금융권 인공지능(AI) 활용 활성화 및 신뢰확보 방안

2022. 8.

금융혁신기획단
금융데이터정책과

목 차

Ⅰ. 추진배경

◇ AI가 산업 전반의 혁신을 주도하는 AI 시대 본격화

□ 인공지능(AI)은 **4차 산업혁명**을 이끄는 **핵심기술**이자 **범용기술**로서 경제·사회 全 분야의 **디지털 전환**(Digital Transformation)을 **촉진**

○ 빅데이터, 클라우드 등 **디지털 기술의 발전**과 **코로나로 인한 비대면화** 등과 맞물려 산업 전반에 걸쳐 **AI 도입**이 **가속화**되는 추세

□ **AI** 활용은 **산업 전반**의 **생산성**을 **제고***하여, AI에 적극적인 **선도국가**와 **후발주자**간 **생산성 격차가 확대**될 것으로 예상

* 인공지능(AI) 도입을 통해 생산성이 40% 향상되고, 2030년까지 직업의 약 1/3이 AI 기술로 대체될 것으로 전망(맥킨지, 2018)

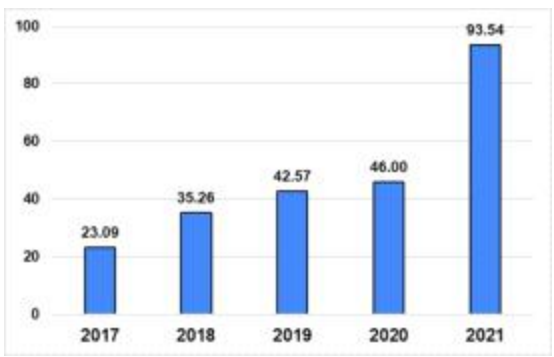
◇ 미국 등 세계 주요국의 AI 투자는 대폭 확대 추세

□ 미국 등 **세계 주요국**들은 **AI**가 향후 경제·사회·안보 등 全 분야에 걸쳐 **국가경쟁력을 좌우**할 것으로 판단하고,

○ **AI 주도권 확보**를 위한 **투자, 기술 개발** 등을 **대폭 확대**중

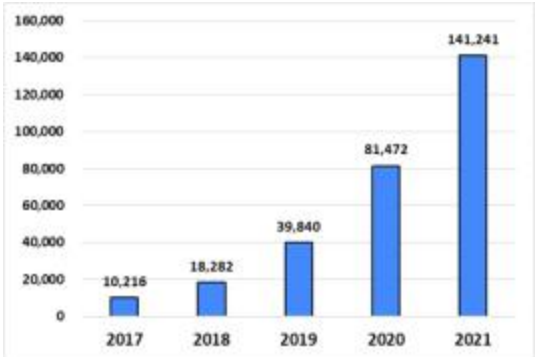
* '21년 전 세계 AI분야 **민간투자액**은 **전년 대비 약 2배 증가**한 935억 달러, **특허 출원**은 전년 대비 **약 1.7배 증가**한 141,241건으로 조사

(단위 : 십억 달러)



'17~'21년 전 세계 AI분야 민간투자액

(단위 : 건)



'17~'21년 전 세계 AI분야 특허 출원 건수

* 출처 : 2022 AI Index Report(스탠포드대학교 HAI)

◇ 국내도 AI를 국정과제로 선정하여 총력 지원

□ 우리나라도 디지털 전환 시대 **디지털 강국으로 도약**하기 위해 **신기술·신산업** 분야에서 **최고 수준**의 **국가역량 확보에 총력**

○ 「AI 초일류 국가」를 **국정과제**로 선정하고, **AI**에 대한 R&D 중점 투자 및 관련 **인프라 구축** 등 **다양한 지원책**을 마련·추진 ('22.6.16, 새정부 경제정책방향)

※ '22.6.16, 새정부 경제정책방향에서 발표된 AI 활성화 정책

■ (과학기술·R&D 혁신) 신기술 확보·신산업 육성에 중점을 둔 과학기술·R&D 정책 수립
- 국가·사회적 난제 해결을 위한 AI 등 메가프로젝트에 R&D 중점 투자

■ (미래대비 선도경제) 인공지능 등 유망 신사업 전략적 육성 지원
- 차세대 AI, 데이터로 이어지는 디지털 초일류 기반 조성

◇ 금융분야는 AI가 가장 잘 활용될 수 있는 분야중 하나

□ **금융분야**는 **마이데이터** 도입*('20.8월), **데이터 결합 활성화**** 등 **빅데이터 활성화 기반이 구축되**어 데이터 활용이 가장 활성화된 분야

* '20.8월 개정 신정법 시행을 통해 금융분야에서 최초로 마이데이터 도입 ('21.7월말 기준 59개 마이데이터 사업자 허가)

** '21년 말 기준 데이터 결합 건수 : 금융분야 112건, 비금융분야 14건

□ 현재 **금융분야**는 AI 활용 **초기단계**로 **로보어드바이저, 챗봇, 상품 추천, 이상거래탐지, 신용평가 및 여신심사** 등에서 **활용**중

○ 향후 **금융산업의 디지털 전환** 및 **생산성 혁신**을 이끄는 주요 기술로서 미래 **금융산업의 지형**을 **근본적으로 변화**시킬 것으로 예상

➡ 금융권 **AI 기반 혁신 가속화** 및 **금융산업 강국**으로 **도약**하기 위해 **금융권 특성에 맞는 AI 활성화 방안** 마련 필요

Ⅱ. 금융권 인공지능(AI) 활용 활성화의 중요성

◇ 최근 우리 **금융산업**은 **규제 혁신**과 **새로운 Player**들의 출현 등으로 **외연**이 **확장**되는 **양적 변화**를 경험

○ 앞으로는 **AI**가 **맞춤형 서비스**를 통한 **소비자 편익 및 금융 중개기능 제고** 등 **금융의 질적변화**를 **주도**할 전망

1 금융소비자 편익 제고

○ AI가 다양한 금융정보*를 **자동**으로 **분석**하여, 개별 **금융소비자**의 **특성**에 맞는 **금융상품 추천** 등 **맞춤형 금융서비스** 제공

* 예: 신용카드 결제내역, 보험료 납부정보, 투자정보, 자동이체정보 등

2 금융 중개기능 제고

○ AI를 활용하여 **데이터 처리**의 **속도와 정확성**을 **개선**하면서 자금이 필요한 곳에 **적시에 충분한 자금을 공급**

3 금융 안정성 제고

○ AI를 통해 **여신심사, 신용평가, 보험 인수심사** 등 핵심적 금융 업무의 심사·평가가 **정교화**되어 금융회사 **리스크 관리기능 제고**

4 금융 포용성 제고

○ AI를 통해 **금융이력부족자**(Thin-filer)의 **신용평가**가 **가능***해져, **합리적 조건**으로 **금융**을 이용할 수 있는 여건 조성

* AI를 활용하여 대량의 비정형·비금융 데이터를 수집·분석하여 대안신용평가 모형 개발·운용 가능

Ⅲ. 현황 및 문제점

◇ ①**양질의 데이터 부족**, ②**제도 미비**, ③**신뢰성 부족** 등이 금융분야 **AI 활성화를 저해**하는 **주요요인**으로 작용

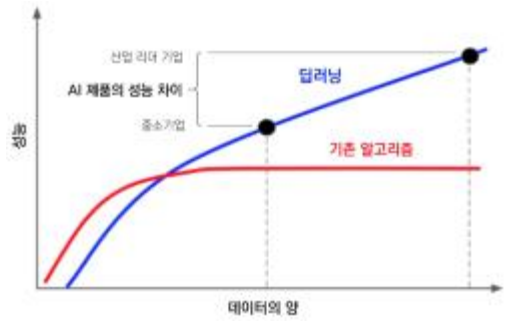
가. 양질의 데이터 부족

□ **데이터**는 **AI의 성능**에 직접적인 영향을 미치는 **핵심**적인 요소

○ **AI의 개발·학습**을 위한 **원활한 데이터 확보**를 위해 **국내·외**에서는 **데이터 댐** 구축 등 관련 정책을 적극 추진중

* **(국내)** 과기부, NIA(지능정보사회진흥원) 등은 데이터를 생산하여 공급하는 **빅데이터 센터**와 데이터 가공·분석 및 유통기반을 제공하는 **빅데이터 플랫폼**을 구축중('19년~)
(국외) 미국 연방정부는 매년 **미국 연방 데이터 전략 시행 계획**를 수립하여 공익적 가치가 있는 데이터를 **AI 연구에 바로 활용할 수 있도록 재가공·개방** 추진중

※ **데이터의 양에 따른 AI 성능 변화 비교**



■ 기존의 AI를 활용하지 않는 디지털 서비스(예: 베스트 수익률 펀드 추천 등)는 데이터 양과 성능간 관계가 제한적이나, 딥러닝 등 AI를 활용한 디지털 서비스(예: 이용자 맞춤형 펀드 추천)는 데이터의 양에 비례하여 성능 향상

* 출처 : AI 학습용 실효성 향상을 위한 정책 방향 (한국정보화진흥원, 2020)

□ 아직 **금융관련** AI 개발·학습 및 테스트 등에 활용가능한 충분한 **양질의 금융 빅데이터 확보**가 **어려운 상황**

※ 국내은행 대상 설문조사시 **인공지능 개발·도입**의 가장 큰 **제약요인**으로 **데이터 부족**을 응답(금융연구원, '22.2월)

❶ **가명정보** 제도 **도입**으로 정보주체 동의없이도 비식별화된 **개인정보**를 **AI 개발** 등에 **활용 가능**하게 되었으나,

- 가명정보는 가명처리시 정한 **목적으로만 활용**하고 **파기**하여야 함에 따라 **대량의 데이터**를 **축적**하여 **활용**하는데 **곤란**

- 데이터를 결합해주는 **데이터전문기관**[*]도 **4곳**(신정원, 금결원, 금보원, 국세청)만 지정·운영되고 있어, **민간중심의 가명정보 결합·활용**이 **활성화**되기에 **불충분**

* 금융회사는 데이터전문기관을 통해서만 데이터결합 가능(금융회사와 일반기업간 결합 포함)

❷ 과기부 등의 정부예산사업을 통해 **챗봇, 음성봇** 등 개발에 활용가능한 **말뭉치**[*] **데이터 셋**이 **점차 확충**[**]되고 있으나,

* **말뭉치(Corpus) :** 텍스트를 컴퓨터가 읽을 수 있는 형태로 구성한 언어 자료
** **AIHub 홈페이지**를 통해 일상생활, 상거래, 교육 등 다양한 분야 말뭉치 제공중

- 금융분야 **AI 개발**을 위해서는 **금융분야**에 **특화**된 **대량**의 전문적·비일상적 **말뭉치**가 필요

나. 제도 미비

□ **현행 금융관련 제도**가 AI 관련 내용을 충분히 **반영하지 못함**

❶ 중소 금융회사 등이 **AI를 도입·활용**하기 위해 **검토**하여야 할 **사항** 등에 대한 **안내 미비**

- AI 개발·활용 **全 주기**[*]에 걸쳐 실무자들이 **참고할 수 있는 주요사례**(Best Practice) 등 **제공 필요**

* 기획·설계단계→ 개발단계 → 평가·검증단계 → 도입·운영·모니터링단계

❷ 해외 주요국의 경우 **소비자 보호**[*]를 위해 **「설명가능한 AI」**에 대해 활발히 **논의 중**이나, **국내 금융분야**는 **관련 검토 부족**

* AI 알고리즘의 특성(Black box)상 판단결과에 대해 충분한 설명이 이루어지지 못할 경우 소비자 피해 발생 가능(예: AI 대출거절 → 거절사유 설명곤란 → 금융접근성 저하 등)

❸ AI **개발·활용**에는 **외부 API**[*] 및 **대규모 자원**(클라우드컴퓨팅 등) **활용**이 **필요**하나 **금융분야**는 **망분리** 등 **규제**로 인해 **활용**에 **제약**

* AI 개발시 일반적으로 인터넷에 공개되어 있는 다양한 오픈 API 활용

다. 신뢰성 부족

□ 금융분야 AI 활용은 아직 **초기단계**로 **기술**의 **투명성, 공정성** 등에 대한 **사회적 신뢰**가 충분하지 않은 상황

○ **'이루다 사건**[*]' 등 AI에 대한 시민사회 우려 등을 감안시 금융분야 AI 활용이 **지속 확대·활성화**되기 위해서는 **사회적 신뢰를 확보**하는 것이 **중요**

* AI 챗봇 '이루다'는 대화시 AI 학습에 활용된 개인정보를 유출하거나, 오염된 데이터 학습으로 여성·장애인·동성애 등에 대한 차별·혐오발언을 하여 논란 발생

□ '21.7월 **「금융분야 AI 가이드라인**[*]**」**을 통해 **신뢰성있는 AI 서비스 운영**을 위한 **원칙** 등이 마련되었으나,

○ 금융회사 AI 서비스의 신뢰성을 **제3자**가 **객관적**으로 **검증하는 인프라**는 다소 **미흡**

* 주요내용: AI 윤리 원칙 마련, AI 거버넌스 구성, 위험관리정책 수립, 학습 데이터 관리, 개인정보 보호, 성능평가, 공정성 제고, 금융소비자 권리 고지 등

□ AI의 **투명성** 및 **공정성 확보**를 **지원하는 인프라** 구축·운영을 통해 **사회적 신뢰**를 보다 **공고히 할 필요**

❶ AI의 **정상적 작동을 검증**할 수 있도록 지원하는 **테스트베드 구축**을 통해 금융회사 등의 **자율적 신뢰성 확보 노력 지원** 필요

❷ **신용평가** 등과 같이 금융소비자에 **중대한 영향**을 미치는 분야의 경우 제3자를 통한 **객관적인 검증체계** 마련 필요

❸ **AI 활용 확대**로 인한 정보유출 등 **보안사고**가 발생하지 않도록 철저한 **AI 보안 검증체계** 구축·운영 필요

□ 또한, 금융감독 등 **공공분야**에서도 AI 활용 **우수사례를 적극 발굴**하여 **변화하는 사회 환경**에 효과적으로 **대응**해 나갈 필요 (감독·행정 신뢰성 제고)

➡ ①**데이터 확보**, ③**제도 정립**, ②**신뢰 강화**를 통해 **금융분야 AI 활용**을 **활성**화 할 수 있는 **기반 확보 추진**

Ⅳ. 추진 방향

◇ 금융분야에서 **AI 활용**이 **안전하게 확대·정착**될 수 있도록 **❶ 양질의 빅데이터** 확보 지원, **❷ AI 활성화를 위한 제도** 정립, **❸ 신뢰받는 AI 활용 환경 구축** 추진

❶ (**데이터 확보**) **AI 개발** 등에 활용할 수 있는 **데이터**를 **원활하게 확보**할 수 있도록 데이터 라이브러리 등 **인프라 구축**

❷ (**제도 정립**) **AI 활용 시도**가 보다 **적극적**으로 이루어 질 수 있도록 관련 **가이드라인 마련** 및 **규제 합리화**

❸ (**신뢰 제고**) **AI 활용**에 대한 **사회적 신뢰**가 **제고**될 수 있도록 **평가·검증체계 구축**

1 AI 활성화를 위한 빅데이터 확보 지원

1 「금융 AI 데이터 라이브러리」 구축

□ (필요성) **데이터는 AI의 성능**을 결정하는 중요한 요소라는 점에서 **양질의 빅데이터**를 원활히 **활용**할 수 있는 환경 구축이 필요

○ 현재는 **데이터**(가명정보) **셋**을 구축해도 **사용 후 파기***(재사용 금지)해야 함에 따라 대량의 데이터 셋 **구축·운영이 곤란**

* 신정법(§20의2②2의2. 등)에 따라 가명정보는 이용 목적을 정하고 해당목적에 맞게 비식별 처리한 후 활용하고, 이용 목적이 달성되면 파기하여야 함

- **旣 구축한 데이터 셋**이 있음에도 **다른 목적**으로 **활용**하기 위해 **데이터**를 **재결합**해야 하는 **비효율** 발생*

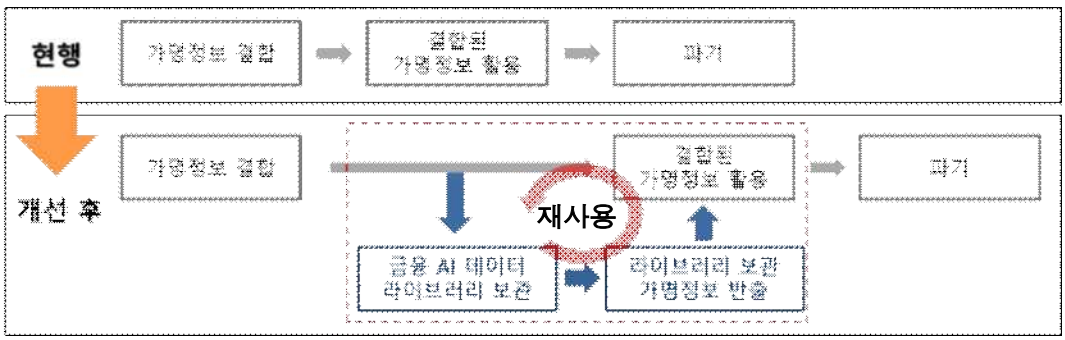
* 예: 신용평가모델 개발 목적으로 데이터 결합을 통해 구축한 가명정보 데이터 셋은 내용상 차주별·산업별 여신현황 분석 등 타 연구에도 활용가능하나, 현재는 재사용 금지 규제로 데이터 셋을 새로 구축하여 활용하여야 함

○ 한편, 데이터 셋 **재사용 허용**시 **개인정보 침해** 발생 가능성* 등 **정보보호 측면**을 **고려**할 필요

* **재사용**을 위해 데이터 셋을 파기하지 않고 **보관**함에 따른 **보안사고 가능성 증가**, **다양한 용도로 활용**될 수 있도록 **최소한**의 수준으로만 **비식별처리** 할 우려 등

⇒ **개인정보**를 **침해하지 않으면서 재사용**을 통해 **대량의 데이터 셋**을 원활히 **구축·활용**할 수 있는 **환경** 마련 필요

□ (추진방안) **규제 샌드박스** 등을 통해 데이터 결합 후 **데이터 재사용을 허용**하는 **「금융 AI 데이터 라이브러리」** 구축



* 가명처리시 정한 목적이 달성되면 해당 가명정보 파기

○ **이종산업**간 데이터 **결합·활용**이 활성화될 수 있도록 **다양한 회사**로 구성된 **컨소시엄**을 통해 **라이브러리 구축**

- 우선, **종합신용정보집중기관**으로서 **개인정보보호에 전문성**을 갖춘 **신정원**을 중심으로 **컨소시엄*** **구성** 추진

* **컨소시엄 참여업체 확정** 후 **'22.3분기** 중 **컨소시엄 출범 추진**

※ 향후 **운영성과** 등을 보아가며 **AI 데이터 라이브러리 운영기관 확대** 및 **신정법 등 개정**을 통한 AI 데이터 라이브러리의 **법적근거 마련 방안** 검토

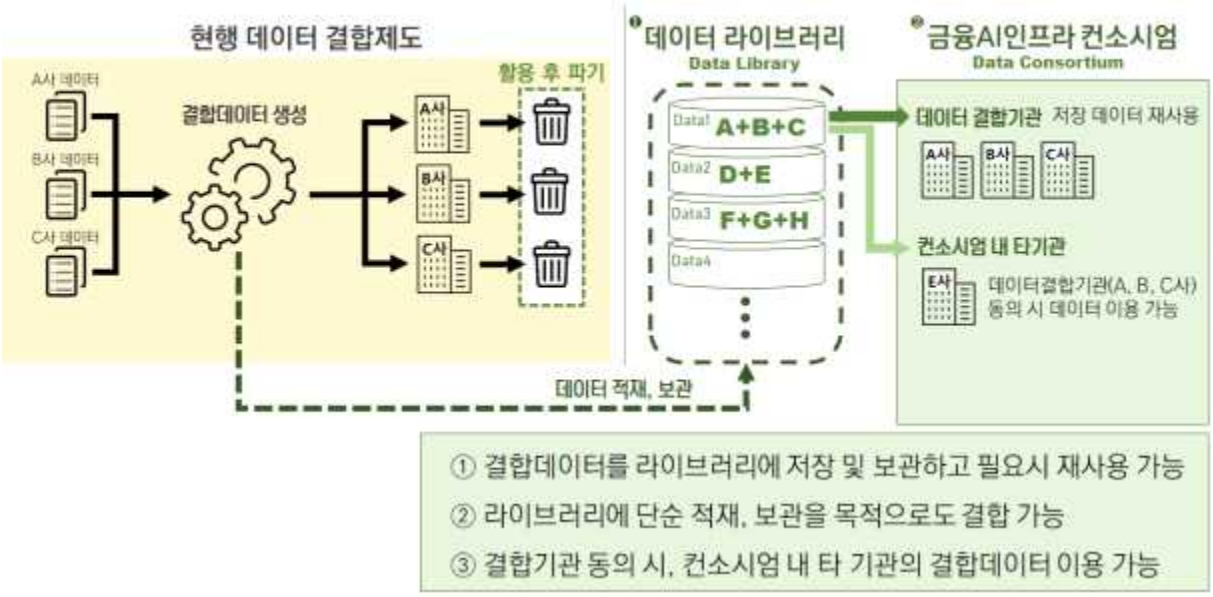
○ 라이브러리에 저장된 데이터는 컨소시엄 **참여기관**이 **필요**시 **인출***하여 **재사용** 가능

* 라이브러리에 저장한 데이터를 **활용목적에 맞추어 가공**하여 인출·활용
(예: 신용평가모형 개발용 데이터를 산업별 부도율 분석에 활용시 신용등급 제거 후 인출 등)

○ **정보유출** 등이 발생하지 않도록 **고도의 데이터 보호체계** 구축

* 예: 철저한 **물리적 망분리 및 업무분리**(Firewall), 접근통제, 각종 보안시스템 운영 등을 포함한 관리적·기술적·물리적 보호체계 수립 및 운영

〈 금융 AI 데이터 라이브러리 구축(안) 〉



2 금융권 협업을 통한 금융 빅데이터 공동 확보

□ (필요성) **챗봇, 이상거래탐지시스템**(FDS) 등의 경우 AI **개발**에 대량의 **비정형***·**전문데이터**가 필요하나, 데이터 확보에 따른 **비용 부담 등**으로 AI 개발·활용에 **애로**

* **AI 개발 등에 개인정보**를 **활용**하기 위해서는 **비식별처리**를 하여야 하나, **비정형데이터**(예: 대화내용)는 비식별처리시 **데이터 내 개인정보 항목**을 **일일이 찾아내야**하는 등 **부담 高**
(정형데이터(예: 거래잔고)의 경우 데이터 내 개인정보를 쉽게 구분가능(예: 계좌번호 등))

○ **금융분야 챗봇·FDS** 등의 경우 사용되는 **용어**와 **데이터**가 **전문적***이므로 **타 분야 빅데이터** 활용이 **어려움**

* 금융투자상품, 신용평가 등 금융서비스에 대한 질의·답변시 전문적 용어 활용

○ 현재 각 금융회사는 **개별적**으로 **빅데이터를 수집·생성***하여 **챗봇·음성봇, FDS 개발** 등에 **활용**하는 상황

* 챗봇 프로그램의 경우 금융회사 직원들이 예상 질문 및 답변을 일일이 작성하거나 거액의 외주용역 등을 통해 관련 빅데이터 셋을 구축

⇒ **금융회사간 협력**을 통해 **공동의 빅데이터**를 **구축**하여 AI 개발 활성화를 도모할 필요

□ (추진방안) **협회, 금융분야 데이터 인프라 기관** 등을 중심*으로 **금융권이 공동으로 사용가능한 AI 빅데이터**(말뭉치 데이터, 사기탐지 데이터 등) 구축

* 예 : 말뭉치 데이터 – 금융협회(은행연 등), 이상거래탐지 데이터 – 금융보안원, 금융결제원 등

○ 구축된 데이터 셋은 원칙적으로 **참여 금융회사가 활용**할 수 있도록 하고, 필요시 **참여자 협의**를 통해 **추가 활용방안*** 검토

* 금융데이터거래소에 상품으로 판매, D-테스트베드 활용 등

3 데이터 전문기관 추가 지정

□ (필요성) **AI 개발 활성화**를 위해서는 이종산업간 **데이터 결합**이 **활성화**될 수 있는 **환경 구축** 필요

○ **데이터전문기관**을 **추가 지정**하여 **데이터 결합**이 보다 편리하고 원활하게 이루어 질 수 있도록 **지원**할 필요

※ 데이터 이용기관의 데이터 결합신청 허용, 샘플링 결합 도입 등 데이터 결합제도 개선은 기 완료('22.7.7. 신용정보업감독규정 개정)

□ (**추진방안**) **데이터 결합 활성화**에 적극적으로 **기여**할 수 있는 **신뢰성, 전문성, 개방성*** 등을 갖춘 기관을 **선별**하여 데이터 전문기관으로 **추가 지정**('22.1월 旣 발표)

* ❶ (신뢰성) 사회적 신뢰 확보가 가능한 기관이 중심역할을 수행
⇒ 보안 및 이해상충방지 체계가 우수한 기관을 지정
❷ (전문성) 데이터 결합 전문성이 높은 기관이 데이터 결합을 선도
⇒ 데이터 분야 업무 역량 및 실적이 우수한 기관을 지정
❸ (개방성) 데이터 산업 경쟁을 촉진할 수 있는 기반 조성
⇒ 데이터 개방·공유에 적극적인 기관을 지정

○ **예비지정신청서**를 **접수***('22.7월)한 기관 대상으로 **'22.3분기** 중 데이터전문기관 **예비지정**

* '22.2월 사전신청서를 제출한 12개 기관(금융기관 6개, 비금융기관 6개) 대상으로 예비지정 신청서 접수

○ 예비지정 이후 **전산설비 구축** 등 데이터전문기관 **운영 준비가 완료**된 기관에 대하여 **본 지정 심사**('22.4분기중 예상)

※ 신정법령에 따라 데이터전문기관 지정이후에도 데이터전문기관의 충실한 업무수행 유도 등을 위해 데이터전문기관의 적격성을 주기적(3년)으로 재확인

2 AI 활성화를 위한 제도 정립

1 금융분야 AI 개발·활용 안내서 발간

□ (**필요성**) '21.7월 금융분야 AI 활용시 고려해야 할 **원칙** 등을 **규정**한 「**금융분야 AI 가이드라인**」이 **마련**되었으나,

○ 실제 구체적인 **AI서비스 도입시 참고**가능한 **기능·서비스별 안내서**가 **필요**하다는 현장 실무자들의 의견 제기

□ (**추진방안**) 금융분야에서 **AI가 가장 많이 활용**되는 **5대 서비스*별 「AI 개발·활용 안내서」**를 제작

* 신용평가 및 여신심사, 로보어드바이저, 챗봇, 맞춤형 추천, 이상거래탐지(FDS)

○ **금융업권** 및 **기능·서비스별 특성**을 고려하여 개발·활용 단계별* **세부 안내서** 마련

* 기획·설계→개발→평가·검증→도입·운영·모니터링 등 각 단계별로 상세히 안내

○ **실무자**들이 **쉽게 이용**할 수 있도록 **체크리스트 형태**로 구성

※ **(참고) 안내서 내 체크리스트 문항 예시**

단계	구분	내용
거버넌스 구축	점검 항목	고위험 서비스에 AI를 활용하기 위한 승인 절차와 승인 책임자 지정 등 적절한 내부통제가 마련되어 있는가?
	체크 리스트	2) 고위험 서비스의 경우, 적절한 내부통제가 마련되어 있는가? - 고위험 서비스를 도입하는 경우, CRO 등 책임있는 업무 수행이 가능한 자 또는 관련 위원회가 검토 후 승인할 수 있는 내부절차가 마련되어 있는지를 확인한다. YES □ | NO □ | N/A □
	준수 사례 (예시)	- A보험사는 개인의 권리에 중대한 위험을 초래할 수 있는 고위험 서비스 관리를 위해 승인 책임자 지정 또는 관련 위원회를 구성하여 내부통제 및 승인 절차를 준수하고 있으며, 고위험 서비스의 주요 내용 변경 시 내부통제 절차를 따라 수행한다.

○ 각 **금융협회**를 중심으로 **실무 작업반***을 구성하여 '**현장의 목소리**'를 **충실히 반영**

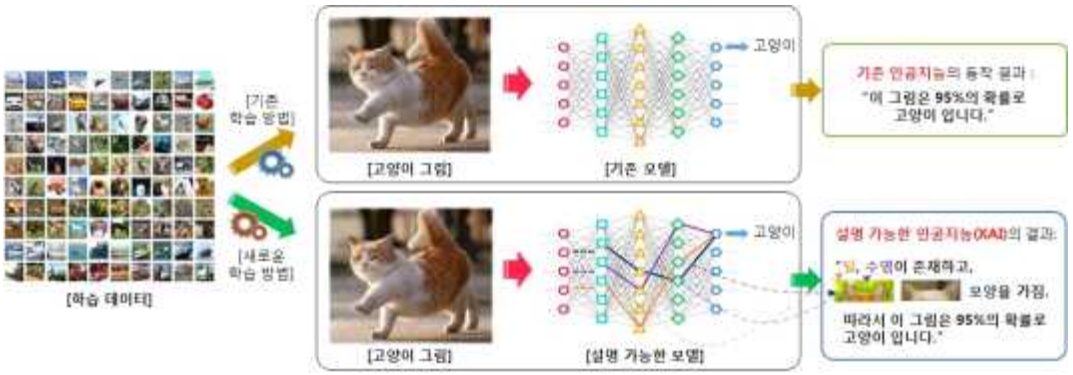
* 5대 서비스별 작업반 구성 : 은행연합회(신용평가 및 여신심사), 금융투자협회(로보어드바이저), 생보협회(챗봇), 손보협회(맞춤형 추천), 여전협회(이상거래탐지)

2 설명가능한 AI(XAI : eXplainable AI) 요건 마련

□ (필요성) AI를 통한 **의사결정**으로부터 **소비자 보호** 등을 위해 **국내·외**에서는 '**설명가능한 AI**(XAI)'에 대한 **논의**가 활발히 진행중*

* (국내) 과기부 '설명가능인공지능 연구센터'를 통해 XAI 연구 진행('17.7월~)
(국외) 美 국방성 국방위고등연구계획국(DARPA)의 설명가능 AI 연구(Glass Box)('17년~) 등

※ **(참고) 설명가능한 AI(XAI) 예시**(설명가능한 인공지능 소개, 금융보안원(2018))



○ **금융분야**도 **AI 의사결정**(상품추천, 여신심사 등) **확대**에 **대비***하여 금융회사 등이 **XAI**를 보다 **원활히 도입·운영**할 수 있도록 관련 **기준**** 등을 **검토**해 나가야 할 시점

* 설명가능성이 확보되지 않는 경우 AI가 여신을 불승인해도 소비자는 불승인 사유를 알지 못해 이를 해소하지 못하여 장기간 금융접근성이 제한될 우려

** 예: 통상 딥러닝 등 **AI 알고리즘**은 **블랙박스** 형태로 개발되는 경우가 많아, **유사한 성능**을 가진 **설명 알고리즘** 등을 **별도로 개발·운영** 필요 → **허용가능**한 설명 알고리즘과 原 알고리즘간 **성능**(예측결과) **차이**(예: 90%일치 등) **요건** 등 마련

□ (추진방안) 연구용역을 통해 **설명가능한 AI(XAI) 정의 및 요건, 구현사례**(신용평가분야 XAI 모범사례 등) 등 **안내서 발간**

※ **(참고) 미국 국가표준기술연구원**(NIST)에서는 '**설명가능한 AI의 네가지 원칙***(Four Principles of Explainable Artificial Intelligence)' 초안을 발표('20.8월)

* **보고서에서 제시**된 **설명가능한 AI의 네가지 원칙**

- AI 시스템은 모든 산출물에 대해 **증거 혹은 그 이유**를 제시해야 한다.
- 시스템은 개별 이용자들에게 **의미 있거나 이해할 수 있는 설명**을 제시하여야 한다.
- 해당 설명은 산출물을 생성하는 **시스템 프로세스를 적절하게 반영**해야 한다.
- 시스템은 충분히 **신뢰할 수 있을만한 산출물을 생성**할 것이라는 믿음을 전제하고 운영되어야 한다.

3 망분리 및 클라우드 규제 개선

□ (필요성) 원활한 **AI 개발·활용**을 위해서는 **외부 API 및 클라우드 활용**이 필요하나 **보안** 규제 등으로 **활용이 제한적**

○ (**망분리**) AI 서비스 개발시 **다양한 AI 알고리즘 적용** 및 **검증**을 위해 대다수 기업은 **개발단계**에서 **외부의 AI API를 활용**중

* AI 분야의 경우 다양한 알고리즘을 테스트하여 서비스에 최적화된 알고리즘을 탐색·적용하는 것이 일반적 → 외부 API 활용시 관련 소요 기간·비용 절감 가능

- **금융회사**의 경우 **망분리 규제***로 인해 **외부 API 사용이 제한**되어 개발 **소요기간·비용 등이 증가**

* 금융회사는 업무망(내부망)과 인터넷망(외부망)을 물리적으로 분리하여 운영토록 규제

○ (**클라우드**) **고성능 AI 개발·활용**을 위해서는 **클라우드**를 통한 **대규모 컴퓨팅 자원**이 **필요***하나, **금융회사**의 경우 **복잡한 이용절차****로 인해 클라우드 이용이 **원활하지 못한 측면**

* 알파고('16년)의 경우 CPU 1,202개, GPU 176개로 구성된 클라우드 활용
** 시스템 중요도 평가, 금감원 사전보고 등

⇒ AI 개발·활용 활성화를 위한 망분리 및 클라우드 규제 개선 필요

□ (**추진방안**) 전자금융의 **보안성 및 안정성**은 유지하면서 **금융회사**의 **AI 개발·활용**이 보다 원활토록 **망분리 및 클라우드 규제 개선** (전자금융감독규정 개정 등)

○ 외부 API를 보다 원활히 활용가능토록 **가명정보** 등을 **활용**하는 **개발·테스트 서버**에 대해서는 **규제 샌드박스** 등을 통해 **물리적 망분리 예외 허용***

* 오픈소스 접속·활용 등에 대한 내부기준을 수립·이행할 것 등 보완조치를 전제로 허용

○ 업무 중요도에 따른 **클라우드 이용절차 차등화**, 클라우드 이용시 사전보고를 **사후보고로 전환** 등 **클라우드 이용절차 개선**

3 신뢰받는 인공지능(AI) 활용 환경 구축

1 금융 AI 테스트베드 구축

□ (필요성) 금융 AI의 **정확성 및 신뢰성** 등을 **확인**하기 위해서는 **다양한 데이터**를 통해 검증하는 과정이 중요

○ **금융회사** 등은 **개발**에 활용하는 **데이터도 부족**한 상황으로 다양한 테스트 데이터를 통한 **검증**에 애로

* AI 개발시 정확한 AI 성능 확인을 위해 일반적으로 보유 데이터를 개발과 테스트 용도로 나누어 사용하나, AI 성능 확보를 위해 필요한 절대적 데이터 양도 부족하여 별도의 테스트용 데이터 할당 애로

○ **대량의 데이터**를 활용하여 효과적으로 **AI를 테스트**하기 위해서는 **고성능의 컴퓨팅 자원이 필요**

⇒ **금융회사** 등의 AI 활용 **신뢰성 확보** 등을 위해 **다양한 테스트용 데이터** 및 **컴퓨팅 자원을 제공**하는 **테스트베드 구축** 필요

□ (추진방안) **금융분야 인프라 기관** 등을 통해 **다양한** 금융분야 AI 테스트가 가능한 **검증 데이터 및 테스트 환경** 구축

○ 우선, **신용평가 AI**(신정원), **금융사기방지 AI**(금결원), **금융보안 AI**(금보원) **테스트베드 구축**을 **추진**하고 운영현황 등을 모니터링하여 **확대 검토**

- **신정원·금결원·금보원**이 보유한 관련 정보(신정원 : 개인·기업신용정보 등, 금결원 : 이체거래내역 등, 금보원 : 침해사고 정보 등)를 활용하여 관련 **AI**의 **테스트**에 활용할 수 있는 **데이터 셋 구축**

※ 보다 다양한 데이터를 테스트에 활용할 수 있도록 중장기적으로 기관간 협업을 통해 타 유관기관 등의 관련 데이터가 결합된 데이터 셋 확충 추진

- 신정원, 금결원, 금보원의 **컴퓨팅 자원** 등을 통해 **同 데이터 셋** 등을 **활용***하여 **AI 학습** 및 **테스트**를 할 수 있도록 **지원**

* 기업 등이 자신이 보유한 데이터와 同 데이터 셋을 결합하여 활용하는 것도 지원

2 AI 기반 신용평가모형 검증체계 마련

□ (**필요성**) 금융회사와 핀테크·플랫폼 업체들의 **다양한 비금융정보와 AI 방법론**을 활용한 **신용평가모형 개발·운용시도**가 **확대**

○ **기존 신용평가모형**은 **제한된 변수*** 사용, **직관적 방법론 활용** 등을 통해 **평가결과의 직관적 해석과 설명****이 **용이**한 반면,

* 연체이력, 금융거래실적, 금융부채규모 등 10개 내외의 제한된 변수만을 평가에 반영

** 연체이력, 금융거래실적 등 평가항목별 배점(예: 5년내 연체이력이 없으면 10점, 1건이면 5점 등) 부여 → 다른 평가항목이 동일하다면 1건 연체이력 해소시 신용평점 5점 상승 예상

○ **AI 신용평가모형**은 일반적으로 보다 **다양한 변수**를 **평가항목**으로 **반영***하고 있으며, **딥러닝** 등 내부구조가 복잡한 알고리즘을 활용함에 따라 **직관적 해석과 설명**이 **다소 어려움**

* A은행의 경우 AI모델에 멤버십 포인트, 휴대폰 소액결제, 통신요금 등 다양한 비금융정보 반영

⇒ AI 신용평가 **신뢰성 확보** 및 금융소비자 **권리 보호**를 위해 **객관적인 검증체계**를 **마련**할 필요

□ (**추진방안**) **개인신용평가체계 검증위원회***(신정법 제26조의3)를 통해 **AI 기반 신용평가모형 검증체계**를 마련·운영

* 신정법에 따라 CB사의 신용평가모형에 사용되는 기초정보와 평가모형의 예측력, 안정성 등에 대한 검증업무 수행중(신정원 운영)

○ 연구용역을 통해 **AI 신용평가 검증방법론을 개발**하여 하반기 중 '**AI 신용평가모형 검증업무 가이드라인***' 마련

* 학습데이터 적정성, AI 알고리즘 선정 절차, 신용평가모형 해석 및 평가결과에 대한 설명가능성 등 AI 신용평가의 주요 특징에 해당되는 영역을 검증하는 방법 등 규정

○ 同 가이드라인에 따라 **CB사 AI**를 **우선 검증**하고, **금융회사** 등의 **수요, AI 활용 현황** 등을 보아가며 **확대** 검토*

* 금융회사가 자체 신뢰성 확보 등을 위해 검증을 신청하는 경우에 한해 검증 등

3 금융분야 AI 보안성 검증체계 구축

□ (**필요성**) AI 활용시 **개인정보 유출**, 알고리즘 오작동, 학습데이터 조작 등 **위험요소 상존** ([참고] 보안위험 예시)

○ **AI 신뢰성 확보 및 유지**를 위해서는 이러한 보안위험이 **사고**로 이어지지 않고, **위험요소**를 **사전**에 **탐지·검증·해결**할 수 있도록 지원하는 **보안관리체계 마련** 필요

* EU 「인공지능법안('21.4월 발의)」: 고위험 AI 시스템 제공자는 지속적으로 AI 시스템의 오류 및 제3자의 악의적 성능변경에 대응토록 의무 부여

□ (**추진방안**) 금융회사의 **자체검증**, **금보원 검증**을 통해 AI 활용에 따른 **보안성을 확인**하는 「**금융 AI 보안성 검증체계**」 구축·운영

○ (**자체검증**) 금융회사의 자체 보안성 검증에 참조할 수 있도록 **보안성 검증 기준 및 방법** 등 「**금융분야 AI 보안 가이드라인**」 마련·배포*

* 산·학·연 AI 전문가로 AI 보안성 검증위원회를 구성·운영하여 AI 기술에 대한 최신 공격기법 등을 대응하기 위한 보안성 검증 기준 및 방법 등을 검토·반영

- **AI 전문인력 부족** 등으로 **자체 보안성 검증이 어려운 중소형 금융회사**에는 **보안성 검증 및 기술 자문** 등을 **지원**(금보원)

○ (**제3자 검증**) **금융보안원**이 「**AI 보안성 검증시스템***」 구축 → **금융회사 등 요청시** 同 시스템을 통해 **AI 알고리즘 보안성 검증**

* AI 보안성 검증에 필요한 검증 데이터셋, 검증 도구, 컴퓨팅 자원 등 제공

- **'23년부터 금융분야 챗봇서비스**에 적용된 AI 알고리즘에 대한 **검증**을 **실시**하고, **단계적 확대 추진***

* 국내외적으로 AI 보안성 검증 기술이 아직 연구개발 초기 단계인 점을 감안하여 단계적인 검증범위 확대 필요

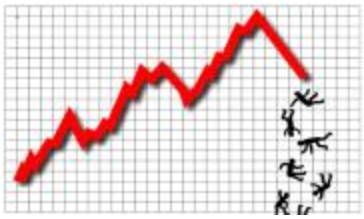
참 고 **AI 기술의 활용에 따른 보안위험 예시**

AI 챗봇 서비스 학습에 사용된 대화 메시지 유출

- 사용자 대화를 비식별 과정 없이 그대로 활용하여 데이터에 포함된 사용자 개인정보 유출
- AI 학습데이터에 대한 비식별 조치, 사용자 동의 등 프라이버시 확보를 위한 사전 검증이 필요

* 예: AI 이루다 서비스의 경우 사용자 대화를 비식별화 처리 없이 그대로 학습에 활용하여 이루다가 사용자의 개인정보(대화 속 사용자 집주소 등)를 대화에 활용

AI 주식 매매 알고리즘의 오류, 오작동에 따른 증시 폭락



- 외부 공격자가 AI 주식 매매 알고리즘의 취약점을 이용하여 AI 주식 매매 알고리즘의 오류, 오작동 유도
- 알고리즘의 대규모 오작동 발생 시, 증시 폭락 등 자본시장 전체에 큰 혼란 발생 가능

* 예: '18.2월 AI 알고리즘이 한꺼번에 매물을 쏟아내어 미국 다우지수 폭락(4%) 발생

AI 학습데이터 오염에 따른 AI 모델의 오작동 발생으로 금융사고 발생

<원본 고양이 사진>

<적대적 예제>

- 외부공격자가 AI 모델이 원본 '고양이' 사진을 '강아지' 사진으로 오인하여 잘못 분류하도록 적대적 예제*를 생성하는 등 학습데이터를 오염시켜 AI의 잘못된 학습을 유도

* AI 모델의 잘못된 학습을 유도하기 위해 원본 데이터에 노이즈를 섞어 만든 학습용 데이터로써 사람의 눈으로 원본과 차이 구별 곤란

- 금융권에 사용 중인 AI 기반 OCR(광학문자판독)도 동일한 공격으로 금융사고 발생 가능성 상존

* 예: 숫자 "7"에 사람의 눈으로는 식별하기 어려운 노이즈를 추가하여 AI가 "1"로 인식하게 유도

4 AI를 활용한 효율적 감독체계 구축

□ (**필요성**) 금융산업 내 인공지능(AI)·빅데이터 등 **신기술 활용**이 확대되면서 **금융서비스**가 **자동화**되고 **규제 환경**도 **복잡화**

○ **고도화**되는 **금융감독 수요**에 **효과적**으로 **대응**하기 위해 **AI·빅데이터**를 활용한 **섭테크**(Supervisory Technology) 도입 필요성 증대

○ **해외 금융감독기관** 역시 **디지털 감독혁신**을 위해 금융감독 업무에 **섭테크 도입**을 적극 추진* 중

* ECB(유럽중앙은행) : 전담조직(Suptech Hub)을 구성하여 디지털 문화 육성, 혁신 생태계 조성, 데이터 및 AI 활용 촉진, 자동화 프로세스 도입 등 섭테크 비전 수립·추진
Bafin(독일금감원) : 금융 디지털화에 대응하기 위해 감독정책, IT감독, 내부 디지털화 등 3가지 영역에 대한 디지털 전략을 발표·추진

⇒ 금융감독에 **AI 도입·활용**을 **확대**하여 **금융감독역량**을 **강화**하고 **변화**하는 **감독환경**에 **대응**

□ (**추진방안**) 旣 구축된 **금융감독 AI 시스템 성능**을 **고도화**하는 한편, **대상업무 확대 발굴** 등을 통해 **섭테크 혁신**을 **지속**

○ AI모델의 **정밀도 검사·평가, 새로운 데이터**를 이용한 **재학습** 등을 통해 **현재 운용중인 섭테크 시스템**의 **인식률·정확도 개선**

※ **(참고) 주요 섭테크 금융감독시스템 현황**

시스템명	주요 내용
대부업 불법추심 판별지원	음성변환(STT) 기술을 활용, 녹취파일을 문자로 자동변환시키고 키워드 검색 등을 통해 불완전판매 및 불법추심 판별을 지원
보험 TM 불완전판매 식별지원	
인터넷 불법금융광고 감시	텍스트분석(TA) 기술을 활용, 인터넷상 수집된 광고 중 미등록대부, 통장매매 등 불법행위가 의심되는 게시글을 분석하고 불법성 여부를 판별
민원분류추천	텍스트분석(TA) 기술을 활용, 민원내용을 분석하여 적합한 민원유형 및 유사민원 등을 업무담당자에게 자동 추천하여 민원처리업무의 효율 제고

○ **해외 섭테크 사례, 금융감독 수요** 및 **AI 도입·활용 타당성** 등을 고려하여 **신규 섭테크 추진과제 지속 발굴·도입**

Ⅳ. 향후 추진계획

◇ **신속한 금융권 인공지능 활성화**를 위해 **속도감 있게** 관련 과제 추진

1 (**AI 데이터 라이브러리**) **신정원**을 중심으로 'AI 데이터 라이브러리 **컨소시엄**' 구성('22.3Q) 후 **'23.2Q 중** 컨소시엄을 통해 **라이브러리 구축**

2 (**금융 빅데이터 공동 확보**) 은행연, 금보원, 금결원 등 **관계기관을 중심으로 TF**를 구성('22.3Q)하여 **'23.2Q 중 데이터 셋 구축·공유**

3 (**데이터전문기관 추가지정**) **'22.4Q** 중 **데이터전문기관 추가지정**

4 (**AI 개발·활용 안내서 발간**) **'22.8월** 중 **안내서 발간**

5 (**XAI 요건 마련**) **연구용역 발주**('22.3Q) 후 연구용역 결과에 따라 **'23.1Q** 중 **안내서 발간**

6 (**망분리클라우드 규제 개선**) **전자금융감독규정 개정**('22.8월), **클라우드 가이드라인 개정**('22.10월), 금융회사 등 **내부통제 점검** ('22.12월) 등을 거쳐 **'23.1월** 중 **시행**

7 (**금융 AI 테스트베드 구축**) **신정원, 금결원, 금보원 각 기관별로 세부 구축방안을 마련**('22.4Q)하여 **'23.4Q** 중 **구축·운영**

8 (**AI 기반 신용평가모형 검증체계**) **연구용역 결과를 바탕으로 검증 방법론을 마련**('22.3Q)하여 **'23.1Q**부터 **AI 신용평가 검증업무 시행**

9 (**AI 보안성 검증체계**) **우선 AI 보안 가이드라인을 마련**('22.4Q)하고, **'23.2Q** 중 **챗봇서비스**에 대한 **AI 보안성 검증시스템 구축·운영***

* 점차 이상거래탐지, 로보어드바이저, 상품추천, 신용평가 등 타 서비스로 확대 추진

10 (**AI 활용 감독체계**) **섭테크 금융감독시스템 지속 고도화**(상시)

< 과제별 추진일정 >

추진과제 구분		추진일	추진내용	추진기관
AI 활성화를 위한 데이터 확보 지원	AI 데이터 라이브러리 구축	'22.3Q	컨소시엄 구성	신정원 등
		'23.2Q	라이브러리 구축	
	협업을 통한 데이터 공동 확보	'22.3Q	기관별 TF 구성	은행연, 금결원, 금보원
		'23.2Q	데이터 셋 구축	
	데이터 전문기관 추가 지정	'22.4Q	전문기관 추가지정	금융위, 금감원
AI 활성화를 위한 제도 정립	AI 개발·활용 안내서 발간	'22.8월	안내서 발간	금융위, 금감원 등
	설명가능한 AI 요건 마련	'22.3Q	연구용역 발주	금융위, 금감원 등
		'23.1Q	안내서 마련	
	망분리 및 클라우드 규제 개선	'22.8월	전자금융감독규정 개정	금융위, 금감원
		'22.10월	가이드라인 개정	
		'22.12월	내부통제 점검	
		'23.1월	개정안 시행	
신뢰받는 AI 활용 환경 구축	금융 AI 테스트베드 구축	'22.4Q	기관별 방안 마련	신정원, 금결원, 금보원
		'23.4Q	테스트베드 구축	
	AI 기반 신용평가모형 검증체계 마련	'22.3Q	검증체계 마련	신정원
		'23.1Q	검증체계 시행	
	AI 보안성 검증체계 구축	'22.4Q	검증체계 마련	금보원
		'23.2Q	검증체계 시행	
	AI 활용 효율적 감독체계 구축	상시	-	금감원