「금융권 AI 협의회」
운영방안

2024. 3.

금 융 위 원 회

금융데이터정책과

Ⅰ 추진배경

◈ **전세계적**으로 **금융권의 생성형 AI 활용**도는 **높은 수준**이나, **국내 금융회사**의 생성형 AI **활용**은 **미흡**한 상황

□ Morgan Stanley(美), NatWest(英) 등 **글로벌 금융회사**들은 업무 효율화 등을 위해 본격적으로 **생성형 AI를 활용***중이나,

* (Morgan Stanley) 고객 자산관리 자문서비스, (NatWest) 고객별 맞춤형 비서

○ **국내**의 경우, **일부 대형 금융회사**(대형은행 등)만 경량화 AI를 **내부망**에 **자체 구축**하여 실험적으로 **활용**하는 수준에 불과

◈ '24.2월 3차례 업권별(금투, 은행, 보험) **릴레이 간담회를 실시**하여 금융업계의 **생성형 AI 활용상 어려움 등** 의견을 수렴

1 全업권에서 금융권 망분리로 인해, 인터넷망을 통해 **해외서버에서 제공 중인 AI***서비스에서 내부망 자료를 활용하기 어려움을 토로

* GPT-4, Gemini 등 **다수 AI가 해외서버 + 인터넷 환경**에서만 활용 가능

○ 생성형 AI를 활용시 단시간에 만들 수 있는 PPT, 분석자료 등을 **수작업으로 진행**하여 **효율이 떨어짐**

2 AI 학습을 위해서는 "**양질의 데이터**" **확보가 필수**적이나, 현재 **개인정보 보호규제*** 등으로 **내부정보 활용이 어려운** 상황

* 실명정보의 경우, 이용자가 동의한 범위 내에서만 AI 활용·개발 등에 사용 가능

3 생성형 AI의 신뢰도*를 높이기 위해서는 **AI가 제시한 결과값 도출 배경을 설명**해주고, **필요시 개선**할 수 있는 기술이 필요

* GPT-4 등 생성형 AI의 경우, 남녀차별 등 선입견 등을 학습하는 **편향**(Bias), 틀린 대답을 하는 **환각** 현상(Hallucination) 등의 문제가 제기됨

○ 해외에서도 논의되고 있는 "**설명가능한 AI***" 기술을 마련하여, 금융회사들이 활용할 수 있도록 지원 필요

* **XAI**(e**X**plainable **AI**) : AI의 결정 과정을 **사람이 이해할 수 있도록 설명**하는 기술

⇨ 「**금융권 AI 협의회**」를 구성하여 AI 활용 활성화 방안 논의 필요

Ⅱ AI 협의회 구성안

◈ 금융권 AI 활성화를 위한 **공동협의회 발족**(부위원장 주재)

◈ **실무분과 논의**를 통해 **생성형 AI 활성화 방안** 구체화

□ **(금융권 AI 협의회) 금융권 AI 활성화**를 위한 **이슈 전반**을 **검토**하고, **지원·감독 정책 방향** 등을 **논의**

○ **(주재) 금융위 부위원장 / (간사) 금융혁신기획단장**

○ **(구성)** 금감원, 신정원, 보안원, 결제원, 금융연, 자본연, UNIST, 하나·신한은행, KB증권, 미래에셋자산운용, 삼성생명, 현대해상, 하나카드 등

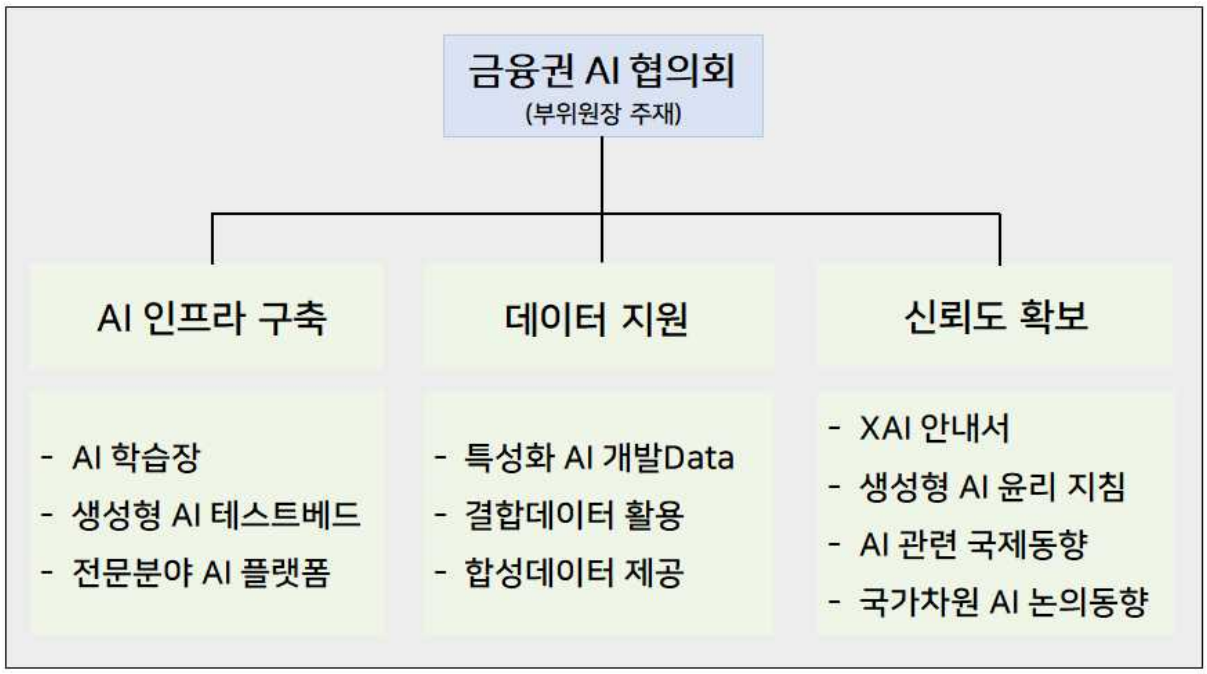
□ **(실무분과) 생성형 AI 활용 인프라 구축,** AI 개발 관련 **맞춤형 DB 구축** 등 **구체적인 당국 지원 필요사항** 등 논의*

* 업권별 AI 추진 현황 및 이슈 등을 수시로 논의하여 전체위원회에 보고

○ **(분과장) 금융혁신기획단장 / (간사) 금융위 데이터과장**

○ **(구성) 업권별 협회, 금융보안원, 신용정보원, 금융결제원** 및 **AI 관련 학계 및 업계 AI 전문가** 등

< 금융권 AI 협의회 조직도(안) >



Ⅲ 논의 주제

1. 금융권 생성형 AI 활용 인프라 구축

가. 금융권 보안규제 下 AI 활용 체계 마련

□ **(현황) 망분리[1] 및 해외서버 이용금지[2] 규제**를 고려할 때 금융회사의 **생성형 AI 활용이 사실상 어려운** 상황

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

2) 클라우드를 통해 **개인신용정보·고유식별정보를 처리**하는 경우, 그 **서버는 국내에 설치**되어야 함(전금規 §14의2⑧)

□ **(이슈)** 다수의 **생성형 AI**가 **해외서버**에 구축되거나, **인터넷 환경***에서 서비스되는 바, 이를 **활용할 수 있는 인프라 구축** 필요

* (해외서버 + 인터넷 필요) GPT-4(OpenAI), Gemini(구글), Hunyuan(混元, 텐센트)
(국내서버 + 인터넷 필요) HyperCLOVA(네이버), EXAONE 2.0(LG)

나. 생성형 AI 테스트베드의 구축 필요성

□ **(현황)** 보안원, 신정원, 결제원은 분석형 AI 관련 테스트베드를 구축('24.1월)했으나, **생성형 AI 테스트베드는 미흡**한 상황

□ **(이슈)** 향후 구축하게 될 **생성형 AI 활용 인프라의 특성***을 반영하여, **각 기관별 테스트베드 구축** 방식 등 논의 필요

* 생성형 AI의 활용·개발시 일시적인 GPU(그래픽 메모리) 사용량 폭증을 고려하여 금융사 개별 내부서버 보다는 공용 클라우드 환경에 구축할 필요

다. 손쉬운 이용을 위한 특성화 AI 플랫폼 구축

□ **(현황)** 일반적으로 논의되는 GPT-4 등 **생성형 AI 기초 모델***은 일반 금융회사 직원이 **업무에 직접 활용**하기 **불편**한 측면

* Foundation Model : 대규모 데이터 셋으로 학습한 거대 언어 모델(LLM, Large Language Model)로 해당 모델을 미세조정하여 특화 AI(예: Chat-GPT, Sora 등) 개발 가능

○ 이에 따라 대부분의 AI 제공사·전문업체(MS Azure 등)는 **업무에 활용하기 쉬운 특화 AI 플랫폼***을 제공중

* 금융상품 내용 및 논문·전문서적 등 요약, 각종 언어 번역, PPT 작성 AI 등

□ **(이슈) 망분리 규제하**에서 (중소)금융회사 등이 손쉽게 **전문분야 AI 플랫폼 수준**으로 **AI를 활용할 수 있는 방안** 검토 필요

2. 양질의 Data 지원

가. 특성화 AI 개발 지원을 위한 DataBase 구축

□ **(현황) 특성화 AI**(Fine-tuned AI) **개발에 사용**할 수 있는 "**양질의 데이터**"를 개별 금융회사가 대량 **확보하기 어려운** 상황

□ **(이슈)** 계속 발생하는 **다량의 Data**에 대해 금융회사들이 원하는 양질의 Data를 **지속적으로 확보·활용**할 수 있는 **체계** 검토 필요

○ 우선적으로 확보가 가능한 보험사기방지 등 **공공 Data를 중심**으로 **금융회사 보유 Data활용 방안**(가명처리 등) 논의

나. 자사·타사 정보 활용을 위한 가명 및 결합정보 활용 지원

□ **(현황)** 현재 **가명정보**의 **생성·결합*·평가 절차**에는 **약 2개월**이 소요되어 **Data의 최신성 확보** 등이 **어려운** 상황

* 데이터셋을 확대하기 위해 서로 다른 금융회사가 보유한 가명정보를 결합

○ 또한, **결합데이터 재사용***을 위한 「금융 AI 데이터 라이브러리」에는 **생성형 AI 학습·개발 용도**로 **활용**하기 위한 **인프라**가 **부재**

* 결합데이터는 사용 직후 파기가 원칙(신정令 §14의2③vi)이나, 「금융 AI 데이터 라이브러리」는 파기하지 않고 저장 가능

□ **(이슈)** 가명처리 및 적정성평가 등의 **소요기간 축소 방안** 및 「금융 AI 데이터 라이브러리」의 **생성형 AI 학습환경 연계** 등 논의

다. AI 학습을 위한 합성데이터 등 충분한 Data 확보

□ **(현황)** 합성데이터*의 **생성 방법, 익명성·유용성 판단기준** 등에 대한 **가이드라인 등이 부재**함에 따라 **AI 활용에 제한**

* 실제 데이터의 특성과 통계적 특성을 반영하여 인공적으로 생성한 데이터

□ **(이슈)** 합성데이터 **생성 주요 기법*** 및 **식별정보 노출 위험 평가**, 업권별 **합성데이터 연구 사례, 활용 전략** 등 논의 필요

* 합성데이터의 익명성과 유용성은 상충 관계로, 적절한 최적점을 찾는 것이 중요

3. AI 신뢰도 확보

가. 설명가능한 AI(XAI) 기술을 통한 생성형 AI 수준 제고

□ **(현황) AI가 산출한 결과**에 대해 ①금융회사가 **오류를 즉시 시정**하거나, ②**고객에게 설명**하기 어려워 **AI 자체의 신뢰도**가 낮음

○ **(금융회사)** AI 자체의 **정확성을 지속적으로 개선**하기 위해 **AI 신뢰성 평가기준***이 필요하나, **가이드라인** 등이 **부재**

* AI 모델에 입력되는 변수가 결과값에 미치는 영향도를 정량화하여 일정 기준을 넘는 AI에 대해서만 활용 가능

○ **(금융소비자)** 금소법 등 **보호범위 내**에서 **AI 서비스**를 받기 위해 **AI의 판단 근거에 대한 설명*** 등이 **필요**하나 **미비**

* 예) XAI 기술 미적용시 대출심사에 거절되는 경우 이유에 대한 설명이 불가능한 반면, XAI 기술 적용시 어떠한 이유로 거절되었는지 설명 가능

□ **(이슈) 금융회사** 및 **금융소비자**가 **AI의 판단 배경**을 이해할 수 있도록 「**설명가능한 AI 안내서**」 마련 등에 대해 논의 필요

나. 생성형 AI 윤리지침을 통한 올바른 AI 교육 지원

□ **(현황)** AI가 잘못된 지침, 편향된 정보 등 **비윤리적인 Data로 학습**할 경우 소비자·금융사의 **피해***가 발생할 수 있으나 **대책 부재**

* AI는 아이와 같아서 어떤 Data로 학습하는지에 따라 천차만별로 달라지며, 불완전판매 등을 유도하는 AI로 진화할 가능성 존재

□ **(이슈)** AI의 학습 단계부터 **윤리기준*을 적용**하여 AI가 **도출한 결과**가 **윤리적 기준에 부합**하도록 하는 **방안** 논의

* AI의 책임감 있고 공정한 사용을 규율하는 원칙

○ 「**생성형 AI 윤리 지침***」 제정시 **소비자 및 금융사 피해 예방**을 위해 포함되어야 하는 내용 등 논의

* '21.7.8일 발표한 「금융 AI 가이드라인」에는 생성형 AI(Chat-GPT, '22.11.30일 발표) 관련 윤리지침까지는 반영하지 못함