금융권 생성형 AI
활용 지원방안

2024. 12.

금 융 위 원 회

목 차

1. 추진배경

□ 전세계적으로 AI가 산업 전반의 혁신을 이끄는 **핵심 동력**으로 부상하면서 **금융권**에서도 AI 활용에 대한 **수요**가 증가

○ 특히, 주로 분류·예측에 활용되는 전통적인 AI와 달리 **새로운 컨텐츠**를 생성할 수 있는 **생성형 AI의 등장**으로 관심 증대*

* 미국, 영국, 싱가포르 등 9개국 956명 금융회사 임직원 대상 설문조사 결과, 생성형 AI를 이미 적용하였거나 적용을 고려 중인 응답자가 전체의 83%(Finastra, 2023)

□ 그러나, **국내 금융회사**들은 ❶**형식적·구체적 보안규제,** ❷양질의 **AI 학습데이터** 부족, ❸**불확실성** 위험 등으로 AI 활용에 한계

❶ **(보안환경)** GPT-4 등 AI서비스는 주로 **인터넷망**을 통해 **해외 서버**에서 제공되어, **망분리*** 등 보안환경 하에서 활용이 어려움

* 금융회사의 **내부 업무용시스템**은 **외부통신망과 분리·차단**(전금規 §15①iii)

❷ **(데이터 부족)** AI 학습데이터가 대부분 **일반분야, 영미권 언어**로 되어있어, **금융 분야**에 특화된 **한글 데이터** 부족

❸ **(불확실성)** 거짓정보 생성(hallucination), 편향(bias), 정보 유출 등 AI 위험에 대응하기 위한 **AI 거버넌스 및 가이드라인** 불명확

현장의 의견

- ○○증권 : 내부 업무망에서 **외부 인터넷망**의 **AI 개발**을 위한 **웹페이지**(Hugging face, Github 등) **접속**이 **불가능**해 AI모델·어플리케이션 설치 및 활용 곤란
- □□은행 : 최신 오픈소스 AI모델을 **테스트**할 수 있는 PoC(아이디어 검증) **환경** 필요
- △△생명보험 : 국내 **금융용어, 약어 등**에 대한 AI학습용 **한글 특화 데이터**가 필요, AI의 **성능평가·검증**을 위한 **데이터셋** 부족, 회사 내부 데이터 활용에 한계
- ◇◇은행 : AI 이용에 대한 표준화된 가이드가 없어 적극적인 활용이 어려우므로 금융 공통의 **AI 거버넌스, 가이드라인** 필요

⇒ **'24.3월 「금융권 AI 협의회」를 구성하여 8차례 실무분과 회의를 통해 금융회사, 유관기관, 업권별 협회 등 의견수렴 및 지원 방안 논의**

2. 추진 내용

〈 추 진 방 향 〉

◇ 금융권 생성형 AI 이원(Two-track) 활용 체계 마련을 위해 ❶금융권 AI 플랫폼 구축, ❷금융분야 특화 데이터 지원, ❸금융분야 AI 가이드라인 개정 추진

	세부과제		내용	비고
1. 금융권 AI 플랫폼 구축	가. 오픈소스 AI 모델, 데이터 등 선별 제공	⇒	▸ 전문가 그룹이 선별한 오픈소스AI 모델, 데이터 등 제공	'25년 상반기
	나. 테스트(PoC) 환경 지원	⇒	▸ 오픈소스 AI 통합 개발 환경 지원	
	다. 내부망 설치 인프라 구축	⇒	▸ 데이터 허브를 통해 AI 모델, 데이터 등 내부망 설치 지원	
2. 금융분야 특화 데이터 지원	가. 금융권 특화 한글 말뭉치	⇒	▸ AI의 금융 전문성 향상을 위한 RAG용, 추가학습용 데이터 ▸ AI 윤리·성능 평가지원용 데이터	'25년 1분기 이후 단계적 지원
	나. 공익 목적 데이터	⇒	▸ 금융사기 방지, 신용평가, 금융보안 데이터 제공 채널 일원화	'25년 상반기
3. 금융분야 AI 가이드라인 개정	가. 금융 AI 원칙 제시	⇒	▸ 거버넌스, AI 개발·활용 단계별 원칙 제시	'25년 1분기
	나. 안내서 개정	⇒	▸ 금융 AI 원칙의 적용을 위한 구체적인 설명 제시	'25년 상반기

가. 금융권 AI 플랫폼 구축

□ **(현황)** 생성형 AI는 **인터넷망**에서 제공되는 ❶**상용 AI**와 회사 **내부 시스템**에 설치하는 ❷**오픈소스 AI**로 구분

< 금융권 AI 이원(Two-track) 활용 체계 >

○ 국내 금융회사들은 **AI 활용 목적**, **비용 효율성** 등을 고려하여 두 가지 **모두**를 **전략적으로 사용**하는 방안을 고민 중

* 글로벌 데이터 및 AI 기업 SAS는 2025년에는 대규모 언어모델(LLM)과 소규모 언어모델(SLM)을 전략적으로 선택하는 분산형 AI 환경이 촉진될 것으로 전망

○ **인터넷** 환경에서 제공되는 **상용AI 활용**은 **망분리**에 대한 **규제 샌드박스**를 허용하여 점차 활성화될 것으로 기대되나,

- **오픈소스 AI**의 경우 외부망에서 **내부망**으로 바로 설치하기 어렵고 **수많은 모델이 난립***하여 **성능 및 안전성 검증**에 한계

* 세계 최대 AI 플랫폼 '허깅페이스'에 등록된 오픈소스 AI 모델은 약 110만 개

⇒ ❶상용 AI는 망분리 규제 샌드박스를 통해 폭넓게 허용하는 한편, ❷오픈소스 AI는 금융권 AI 플랫폼을 통해 **내부망에 손쉽게 설치·활용**할 수 있도록 **금융권 AI 이원(Two-track) 활용 체계** 구축

□ **(추진내용) 「금융권 AI 플랫폼」**을 통해 AI 설치·활용 통합지원

❶ **(오픈소스 AI모델 선별)** 금융분야에 적합한 **성능과 안전성**을 지닌 오픈소스 **AI모델, 데이터**를 전문가 그룹*이 **선별**하여 제공

* (구성) 은행·보험·증권 업권별 AI 전문가, 금융 유관기관(신정원, 금보원, 금결원) 등

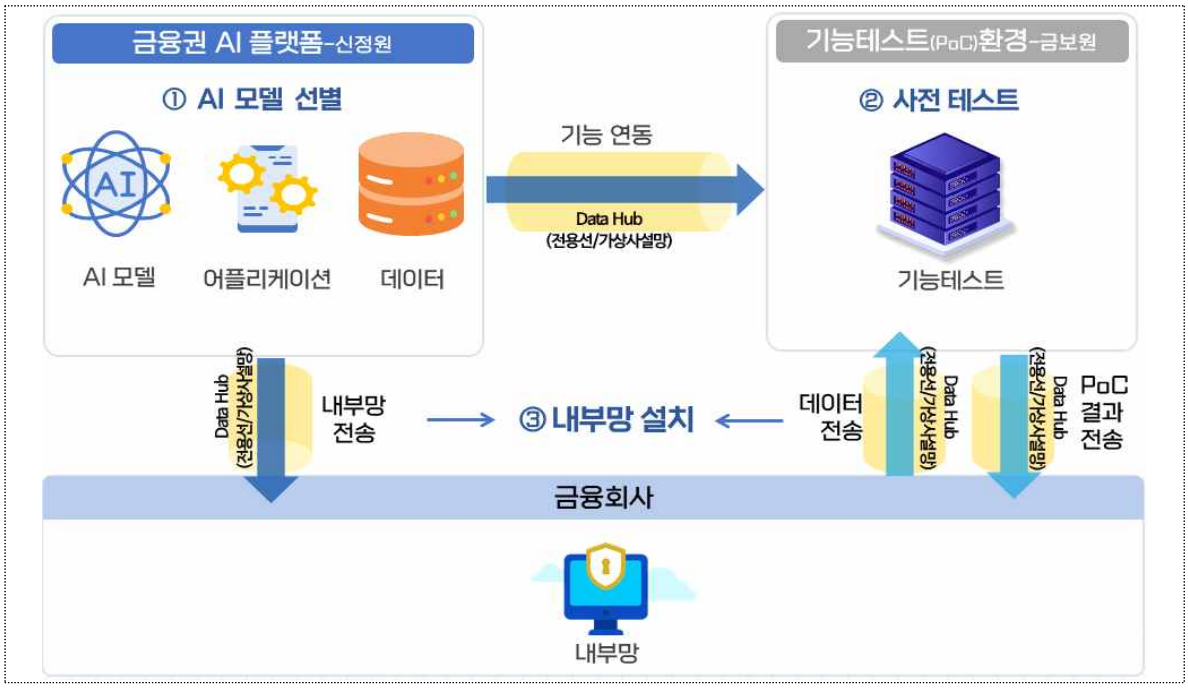
❷ **(사전 테스트)** 금융회사별 **혁신적인 생성형 AI 서비스**에 적합한 AI모델, 데이터를 **실험·선별**하기 위한 **기능테스트(PoC*) 환경** 제공

* PoC(Proof of Concept) : 실제 서비스 개발을 시작하기 전, 서비스·기술 등의 실현 가능성을 입증하는 절차

❸ **(내부망 설치)** 전문가 그룹이 선별한 오픈소스 **AI모델, 데이터**를 금융사 **내부망**에 바로 설치할 수 있도록 지원하는 인프라* 구축

* 보안성이 확보된 데이터 허브망(금융보안원 제공)을 통해 금융회사 내부망과 AI 플랫폼을 연결하여 안전하고 간편하게 데이터 송·수신 가능

< 금융권 AI 플랫폼 구축 개요도 >



□ **(기대효과)** 금융회사가 **오픈소스 AI**를 **내부망에 설치**하여 **망분리·해외서버 이용금지** 등 규제 **제한 없이** 생성형 AI **활용** 가능

○ 금융권 **공동 인프라** 구축으로, 개별 추진시 데이터 확보 및 AI 테스트 환경 구축 등에 소요되는 **비용을 절감**하여 AI 활용을 촉진

나. 금융분야 특화 데이터 지원

□ **(현황)** Llama3(Meta社) 등 주요 오픈소스 AI 모델은 **영미권 언어, 일반 데이터**를 학습하여 **한국어 능력, 금융분야 전문성** 부족

○ AI 성능 향상을 위해 **금융분야 한글 빅데이터**가 필요하나, 개별 회사가 금융 관련 양질의 빅데이터를 구축하는데 한계

* 특히, 편향(bias), 환각(hallucination) 등 생성형 AI의 부작용을 최소화하기 위한 신뢰성 있는 **AI 성능·윤리 평가 데이터**도 부재

○ **이상거래 탐지(FDS), 신용평가, 금융보안** 등 공익 목적의 AI 데이터 활용을 위한 **주기적인 공급 체계**도 부재

□ **(추진내용)** ❶**금융권 특화 한글 말뭉치** 및 ❷**공익 목적 데이터** 제공

❶ **(금융권 특화 한글 말뭉치)** 금융 법규, **업권별 보도자료, 연수기관 교육자료** 등을 기반으로 금융권 공동 활용 빅데이터 확보

⇒ 생성형 AI 개발과 활용 방식에 따라 필요한 **다양한 형태**로 제공

RAG*용	생성형 AI가 답변의 정확도 향상을 위해 검색·참조하는 데이터
추가학습용	금융특화 모델 개발을 위한 AI 추가학습(Fine tuning)에 활용
평가지원용	AI모델의 성능(정확성) 및 윤리(공정성) 평가에 활용

* RAG(Retrieval-Augmented Generation) : AI 모델이 외부 특화 데이터, 최신 정보 등을 검색한 후 결과물을 생성하도록 하여 정확도와 신뢰도를 향상시키는 방법

❷ **(공익 목적 데이터) 금융사기방지(금융결제원), 신용평가(신용정보원), 금융보안(금융보안원) 데이터** 등 **공익 목적 데이터**의 **제공 채널**을 「금융권 AI 플랫폼」으로 **일원화**

□ **(기대효과)** 오픈소스 AI 모델의 금융분야 전문성 향상, 성능 및 윤리 평가지원 데이터를 활용하여 AI 서비스의 신뢰도 제고

○ **이상거래 탐지(FDS)** 등 **특성화 AI** 개발을 위한 **데이터**를 「금융권 AI 플랫폼」을 통해 **정기적으로** 편리하게 **활용** 가능

다. 금융분야 AI 가이드라인 개정

□ **(현황)** 그동안 금융당국과 금보원은 금융분야 AI 활용을 활성화하기 위하여 3개의 가이드라인, 안내서 등을 제시*

* 「금융분야 AI 운영 가이드라인」('21.7월), 「금융분야 AI 개발·활용 안내서」('22.8월), 「금융분야 AI 보안 가이드라인」('23.4월)

○ **생성형 AI** 출현 등 급격한 **기술 발전**과 금융권 **내부통제** 강화 등 **제도 변화**에 따라 가이드라인도 개정이 필요

□ **(추진내용)** 생성형 AI기술과 내부통제 등 금융권 제도 변화 등을 반영하여 **「금융분야 AI 가이드라인」**으로 **단일화**

○ **(금융 AI 원칙) 최고경영자 및 경영진**의 역할과 책임, **AI의 보조 수단성**, **금융 안정성**에 대한 **위험의 최소화** 등을 반영

< 금융 AI 7대 원칙(안) >

분야	7대 원칙
거버넌스	① **최고경영자를 포함한 경영진**은 AI 개발·활용에 대한 **관심**을 갖고 **역할과 책임**을 분담해야 함
	② AI 활용 전단계에서 금융·AI 등 **관련 법규**를 **준수**해야 함
	③ 현 단계에서 AI는 업무의 **보조 수단**이므로 **최종 의사결정**과 그에 따른 책임은 **임직원**이 수행함
AI 개발 단계	④ AI 개발 과정에서 **신뢰할 수 있는 데이터와 모델**을 사용해야 함
	⑤ AI 설계·학습 등 전과정에서 **금융 안정성** 위험을 최소화해야 함
AI 활용 단계	⑥ AI 활용 시 **금융소비자의 이익**을 최우선으로 해야 함
	⑦ AI활용 시 **보안성** 기준 및 점검·개선 체계를 마련해야 함

○ **(안내서)** 금융 AI 원칙의 적용, **생성형 AI** 관련 보안 위협과 **보안성 검증 기준** 등을 포함한 상세 설명과 사례를 제시할 예정

□ **(기대효과)** 금융회사들이 실제 업무에 AI를 활용할 때 **불확실성**을 **상당 부분 해소**하고 금융 분야 AI의 **신뢰도 제고**

3. 향후 추진계획

◇ 금융회사들의 신속한 지원 요청에 따라 과제들을 속도감 있게 추진

1 금융권 AI 플랫폼 : '25.상반기 구축

○ 오픈소스 AI모델, 데이터 등 선별을 위한 전문가집단 구성 및 추진체계 마련('25.1분기)

○ AI 플랫폼 시스템 및 PoC 환경 구축('25.상반기)

2 금융권 특화 데이터 : '25.1분기부터 단계적 지원

○ 금융 일반·용어 등 공개 데이터, 금융권 연수자료 기반의 「말뭉치 구축·제공 시범사업」('25.1분기)

○ 금융권(은행·금투·보험·여전 등) 보도자료, 금융권 연수기관 교육자료 기반 말뭉치 및 평가지원 데이터 제공('25.상반기)

○ 업권별 특화 데이터 제공 확대('25.하반기)

3 금융분야 AI 가이드라인 개정 : '25.상반기 완료

○ 금융AI 원칙에 대한 의견수렴('25.1분기)

○ 가이드라인 세부 내용 마련('25.상반기)

과제	내용	일정
1 금융권 AI 플랫폼	전문가집단 구성 및 추진체계 마련	'25.1분기
	금융권 AI 플랫폼 구축	'25.상반기
2 금융권 특화 데이터	공개데이터·연수자료 기반 말뭉치 시범사업	'25.1분기
	보도자료·연수자료 기반 말뭉치 제공	'25.상반기
	업권별 특화 데이터 제공 확대	'25.하반기
3 금융분야 AI 가이드라인 개정	금융 AI 원칙에 대한 의견수렴	'25.1분기
	가이드라인 세부 내용 마련	'25.상반기

참고1 인공지능 관련 용어 설명

□ **(인공지능의 개념)** 컴퓨터 등 기계를 활용하여 **인간의 인지, 학습, 추론 능력** 등을 **인공적**으로 **구현**하는 기술

○ **(분석형 AI) 머신러닝*** **등**을 활용하여 데이터를 학습한 후 **향후 발생**할 수 있는 상황을 **분류·판단**하는데 중점을 둔 AI 유형

* 사람이 직접 프로그래밍하지 않아도 컴퓨터가 데이터를 통해 스스로 학습하고, 그 학습을 기반으로 새로운 작업이나 문제를 해결하는 기술

○ **(생성형 AI) 텍스트, 이미지, 미디어** 등 **기존 데이터**를 학습한 후 **새로운 콘텐츠**를 **생성·구축**하는데 중점을 둔 AI 유형

→ **가장 널리 쓰이고 있는 생성형 AI 기술**은 **'거대언어모델'**(LLM, Large Language Model)이며 **상용 LLM**과 **오픈소스 LLM**으로 분류

□ **(거대언어모델)** **인간의 언어**를 **이해·학습**하며 **생성**할 수 있도록 대규모의 **텍스트 데이터**로 **훈련**된 생성형 AI 모델

○ **(상용 모델) AI 전문기업(OpenAI, Google 등)**에서 **최신 기술**과 **대규모 데이터 셋**을 **활용**하여 **상업적으로 개발**한 모델

- 대형 모델로 **기본 성능이 우수**하나, 사용자 필요에 맞는 모델 **수정이 어렵고** 서비스 활용 시 **입력 정보가 제공기업**으로 **전송**

* 예) GPT-4(오픈AI社), HyperClova(네이버社), Claude(앤트로픽社), Gemini(구글社)

○ **(오픈소스 모델)** **누구나 다운로드 받아 사용**할 수 있는 **공개** 모델로 **금융회사 내부망에 설치·활용**할 수 있으며, 상대적으로 **작은 컴퓨팅 자원**으로도 **활용**이 **가능**한 모델

- 상용 모델 대비 모델의 크기가 상대적으로 작으나, **수정이 용이**하여 **최적화**에 **유리**하고 **외부로 정보 전송 없이 활용** 가능

* 예) LLama(메타社), Grok(xAI社), Gemma(구글社)

참고2 기존 금융분야 AI 가이드라인 등 주요 내용

① 「금융분야 AI 가이드라인」 (금융위원회, '21.7월)

□ **(개요)** 모든 금융권에 적용되는 가이드라인으로서, AI에 대한 신뢰성·공정성 확보, 위험관리, 소비자 권리보장 등의 내용을 포함

- ㅇ **(거버넌스 구축)** AI윤리 원칙수립, AI 조직구성*, 위험 관리정책** 수립
  - * AI시스템의 전 과정에 대한 구성원의 역할·책임·권한 등 정의, AI 윤리위원회 설치 등
  - ** 소비자 권리보장, AI모델 및 학습데이터 관리, 사고발생시 감독당국가 소통 등 포함
- ㅇ **(고위험 서비스)** 개인의 권익, 안전, 자유에 중대한 위험을 초래할 수 있는 서비스에 대한 승인절차 및 승인책임자* 지정
  - * 최고위험관리책임자, 신용정보보호·관리인, 최고정보보호책임자 등 유사업무 겸직 가능
- ㅇ **(개인정보보호)** 민감정보 등 활용시 관련법상 사전동의 획득, 비식별 조치 등 준수, 정보의 재식별, 유출, 악용가능성이 없도록 조치
- ㅇ **(설명가능성)** 관련법상 고객 설명의무가 있는 서비스, 고위험 서비스 등에 AI 활용시 설명가능 인공지능 기술 적용 등 설명가능성 고려
- ㅇ **(소비자 권리보장)** 고객에 AI 이용여부, 설명·이의제기권 등 권리구제 방안 고지, 소비자 피해 발생시 명확한 책임조항 및 손해배상 처리 절차 등 마련

② 「금융분야 AI 개발·활용 안내서」 (금융위원회, '22.8월)

□ **(개요)** 모든 금융권에 적용되는 가이드라인으로서, AI에 대한 신뢰성·공정성 확보, 위험관리, 소비자 권리보장 등의 내용을 포함

- ㅇ **(공통사항)** 목적·적용범위, 거버넌스 구축, AI 업무위탁 시 주의사항 등
- ㅇ **(기획·설계)** AI 윤리원칙 검토, 의사결정 대체시 인간의 감독·통제절차 마련 등
- ㅇ **(개발)** 학습데이터 출처·품질·편향성 검증, 데이터 관련 컴플라이언스 준수, 설명가능한 AI 기술 도입 노력 등
- ㅇ **(평가·검증)** 적절한 성능·공정성 지표 선정·관리, 설명가능성 개선 노력
- ㅇ **(도입·운영·모니터링)** AI 이용 여부, 설명·이의제기권 등 소비자 권리구제방안 고지, 주기적인 성능관리, 데이터 오염 등 적대적 공격에 대한 보안대책 구축

② 「금융분야 AI 보안 가이드라인」 (금융보안원, '23.4월)

□ **(개요)** AI 모델 개발 시, 고려해야 할 보안사항을 단계별로 제시하고, AI 챗봇 서비스에 대한 보안성 체크리스트 제공

- ㅇ **(데이터 수집)** 데이터는 신뢰할 수 있는 출처로부터 수집하며, 이력 관리 수행
- ㅇ **(데이터 전처리)** 학습데이터 내 이상치, 변조값 등을 확인하며, 적대적 예제 데이터 등을 추가하여 AI 모델의 강건성 확보
- ㅇ **(모델 설계·학습)** 다양한 방식(앙상블 기법, 모델 튜닝 등)을 통해 모델의 강건성을 확보하며, 사전학습 모델 활용 시 출처를 유의하여 활용
- ㅇ **(모델 검증·평가)** 적절한 성능지표를 통해 성능을 검증, 적대적공격 테스트, 최종 출력값 확인, 안전장치(출력 횟수 제한, 적대적 공격 탐지 등)를 통해 보안성 확보