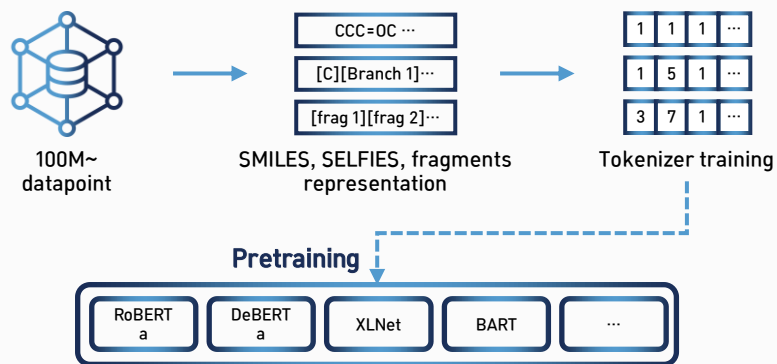
그래프 파운데이션 모델(Graph Foundation Model, GFM) 학습 전략



|핵심 연구 내용 및 기술| LLM/GFM 사전 학습 및 미세 조정을 통한 FAM 확보

LLM-GFM 앙상블 전략 수립

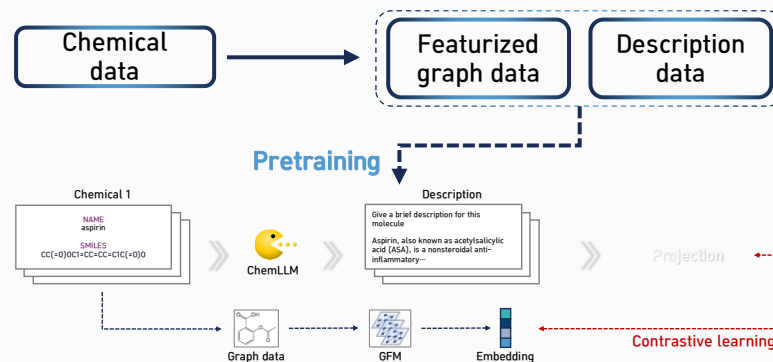
Large Language Model(LLM)



Benchmark task finetuning

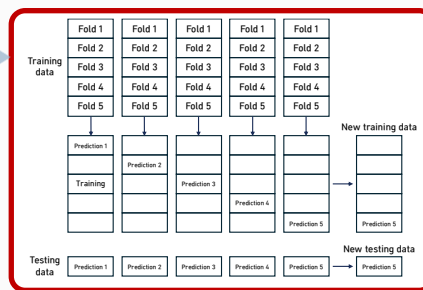
Finetuned LLMs

Graph Foundation Model(GFM)



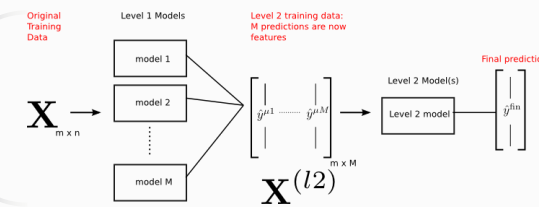
Auxiliary task finetuning

Finetuned GFMs



Stacking ensemble

Final predictions for ADMET




| 연구 산출물 | 연구성과지표

ADMET 태스크 예측 평가 지표 제시 및 성능 향상 지향


성과지표명		단계	목표치 (24-26)	가중치 (%)
전담기관 등록·기탁 자료	학술지, 제품인증	SCI(E) 학술지 등록 건 수	2	10
		기술 및 제품 인증 등록 건 수	2	5
	특허	특허 출원 건 수 : 국내 출원	2	5
		특허 등록 건 수 : 국내 등록	2	5
	보고서, 소프트웨어	보고서 원문 등록 건 수	1	10
		소프트웨어 등록 건 수	2	10
	고용 창출	고용 창출 인원	1	5
	연구시설장비 등록 건수	연구시설장비 등록 건 수	2	5
	시제품 제작	시험제품제작 : TRL 5단계	2	5
		시험제품제작 : TRL 6단계	3	5
		시험제품제작 : TRL 7단계	1	10
		시험제품제작 : TRL 8단계	1	10
연구개발 과제 특성 반영 지표	특성 반영 지표	ADMET 태스크 예측 성능 향상	-	10
		모델 사용자 매뉴얼 개발 (1건 이상)	0	2
		FDD 플랫폼 및 FAM 활용방안 보고서 작성(1건 이상)	0	3
계			21	100

리더보드 평균 3위 이내 달성



TDC ADMET benchmark group

Caco2_Wang	Bioavailability_Ma
Lipophilicity_AstraZeneca	Solubility_AqSolDB
HIA_Hou	Pgp_Broccatelli
BBB_Martins	PPBR_AZ
VDss_Lombardo	CYP2C9_Veith
CYP2D6_Veith	CYP3A4_Veith
CYP2C9_Substrate_CarbonMangels	CYP2D6_Substrate_CarbonMangels
CYP3A4_Substrate_CarbonMangels	Half_Life_Obach
Clearance_Hepatocyte_AZ	Clearance_Microsome_AZ
LD50_Zhu	hERG



2 세부:
신약개발 데이터 활용 및 품질 관리

테스트 데이터세트 구축 및 검증을 통해 실전 효용성 입증