

## 금융 부문의 생성형 인공지능

※ Generative Artificial Intelligence in Finance (OECD, '23.12월)

- ◇ **[생성형 AI 특징]** 생성형 AI는 학습 데이터를 기반으로 프롬프트에 반응하여 새로운 콘텐츠를 생성
  - ‘전통적인’ AI와 생성형 AI의 가장 큰 차이점은 후자가 독창적인 콘텐츠(original content)를 제작할 수 있다는 점에 있음
  - 생성형 AI는 사용자 친화성, 직관적 인터페이스, 대화적 요소 등 특징이 있어 일반인이 쉽게 접근할 수 있는 바, 대중에 빠르게 확산
- ◇ **[금융 부문의 AI 도입]** 사람의 개입 없는 완전한 자동화와 같은 금융 시장의 고급 AI 사용 사례는 대부분 개발단계에 머물러 있음
  - 실제 AI 사용 사례는 대부분 금융서비스 제공업체의 백오피스(운영)와 미들오피스(규정 준수 및 리스크 관리)의 프로세스 자동화에서 관찰
  - 금융 부문의 AI 기술 도입이 더딘 이유는 금융업이 고도의 규제를 받는 분야이고 생성형 AI 활용이 상당한 위험성을 수반하기 때문임
- ◇ **[금융 부문의 생성형 AI 위험]** 생성형 AI는 강화된 기능으로 인해 AI 관련 리스크를 악화시키고 여러 가지 새로운 문제를 추가로 제기
  - ▶ 편견, 차별, 불공정한 결과의 위험성, ▶ 설명 가능성 부족, ▶ 데이터 관련 위험(데이터 품질, 개인정보 보호 등), ▶ 모델의 견고성, 산출물의 신뢰성, 시장조작 위험성, ▶ 거버넌스 관련 위험(책임성, 투명성 부족), ▶ 금융 안정성 관련 위험(쏠림, 변동성, 상호 연결성 증가), ▶ 경쟁 관련 위험, ▶ 기타 위험(고용, 환경 등) 등의 문제 제기
- ◇ **[정책적 고려사항]** 향후 정책대응은 상황에 맞는 비례적 프레임워크 하에서 검토되어야 하며 미래지향적 관점이 필요
  - ▶ 데이터 거버넌스 강화, ▶ 편견/차별 극복을 위한 안전장치 마련, ▶ 설명 가능성 개선, ▶ 투명성 및 공개 강화, ▶ 모델 거버넌스 및 책임 메커니즘 촉진, ▶ 모델 안전성·복원력 제고, 시장조작 위험 완화, ▶ 인간 중심의 접근 장려, ▶ AI 교육 강화 및 인식 제고, ▶ 국제적·다학제적 협력 등이 필요

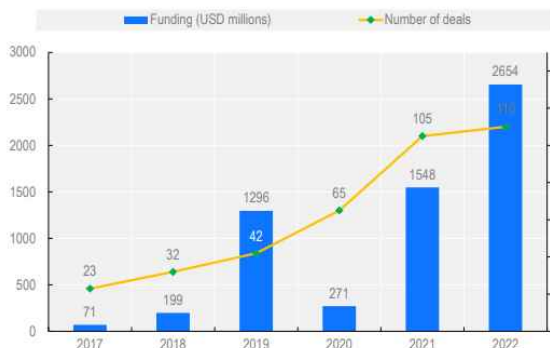
# 1. 생성형 인공지능(GenAI)과 금융

## (1) 생성형 AI의 특징

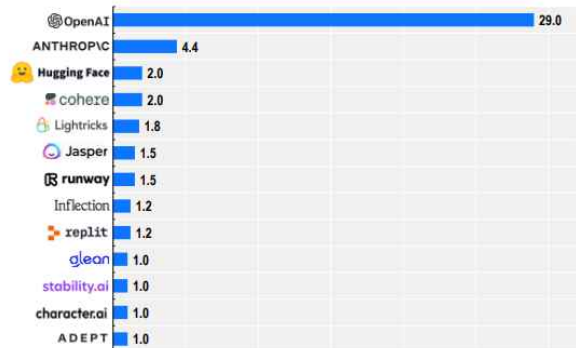
- 생성형 AI는 학습 데이터를 기반으로 프롬프트에 반응하여 새로운 콘텐츠를 생성할 수 있는 모델
  - 언어/텍스트 생성(예: ChatGPT)뿐만 아니라 시각적 출력, 오디오, 이미지, 코드 등 다른 유형의 콘텐츠 생성에도 관여
  - 생성형 AI는 사용자로부터 받은 피드백을 포함하여 방대한 양의 비정형(unstructured) 데이터 세트를 처리·학습 가능
- ‘전통적인’ AI와 생성형 AI의 가장 큰 차이점은 후자가 독창적인 콘텐츠(original content)를 제작할 수 있다는 점에 있음
  - ‘전통적인’ 모델은 주로 패턴 식별, 분류·예측에 사용되나 생성형 AI는 사람이 만든 콘텐츠와 구별하기 어려운 독창적인 결과물을 만들어 낼 수 있음
- 최근 AI의 빠른 발전은 데이터의 기하급수적 증가, 계산 능력의 빠른 발전, 저렴한 비용으로 이용 가능한 데이터 처리 능력에 의해 주도됨
  - '22년 110건의 거래에 26억 달러가 투자되는 등 민간 투자의 증가도 생성형 AI의 개발을 뒷받침

### < GenAI와 GenAI 유니콘으로 유입되는 민간 자금 규모 >

in USD million (LHS), in absolute figures (RHS)

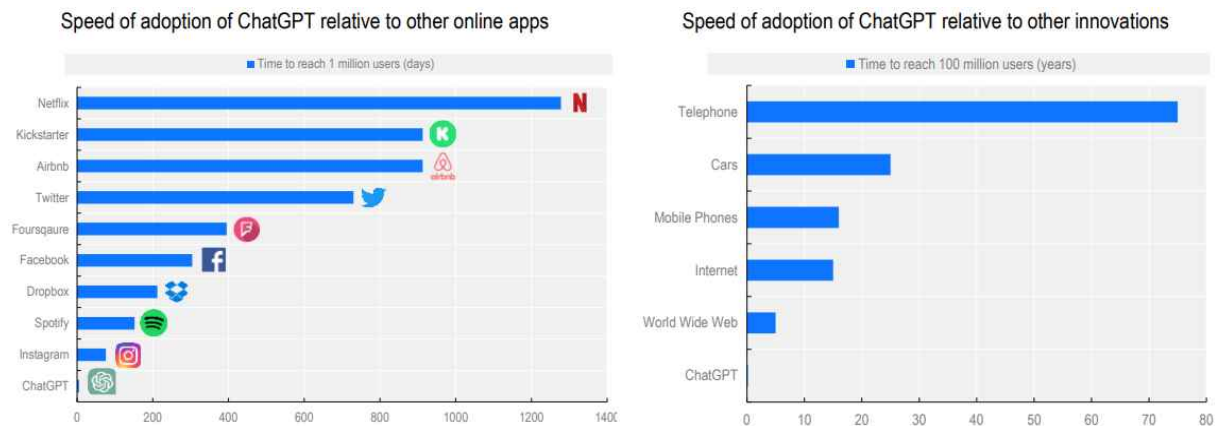


In USD bn, as of 08/05/2023



- 생성형 AI의 사용자 친화성과 직관적인 인터페이스, 인간의 인지를 모방한 거대언어모델(LLM)의 대화적 요소는 수요 측면의 도입 속도를 가속화하고 대중의 관심을 불러일으킴
- ChatGPT 등은 일반 소비자가 쉽게 접근할 수 있고 대화형이며 참여하기 쉬운 최초의 AI 모델로서 온라인 버전이 무료로 제공되면서 대중에 빠르게 확산

**< ChatGTP 도입 속도 >**



**(2) 더디게 진행되는 금융 분야의 AI 도입**

- AI에 대한 과대광고에 불구하고, 사람의 개입 없이 완전한 end-to-end 자동화를 포함하는 금융시장의 고급 AI 사용 사례는 대부분 개발 단계(테스트, 실험)에 머물러 있음
- 실제 AI 사용 사례는 대부분 금융서비스 제공업체의 백오피스(운영)와 미들오피스(규정 준수 및 리스크 관리)의 효율성·생산성 제고를 위한 프로세스 자동화에서 찾아볼 수 있음
- 트레이딩 부문은 AI가 더디게 사용되고 있는 대표적인 예에 해당
- 생성형 AI의 경우 금융시장 종사자들은 오프라인 또는 비공개로 제한된 버전의 거대언어모델을 실험 중인 바, 공개된 데이터를 기반으로 내부 프로세스 및 운영을 지원하는 도구(예: 요약, 번역)로 사용 중인 것으로 파악됨

□ 금융 분야에서 AI 기술 도입이 더딘 이유는 금융시장 활동이 고도의 규제를 받는 분야이기 때문임\*

\* 금융 분야의 규제 프레임워크는 시장의 무결성(integrity), 소비자 보호, 금융 안정성, 위험관리, 모델 거버넌스, 투명성 등에 있어 다양한 의무를 요구

○ 또한, 기만적인 결과물, 잘못된 정보, 데이터 프라이버시 침해 등 생성형 AI가 수반하는 상당한 위험성도 금융에서의 AI 활용을 저해

□ AI는 금융서비스 제공업체가 고객과 상호작용하는 데 있어 간접적인 도움을 줄 수 있음(예: 고객 프로필에 맞는 포트폴리오 권고 준비)

○ 모델이 금융상품을 직접 추천하거나 제안된 추천을 자동으로 실행하는 등 금융서비스 제공에 직접적으로 관여하는 것은 인간적인 변수를 제거하므로 매우 위험

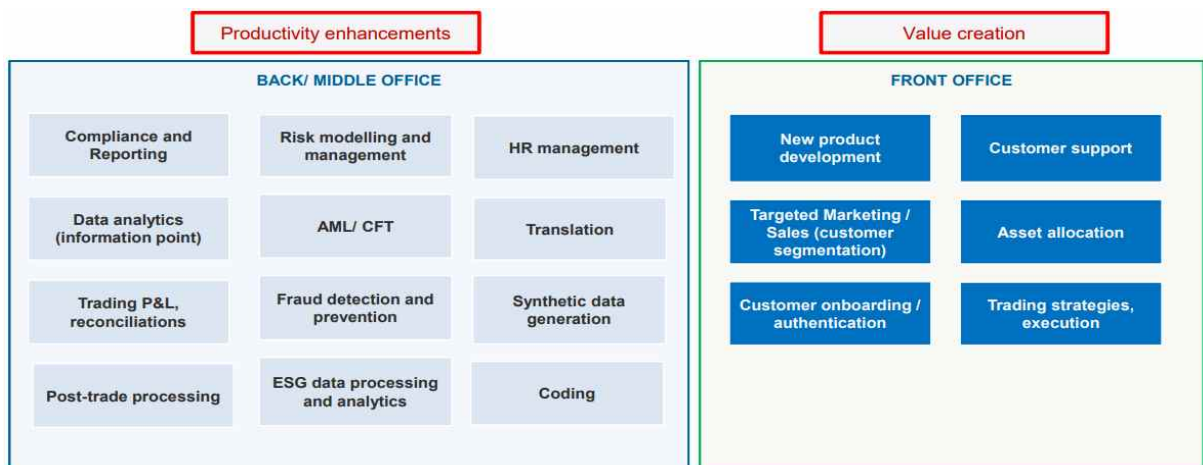
### (3) 금융 분야의 AI 및 생성형 AI 사용 사례와 혜택

□ 자산관리(종목 선정, 리스크 관리 및 운영), 알고리즘 및 초단타 매매(유동성 관리, 거래체결), 소매·기업뱅킹(온보딩, 신용도 분석, 고객 지원), 결제(자금세탁방지, 사기 탐지) 등 여러 금융업종에 걸쳐 AI 기술이 도입

○ 생성형 AI와 거대언어모델의 기능은 금융기관의 AI 기반 서비스 역량을 더욱 강화\*할 것으로 기대

\* 영업·마케팅, 고객 지원, 데이터/정보 관리 및 소프트웨어 개발을 위한 코딩 등

#### < 금융 분야의 생성형 AI 사용 사례 >

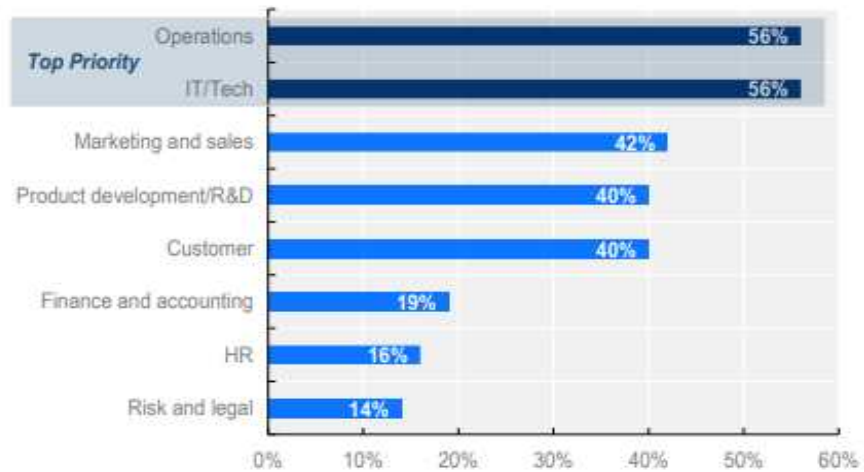


□ 자동화를 통해 생산성을 높일 수 있는 잠재력을 가진 AI 기술은 다양한 금융시장 종사자들의 **백오피스에서 사용 중**

○ **운영 최적화(operational optimization)**는 오늘날 금융 분야에서 가장 널리 보고된 AI 활용 사례 중 하나로서 **업무흐름의 효율성과 정확성을 높이고 성과를 향상시킬 수 있는 잠재력을 지님**

**< 2023년 금융 분야의 생성형 AI 활용 사례 >**

In % of industry survey respondents



**< 금융서비스 회사의 생성형 AI 적용 유형 >**

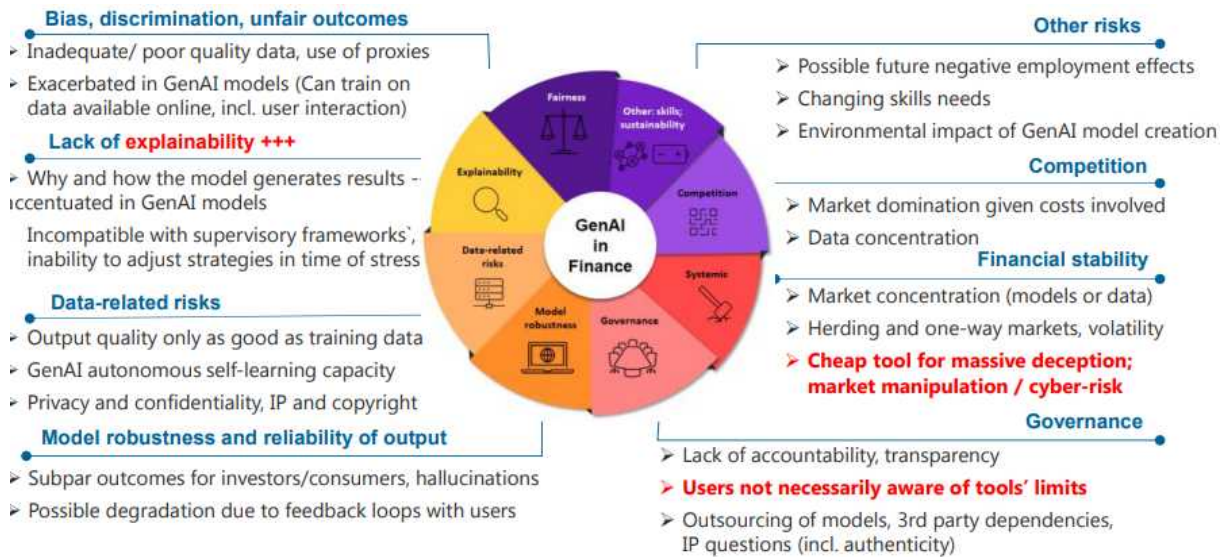
	Segment	Service	Description
<b>Goldman Sachs</b>	Corporate and Investment Banking	Code generation	ChatGPT-style AI in-house to assist developers with writing code
<b>JPMORGAN CHASE &amp; CO.</b>	Corporate and Investment Banking	Code generation	Toolkit called Senatus to facilitate the software development process through features such as code recommendations
<b>Deutsche Bank</b>	Corporate and Investment Banking	Financial analyst assistant	Testing Google's generative AI and large language models (LLMs) at scale to provide new insights to financial analysts
<b>Bloomberg</b>	Financial research	Financial assistant	Finance-specific LLM trained on Bloomberg data
<b>Brex</b>	FinTech	Expense management	ChatGPT-style CFO tools. Provides insights on corporate spend and answer critical business questions in real time
<b>البن ALBAN</b>	FinTech	Expense management	Integration of OpenAI's advanced AI technology onto its user platform to provide insights on corporate spend and answer critical business questions in real-time

⇒ 현재 생성형 AI는 ①대부분 운영 지원용으로 배포, ②생성형 AI와 고객과의 상호작용은 간접적으로 이루어짐, ③오프라인/비공개로 제한된 버전 사용

## 2. 금융 부문의 생성형 AI 위험

- 금융 분야에서 AI의 광범위한 도입은 금융시장에 이미 존재하는 위험을 증폭시키는 동시에 새로운 도전과 위험을 야기할 수 있음
  - 생성형 AI는 강화된 기능으로 인해 AI 관련 리스크를 악화시키고 여러 가지 새로운 문제를 추가로 제기

### < 생성형 AI의 위험과 도전요인 >



#### (1) 편견, 차별, 불공정한 결과

- AI 기반 모델은 모델 학습에 저품질 또는 부적절한 데이터 세트 사용 시 의도치 않게 **편향되거나 차별적인 결과를 생성할 수 있음**
  - 특히, 생성형 AI의 경우 온라인에서 사용 가능한 거의 모든 데이터로 자기 학습할 수 있는 바, 학습에 사용된 데이터에 이미 존재하는 **편견을 강화·악화시켜 웹에 반영된 역사적 편견을 영속화할 가능성**
  - 생성형 AI 모델이 **사용자의 피드백과 상호작용을 통해 학습할 수 있다**는 사실은 **사용자의 프로필에 따라 편견과 사회적 불평등을 심화시킬 수 있는 위험을 증가시킴**
- 금융에서 차별과 편견의 위험을 보여주는 대표적인 예는 **신용 배분과 AI 기반 모델에 기반한 차별적 또는 불공정한 대출 가능성**임

## (2) 설명 가능성 부족

- AI 기반 모델이 결과를 생성하는 이유와 방법을 이해하기 어렵다는 점, 즉 설명 가능성은 금융 분야에서 중요한 위험을 초래
  - AI 모델의 발전은 비선형·다차원 AI 모델의 복잡성 증가와 인간의 인지 수준에 맞는 추론 및 해석 요구 사이의 불일치를 더욱 심화
  - 금융에 사용되는 AI 모델의 설명력이 부족하면 실적이 부진할 때 투자 또는 트레이딩 전략을 수정하기 어려움\*
    - \* 추적 가능한 선형관계가 없거나 모델의 결과를 이끈 매개변수를 모르기 때문
  - 대출의 경우 소비자는 불공정한 신용 결정을 식별하고 이의를 제기할 수 있는 능력이 제한됨
- 그간 AI 기반 모델의 설명 가능성에 진전이 있었으나 급속한 AI 개발 속도에 뒤처지는 모습
  - 설명 가능성 수준이 제한적일 경우 고객과 시장 참여자의 AI 기반 금융서비스 제공에 대한 신뢰도가 낮아질 우려

## (3) 데이터 관련 위험

- 모델 산출물의 품질은 모델 학습에 사용된 데이터에 좌우되며 모든 유형의 모델은 'garbage in, garbage out' 위험이 존재
  - AI가 학습할 데이터 세트의 선택은 모델의 산출물에 보이지 않는 막대한 영향을 미치며 생성형 AI의 경우에는 사용자가 모델과의 상호작용에서 모델에 입력하는 데이터(프롬프트)도 산출물에 영향\*
    - \* 생성형 AI는 자율적인 자가 학습 능력 보유(unsupervised learning)
- 향후 모든 온라인 콘텐츠에 접근할 수 있는 생성형 AI 버전이 널리 도입되면 개인 정보 및 기밀성 침해 위험이 증가할 가능성
  - 사용자가 프롬프트에 제공하는 데이터가 구체적일수록 개인 정보 또는 기밀 데이터 유출의 위험이 높아짐

- 사이버 보안 위협, 해킹 위험도 존재하는 바, 사기꾼은 생성형 AI 기능을 최대한 활용하여 맞춤형 개별 사기 공격을 수행할 수 있음\*

\* 생성형 AI는 악의적인 행위자에게 소셜 엔지니어링, 이메일 피싱, 문자 메시지 스미싱 공격을 수행하는 세련된 방법 제공

#### (4) 모델의 견고성 및 산출물의 신뢰성

- AI 기반 모델 결과물의 신뢰성·정확성이 낮으면 관련된 모든 유형의 금융 사용 사례에서 기대 이하의 결과가 발생(예: 부적절한 투자 자문)

- 생성형 AI 모델에서는 사용자와의 상호작용 및 자율적 자가 학습 모델의 피드백 루프가 모델의 정확도 저하를 야기할 수 있음\*

\* 실증분석에 따르면 동일한 LLM 모델의 산출물이 비교적 짧은 시간 내에 크게 변화

- 생성형 AI는 산출물의 질·신뢰성과 관련하여 기만과 허위 정보·오류 위험을 수반하는 바, 금융서비스 제공을 위해 구축된 모델이 잘못된 정보와 조언을 제공할 경우 개인투자자에게 부정적 영향을 미칠 우려

- 악의적 행위자는 대규모로 생성형 AI 기능을 사용하여 금융시장 전반을 의도적으로 조작할 수 있음

- 금융기관이 사용하는 생성형 AI에 웹(예: 소셜 미디어)의 정보를 실시간으로 지속 공급할 경우 시장조작의 위험은 더욱 커질 우려\*

\* 현재는 웹과 상호작용하지 않고 오프라인 또는 프라이빗 클라우드 환경에서 특별히 제한된 버전의 생성형 AI 모델을 사용하고 있어 시장 조작 상황은 발생하지 않음

#### (5) 거버넌스 관련 위험

- 고급 생성형 AI 모델의 책임성·투명성 제고를 위해서는 모델과 모델에 사용된 데이터에 대한 광범위한 정보 공개\*가 필요

\* 예: 데이터 소스, 모델의 한계, 예측가능한 위험과 위험관리를 위해 취한 조치 등

- 그러나 생성형 AI는 높은 투명성 수준을 달성하기 어려울 수 있음\*

\* 예: 학습 데이터에 인터넷에서 선별한 비정형 정보가 포함된 경우 저작권에 대한 정보를 공개하기 어려울 가능성



- 생성형 AI 솔루션이 금융서비스 업계 전반에 걸쳐 보편화될 경우 관련 지식이 부족한 금융기관 종사자는 생성형 AI 기술을 사용 중이라는 사실을 모른 채 이러한 도구를 사용할 수 있음
  - AI 도구의 사용 위험에 대한 실무자의 인식이 부족하면 최종적으로 고객이 중요한 위험에 노출될 가능성이 커짐
- AI 모델 관련 서비스 및 인프라를 제3자에게 아웃소싱할 경우에는 모델의 부정적 결과에 대해 누가 책임을 져야 하는지 불분명

#### (6) 금융 안정성 관련 위험

- 금융에 AI를 도입하면 일방향 시장, 시장 유동성 및 변동성, 상호 연결성 및 시장 집중과 관련된 잠재적 금융 안정성 위험이 발생 가능
  - 예를 들어, 트레이딩에 AI가 도입되면 대규모 매도/매수가 동시에 실행되어 시장 변동성이 커지고 스트레스 상황에서 유동성 부족과 급격한 폭락을 초래할 수 있음
  - 또한, 시장의 상호연결성이 예상치 못한 방식으로 높아지고 이전에는 상관관계가 없던 변수들의 상관관계와 의존성이 증가할 수 있음
  - 소수의 모델 제공자 또는 사용자 인터페이스 제공자에 의한 시장 지배 위험이 가중될 뿐만 아니라 지배적 사업자의 운영 실패 시 시장에 시스템적 영향이 발생할 가능성도 존재

#### (7) 경쟁 관련 위험

- 생성형 AI 모델의 경우 시장 지배력과 관련된 경쟁 문제가 존재
  - 생성형 AI 모델을 개발·훈련하는 데 필요한 막대한 양의 컴퓨팅 파워와 데이터를 고려할 때, 선발주자 우위를 점하거나 모델의 설계, 훈련, 유지보수를 수행할 수 있는 자원을 보유한 소수의 플레이어에게 활동이 집중될 수 있음
- 데이터 집중은 데이터 세트에 보다 저렴하고 쉽게 접근할 수 있는 기존 업체(예: 소셜 플랫폼)의 지배력을 강화할 수 있는 또 다른 위험요인

- 데이터 접근은 거대언어모델과 같은 AI 모델의 성공에 매우 중요한 바, 빅테크 등에 의한 데이터 집중은 소수기업의 지배 위험을 악화시킬 우려

## (8) 기타 위험

- AI의 사용은 고부가가치 업무에 사용 가능한 리소스를 확보할 수 있으나 동시에 향후 고용시장에 상당한 영향을 미칠 위험이 있음\*

\* 다만 현재까지는 AI가 고용에 상당한 부정적 영향을 미친다는 증거가 거의 없음

- 특히, 생성형 AI는 금융 분야의 다양한 백오피스 및 미들오피스 기능을 자동화할 수 있는 혁신적 잠재력을 가지고 있는 바, 향후 금융업계에서 AI 확산에 따른 고용 문제가 발생할 가능성

- AI 시스템의 컴퓨팅 수요가 증가하면서 지속가능성에 대한 우려가 확산

- 생성형 AI와 거대언어학습은 학습을 위해 막대한 양의 컴퓨팅을 필요로 하며 이와 관련한 많은 에너지 소비와 환경적 영향에 대해서는 추가적인 검토가 필요\*

\* AI가 환경에 미치는 영향에 대한 신뢰할 수 있는 데이터는 아직 제한적

## 3. 정책적 고려사항

### (1) 기본방향

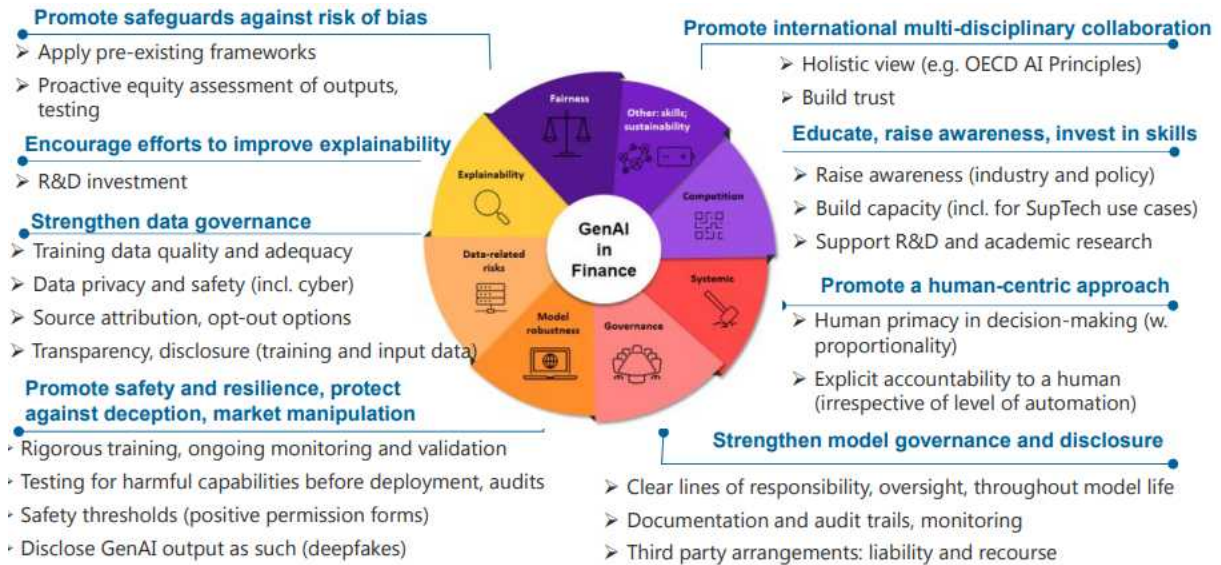
- 향후 정책적 대응은 사안의 중요성, 소비자에게 미치는 영향 등을 감안한 상황에 맞는 비례적인 프레임워크(contextual and proportional framework) 하에서 검토되어야 함

- 또한, AI 분야의 빠른 발전속도를 고려할 때 모든 지침이나 정책은 시간이 지나도 유효할 수 있도록 미래지향적이어야 함

- 금융 규제·감독 당국은 생성형 AI가 금융 안정성 확보, 금융소비자 보호, 시장 건전성 및 공정 경쟁 촉진 등 정책목표에 부합하는 방식으로 활용되도록 보장하는 역할을 수행해야 함

## (2) 정책 대응

### < 금융 분야의 생성형 AI 위험 완화를 위한 정책 대응 >



□ **[모델 개발자/배포자의 데이터 거버넌스 강화]** 데이터의 품질, 사용 목적에 맞는 적절성, 프라이버시, 신뢰성 및 출처/저작권 등을 포함하는 데이터 관리 및 거버넌스 관행에 대한 모범사례를 제시할 필요

○ 개인 데이터가 사용되는 경우 소비자는 자신의 데이터가 생성형 AI 모델 학습에 사용되는 것을 거부할 수 있는 권리가 있어야 함

□ **[편견/차별 극복을 위한 안전장치 마련]** 생성형 AI 모델을 배포하는 기업은 사전 형평성 평가, 모델 결과물에 대한 영향 평가 등을 통해 금융서비스의 기존 공정성 프레임워크가 계속 적용되도록 해야 함

○ 모델에서 사용되는 변수와 데이터를 대표성(representativeness) 관점에서 적절한지 검증하는 것은 잠재적 편향을 줄일 수 있는 방안 중 하나

□ **[설명 가능성 개선을 위한 노력 장려]** 비교적 단순한 머신러닝 모델에서 달성한 설명 가능성 관련 진전은 훨씬 더 복잡하고 설명 가능성이 낮은 생성형 AI 모델에서도 추구되어야 함

○ 설명 가능성 수준의 향상은 금융 분야에서 AI에 대한 신뢰를 구축하는 데 있어 핵심적 사안

- **[투명성 및 공개 강화]** 소비자는 금융서비스 제공에 AI 기술이 사용 되는지, 자신들의 데이터가 모델 개발을 위해 수집·활용되는지 등에 대한 정보를 제공받아야 함
  - 또한, AI 시스템의 성능과 한계, 한계점을 해결하기 위해 취해진 조치 등에 대한 정보가 알기 쉬운 언어로 소비자에게 제공되어야 함
  
- **[모델 거버넌스 및 책임 메커니즘 촉진]** 모델 개발 및 배포에 관여하는 당사자는 모델이 제대로 작동하는지에 대해 책임을 져야 함
  - 거버넌스 프레임워크 강화를 위해 모델 수명주기 전반에 걸친 **명확한 책임 및 감독 규정, 최소 요건·모범사례 지침** 제공이 필요
  
- **[모델 안전성 및 복원력 제고, 사기·시장조작 위험 완화]** 지속적이고 엄격한 모니터링, 검증을 통해 모델 복원력을 높이고 드리프트를 방지·해결하는 한편 모델이 의도한 대로 작동되도록 해야 함\*
  - \* 자율적 비지도 학습을 기반으로 하는 생성형 AI는 허위, 부정확한 정보가 모델에 영향을 미치므로 지속적인 모니터링이 특히 중요
  - 모델 모니터링 및 검증에는 **독립적인 검토와 외부 감사**가 포함될 수 있으며 **프로세스를 문서화**할 경우 감독을 용이하게 할 수 있음
  - 모델을 배포하기 전에 **적대적인 방식으로 모델이 활용되지 않도록 필터링 및 기타 제한을 도입**(예: 유해한 요청에 대한 거부)할 필요
  
- **[인간 중심의 접근 장려]** 대출 등에 있어 **고객 피해 최소화**를 위해 AI 기반 금융시장 활동에 **적절한 수준의 인간 개입(human involvement)**이 필요
  - 고객은 서비스 제공에 AI가 관여하고 있다는 사실을 알아야 하며 AI 사용을 거부하고 **인간에 의한 서비스 제공을 요청**할 수 있어야 함
  - 기업도 모델 중단(model disruption) 시 업무 연속성이 보장되도록 AI 기반 시스템에 대한 **과도한 의존은 지양**해야 함
  - **인간을 계속 참여(keeping human in the loop)**시키는 것은 금융 분야에서 AI 사용에 대한 **확신과 신뢰를 구축**하는 데에 기여

- **[AI 교육 강화 및 인식 제고]** AI 발전에 발맞추어 R&D, 기술 및 역량에 투자하는 동시에 생성형 AI의 위험에 대한 인식을 높여야 함
  - 생성형 AI 모델의 위험(예: 환각)을 완화하기 위한 R&D 투자가 필요하며, 금융업계 종사자 및 소비자에 대한 교육을 통해 AI 모델의 거버넌스를 강화하고 모델의 안전한 활용을 도모할 필요
  - 정부 관계자도 금융 분야의 AI 활동을 감독하고 필요시 개입할 수 있도록 기술적 능력을 갖추어야 함\*
  - \* 정책 당국의 AI 숙련도가 높아질 경우 RegTech/SupTech 솔루션을 활용한 보다 효과적인 금융시장 감독이 가능
  
- **[국제적·다학제적 협력 도모]** OECD AI 원칙에 따라 국제적·다학제적 및 다중 이해관계자 간 협력을 촉진할 필요
  - 경제·사회 전반에 걸친 AI의 혁신적 효과를 감안할 때, 신뢰할 수 있고 안전한 AI 배포를 위한 종합적 관점과 목표가 필요
  - 금융상품·서비스는 글로벌 차원에서 제공되므로 AI 위험관리 및 책임 프레임워크 관련 정책 일관성과 상호 운용성(interoperability)을 확보하는 것도 중요