

생성형 AI 기술의 금융권 적용과 보안 이슈

- 국내·외 생성형 AI 가이드라인을 중심으로 분석 -

김병민 · 김은지 · 김지민 · 문범수 · 안세은 · 이은진 · 황혜경

요 약

최근 생성형 AI가 다양한 분야에서 활용되면서 새로운 보안 위협의 중요성이 주목받고 있다. 본 연구는 금융권에서 AI 활용 시 발생하는 리스크와 해결 방안을 기술적, 규제적 측면에서 분석하였다. 현재 우리나라는 생성형 AI 관련 규제 및 가이드라인을 준비 중이기에 주요 국가들의 규제 및 가이드라인을 분석하여 국제적 추세와 안전한 활용 방안을 도출하였다. 특히 AI 도입 및 활용이 초기 단계에 있는 금융 부문에 중점을 두어 보안 위협과 대표적인 해결 방안을 심도 있게 분석하였다. 아울러, 주요 해외 국가들의 AI 보안 및 리스크 관리 정책을 면밀히 검토하여, 이를 통해 우리나라가 참고할 수 있는 시사점을 도출하였다. 또한, 여러 국가의 AI 규제와 국내 가이드라인을 비교·분석함으로써, 보안 위협을 최소화하면서 기술 발전을 극대화할 수 있는 금융권에서의 생성형 AI 활용 규제 방향을 제시하였다.

목 차

I. 서론	3
II. 본론	5
1. 금융권 보안 위협	5
2. 금융권 생성형 AI 보안 위협 및 대응방안	5
3. 국내 AI 보안 가이드라인 주요 내용	9
4. 생성형 AI 국외 규제 동향	15
5. EU, 미국, 중국 AI 정책의 시사점	25
III. 결론 및 제언	30
(부록) 체크리스트	31

I. 서론

인공지능은 최근 산업 전반적으로 다양한 목적을 위해 사용되고 있으며, 그 적용 범위와 영향력은 빠르게 확산하고 있다. 금융분야는 세계 인공지능 시장에서 약 20% 이상의 규모를 차지하고 있으며, 이후 2026년까지 연평균 약 36% 성장할 것으로 전망한다 [1]. 국내 금융분야에서 역시 디지털 금융 혁신을 목적으로 AI의 적극적인 도입을 고려하고 있다.

이러한 추세에 맞춰 EU, 미국 및 인공지능 산업을 주도하는 주요 국가들은 법·제도를 정비하여 인공지능 금융 혁신을 활성화하고 있지만, 국내의 법·제도는 여전히 금융분야의 인공지능 도입에 많은 제약을 두고 있다 [2]. 이는 금융기관이 고객의 개인정보와 기업 비밀, 입출금 정보 등 매우 높은 기밀성을 요구하는 정보를 취급하고 있으며, 이러한 정보를 인공지능에 사용할 때 정보 유출 등 여러 가지 리스크가 따르기 때문이다. 하지만 AI 기술은 업무의 효율성을 높이고 더욱 경쟁력 있는 서비스를 제공하기 때문에, 금융분야에서도 적극적인 AI 기술의 도입이 필요하다.

BoA(Bank of America)의 Erica는 기존 은행과 AI 기술 접목의 좋은 사례로 꼽히고 있다. 기존 상담원이 전담하던 업무와 더불어 투자 자문, 재무 설계 서비스까지 제공하며 기존 BoA가 보유하고 있던 빅데이터를 활용함으로써 편리함과 정확도를 갖추고 있다 [3].

NH농협은행 또한 최근 전국 1,103개 모든 영업점에 AI 은행원을 배치했다. AI 은행원은 투자상품 판매를 위한 필수적인 상품 설명을 보조하는 역할을 맡고 있다. 또한, 업무를 확대해 영업점뿐만 아니라 모바일이나 디지털 데스크와 같은 온라인으로 활용 범위를 넓히는 것을 검토 중이라 밝혔다. NH농협은행뿐만 아니라 다른 시중은행들도 기존의 챗봇 수준에서 벗어나 실제 고객 응대가 가능한 수준의 AI 은행원 서비스를 적극 활용하여 대면 서비스의 질을 높이고 있다 [4].

이처럼 금융권에서 인공지능 기술의 도입은 금융 서비스의 혁신과 기업의 효율성을 극대화하는 측면에 있어 중요한 역할을 하고 있다. 그러나 이러한 기술의 도입이 항상 긍정적인 결과만을 가져오는 것은 아니다. 특히 AI 활용 시 발생할 수 있는 보안 위협은 금융권에서 매우 중요한 이슈로 제기되고 있다. AI 시스템은 대량의 데이터를 활용해 처리하고 분석하는 과정에서 데이터 유출, 데이터 오염, 환각 등의 보안 위협이 발생할 수 있다.

위에서 언급한 보안 위협들은 AI를 활용하는 금융기관과 기업의 신뢰도와 평판을 크게 하락시키는 주요 원인이 될 것이며, 신뢰를 바탕으로 이루어지는 금융 산업에

심각한 영향을 미칠 것이다. AI 기술의 빠른 발전에 따라 발생할 보안 위협은 단순히 기술적인 문제를 넘어서 복합적인 이슈로 나아가고 있다. 이와 같은 종합적인 보안 위협을 인지한 세계 각국의 규제 기관들은 AI의 활용에 따른 리스크를 최소화하기 위해 다양한 규제와 지침을 마련하고 있다. 이는 금융기관들이 AI를 활용할 때 발생할 수 있는 보안 위협을 효과적으로 관리하고 대응하는 데 중요한 역할을 할 것으로 보인다.

이에 본 논문 제2장에서는 금융보안원이 제시한 주요 보안 위협을 중심으로 금융권이 마주한 보안 이슈들을 알아보고, 금융권에서 AI를 활용할 때 발생할 수 있는 보안 위협을 분석하며, 이러한 위협을 예방하고 대응할 방안을 제시하고자 한다. 추가로 현재 진행 중인 국내외 규제 동향 및 가이드라인을 살펴본다. 이후 제3장에서는 분석한 내용을 바탕으로 금융기관들이 AI 기술을 안전하게 활용할 수 있도록 하는 방안을 모색할 것이다.

II. 본론

1. 금융권 보안 위협

코로나19 팬데믹, 통신 기술의 발전 등에 따라 재택근무, 클라우드 활용 등 혼재된 업무 형태가 증가하였고, 태블릿 기기, 스마트폰 등 업무 기기가 다양해짐에 따라 공격 대상 및 범위도 방대해졌다. 따라서, 금융보안원에서는 온라인과 오프라인 연계를 바탕으로 공격 주체의 노출을 최소화하는 하이브리드 위협이 더욱 고도화될 것이라고 예상하고 있다.

또한, S/W 공급망 공격이 성행되며 SBOM의 중요성이 강조되고 있다. 최근 금융권에서는 오픈소스를 활용한 S/W 개발이 일반화되고 있다. 하지만 공격자가 S/W 개발용 패키지 관리자(NPM)¹⁾에 금융회사 업무와 관련된 악성 패키지를 업로드하고, 이를 다운로드 받아 실행하여 업무 권한을 탈취당한 사례가 있다 [5]. 이처럼 공격자는 오픈소스 저장소에 악성코드를 유입시키거나, S/W 개발 회사를 해킹하여 정상적인 S/W를 변조하는 등 다양한 형태로 S/W 공급망 공격을 시도하고 있다. 따라서 S/W 공급망 공격에 대한 예방책으로 SBOM 활용을 본격화할 것으로 예상된다.

인공지능의 기술과 성능이 빠르게 향상됨에 따라 딥페이크 기술도 함께 성장하였다. 최근에는 사진 한 장만으로도 완성도가 높은 딥페이크 영상을 만들 수 있으며, 3초 분량의 음성으로도 개인의 목소리를 구현해 낼 수 있게 되었다. 그러나 이러한 발전은 딥페이크 사기의 증가로도 이어졌고, 실제로 작년에는 생체인증을 우회하기 위한 딥페이크 시도가 전년 대비 3,000% 증가한 것으로 나타났다. 홍콩에서는 딥페이크를 이용한 사칭 범죄로 약 340억 원이 송금된 사례도 있었다 [6]. 이와 같은 딥페이크 사기 위협의 증가 추세에 비해, 딥페이크 예방 기술의 발전은 더딘 편이다. 여러 가지 혁신적인 딥페이크 탐지 기술들이 개발되고 있고, 국가 차원에서도 딥페이크 활용 선거운동 금지 등의 딥페이크 피해 예방 움직임이 있지만 아직 사용자 수준에서 체감할 수 있는 딥페이크 예방 기술은 상용화되지 않은 것이 현실이다.

2. 금융권 생성형 AI 보안 위협 및 대응방안

(1) 데이터 유출

현재 상용화된 생성형 AI는 방대한 양의 데이터를 기반으로 학습 및 작동되며, 금융권에 도입될 생성형 AI 또한 방대한 양의 금융 데이터를 사용하게 될 것이다. 하

1) NPM(Node Package Manager) : 자바스크립트 프로그래밍 언어를 위한 패키지 관리자로 공개용 패키지 및 개인패키지 DB로 구성

지만 현재 사용되는 생성형 AI는 데이터 유출에서 완전히 자유롭지 못하다. 23년 7월에는 OpenAI의 ChatGPT에서 개인정보 유출 사고가 발생하였고, 23년 3월에는 ChatGPT를 허용한 삼성전자의 기업 내부정보가 유출되는 사고도 있었다. 이러한 사례를 봤을 때, 금융권의 생성형 AI 도입은 더욱 신중해질 수밖에 없다. 금융권은 고객들의 개인정보 및 신용정보, 기업 간의 거래 정보와 같은 민감 데이터를 다루고 있다. 이러한 정보가 유출될 때 금융권에 심각한 피해를 초래할 수 있다. 따라서 이러한 데이터를 생성형 AI에 사용할 경우, 정보 유출 방지가 필요하다. 한창 이슈가 되었던 '이루다' 서비스의 경우, 사용자 대화를 비식별 처리 없이 그대로 학습에 활용하다가 사용자의 개인정보를 대화에 활용하는 문제가 있었다. 이처럼 생성형 AI가 금융권에서 사용된다면 AI 서비스 학습에 사용된 데이터 유출 위험이 있다. 이에 대한 해결 방안으로는 '개인정보 비식별화'가 있다. 개인정보 비식별화는 개인정보 및 민감정보를 제거 혹은 마스킹하여 데이터를 보호하는 동시에 분석·연구에 활용할 수 있게 하는 기술이다. 하지만 개개인의 데이터가 비식별화되어서 취급 및 저장이 되더라도 데이터가 담고 있는 정보를 통해 개인의 신원을 역추적해 특정이 가능할 수 있어 데이터 유출 방지가 중요하다. 따라서 금융권 생성형 AI 서비스 도입 시 학습데이터에 대한 유출 방지, 비식별 조치, 사용자 동의 등 프라이버시 확보를 위한 철저한 사전 검증이 필요하다.

(2) 데이터 오염 및 회피 공격

AI 시스템의 데이터 관리와 관련하여 데이터 유출 등의 문제뿐만 아니라, 유입 데이터를 이용한 공격의 위험도 커질 수 있다. 유입 데이터를 이용한 공격은 대표적으로 데이터 오염(Poisoning)과 회피 공격(Evasion attack)이 있다.

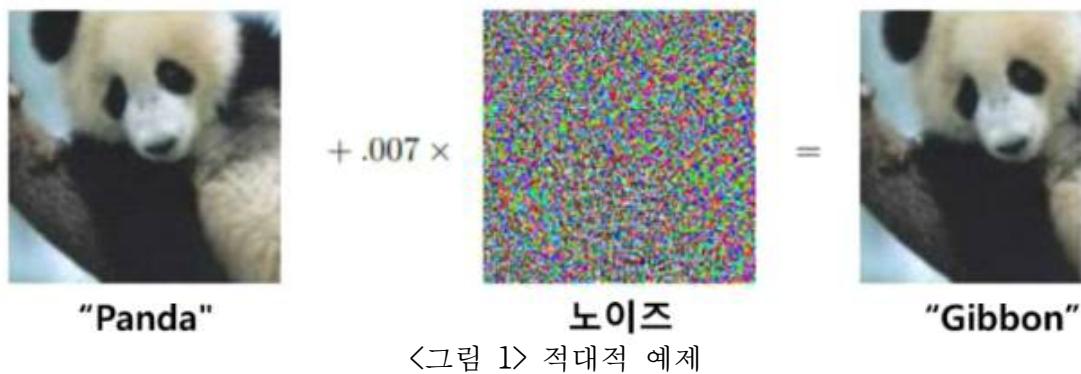
생성형 AI의 경우, 잘못된 데이터가 입력될 때 학습 데이터의 내용이나 데이터를 프로세스 하는 알고리즘, 출력을 해석하고 적용하는 인식 과정에서 보편성과 공정성이 훼손될 가능성이 있다. 또한, 시스템과 보안의 취약점을 악용해 AI 모델을 무력화하거나 출력을 조작하기 위해 외부에서 편향된 데이터를 주입할 가능성도 존재한다. 이러한 공격을 '데이터 오염'이라 하며, 이용되고 있는 AI 모델의 구조와 사이버보안 프로토콜에 대해 알고 있는 관계자가 데이터 오염을 시도할 경우 공격은 수월하게 이루어질 수 있고 피해가 커질 수 있다. 데이터 오염은 적대적 예제¹⁾

1) 적대적 예제 : AI 모델의 잘못된 학습을 유도하기 위해 원본 데이터에 노이즈를 섞어 만든 학습용 데이터로써 사람의 눈으로 원본과 차이 구별이 되지 않는다.

를 사용하여 모델의 오작동을 유도하거나, 엉뚱한 데이터를 주입해 성능저하를 일으킬 수도 있다 [8].

데이터 오염의 대표적 해결 방안으로는 임계치를 설정하고, 임계치를 벗어나는 데이터는 학습에서 배제하는 방법이 있다. 혹은, 클러스팅 기법을 사용하여 유효한 집합만 학습에 사용하는 방법도 있다.

회피 공격은 데이터 오염과 비슷하게 입력 데이터를 악의적으로 조작하는 공격이지만, 데이터 오염과 달리 완성된 모델을 그 목표로 한다. 회피 공격의 경우 악의적으로 조작된 적대적 예제나, AI가 판단하기에 모호한 사진을 입력하여 잘못된 답을 유도하는 공격이다. 이러한 공격은 인공지능의 답변을 예상 못 하게 하거나, 통제 불능으로 만들 수 있는 위험이 있다.



위 사진은 인공지능 모델을 기만하기 위한 회피 공격의 ‘적대적 예제’의 예시이다. 자세히 보면 우측 사진에는 미세한 노이즈가 추가되어 있다. 위와 같이 외부 공격자가 AI 모델이 원본 ‘판다’ 사진을 ‘긴팔원숭이’ 사진으로 오인하여 잘못 분류하도록 적대적 예제를 생성한다면 AI의 잘못된 답변을 유도할 수 있다.

회피 공격을 해결하는 방안으로 적대적 예제 학습, 경사도 마스킹, 입력값 변화 등이 있다. 첫 번째로 적대적 예제를 사전에 생성하여 학습함으로써, 적대적 예제를 통한 회피 공격을 예방하고 차단하는 방법이 있다. 두 번째로는 경사도를 공개하지 않거나 마스킹, 범주화 등의 기법을 활용하여 모델의 정보 노출을 최소화함으로써 회피 공격을 효과적으로 예방할 수 있다. 마지막으로, 입력값 변화 방법이 있다. 이 미지가 입력값이면 회전, 패딩 등의 임의적인 변환으로 모델의 강건성을 높일 수 있다 [8]. 금융기관의 의사결정은 개인이나 기업 고객의 재무에 결정적인 영향을 미칠 수 있으며, 금융시장을 교란하는 파급 효과를 초래할 수 있다. 이에 따라 이해 당사자들

이 데이터 처리 과정에 개입하여 결과를 조정하려는 인센티브가 존재한다. 더불어 AI 모델에 주입된 데이터와 아웃풋의 정확한 인과관계를 설명할 수 없는 AI의 특성도 데이터 편향이나 데이터 오염 등의 공격에 대한 정확한 진단과 교정을 어렵게 하는 요인으로 볼 수 있다.

(3) 환각 현상

입력 데이터의 문제가 없더라도 환각 현상으로 인해 잘못된 산출물이 발생할 위험이 있다. 환각 현상이란 주어진 질문에 대해 사실이 아닌 그럴듯한 거짓된 답변을 꾸며내는 것으로, 생성형 AI를 활용하는 데 실질적인 피해와 신뢰 저하의 문제를 발생시킨다.

생성형 AI의 환각 문제는 금융분야에서 여전히 리스크 요소로 작용하고 있다. RAG등을 사용하여 오류를 줄일 방법이 있지만, 이는 완벽한 해결책이 아니다. 하지만 LLM 기술의 지속적인 발전으로 환각 문제는 점차 개선되고 있다. 즉, 금융분야에서 생성형 AI 도입이 필수적이지만 정확성과 신뢰성 확보를 위한 지속적인 기술 개선과 대응 방안 마련이 필요한 상황이라고 볼 수 있다.

생성형 AI는 작동 원리상 학습을 바탕으로 새로운 결과를 생성하고, 대규모 데이터를 압축하며, 자연스러운 답변을 생성하기 위해 결과를 취사선택하는 성질이 있어 근본적으로 환각을 완전히 제거할 수 없다. 양질의 데이터를 대량으로 학습한다 해도 기존의 데이터와 일치하지 않는 질문이 발생할 수 있고, 저장공간의 한계로 데이터를 압축·복구하는 과정에서 유실이 발생하여 사실과 유사하지만, 거짓된 답변이 생성되는 것은 불가피하다.

금융 시장에 구조적인 변화나 예기치 못한 상황이 발생할 때 생성형 AI가 환각 현상을 일으킬 수 있다. 이는 금융기관에 신뢰 훼손과 같은 피해를 줄 가능성성이 있다. 환각 현상을 해결하는 방안으로는 검색증강생성(RAG)과 휴먼인더루프(HITL)¹⁾가 있다.

Retrieval-Augmented Generation(RAG)은 인공지능의 사전 학습 데이터만을 사용하는 것이 아닌 정보 검색을 통한 데이터를 사용하는 방법이다. 이는 AI가 학습하지 않은 내용을 답변하거나, 학습 데이터에 포함되지 않은 최신 자료들도 답변할 수 있도록 한다. Human-In-The-Loop (HITL)은 인간이 AI의 학습 데이터를 직접 검증 및 수정할 수 있도록 하는 접근 방식이다. 미국의 국가 보건 서비스를 예시로,

1) 휴먼인더루프(HITL) : AI를 활용해 신뢰도가 낮거나 중요한 의사결정을 포함하는 데이터에 대해서는 인간이 직접적으로 개입해 데이터를 수정하여 답변의 품질을 높이는 방법이다.

Amazon A21을 사용한 HITL 프로세스를 통해 매달 5,400만 건에 달하는 데이터를 인간이 직접 개입하여 수정할 수 있도록 하였다. 결과적으로 AI의 예측이 초래하는 위험을 감소시켰으며 더 효율적인 보험금 지급이 가능할 수 있었다 [9].

이는 AI를 활용하는 기업으로서 예상하지 못하고 이해하기 어려운 답변에 대한 환각 현상을 예방하고 사전에 처리할 기회를 제공한다. 특히 금융권에서는 예상치 못한 산출물로 인해 손실이 일어나는 것을 극도로 경계해야 하므로 리스크를 관리하는 하나의 해결책이 될 수 있다.

3. 국내 AI 보안 가이드라인 주요 내용

(1) 국내 가이드라인 분석

국내 AI 산업의 전반적인 진흥 및 규제를 위한 ‘AI 기본법’의 제정이 지연되고 있다. 2023년 2월, 국회 과학기술정보방송통신위원회에서 ‘인공지능 산업 육성 및 신뢰 기반 조성에 관한 법률안’(AI 기본법)이 통과되었지만, 이 법안은 이후 1년 넘게 국회에서 처리가 지연되었다.

국가 인권위원회와 시민단체 등이 현 AI 기본법의 핵심인 ‘우선 허용 · 사후 규제’ 조항에 반대하고 안전과 인권 보호를 더욱 강조하는 방향으로 수정해야 한다며 기본법을 반대했기 때문이다. 결국 21대 국회 임기 만료로 AI 기본법은 폐기되었다 [10].

이 가운데 금융위원회는 2021년 7월 ‘금융분야 AI 가이드라인’을 발표하였고, 2022년 8월에는 ‘금융권 인공지능 활용 활성화 및 신뢰 확보 방안’을 발표하였다.

<표 1> 금융분야 AI 가이드라인(2021)

핵심 가치	내용
금융산업의 책임성 강조 (3중 내부통제장치 마련)	<ul style="list-style-type: none"> (AI 윤리) AI 서비스 개발, 운영시 준수해야할 원칙, 기준 수립 (AI 조직) AI의 잠재위험을 평가, 관리 할 구성원의 역할, 책임, 권한을 서비스 전 단계(기획, 설계, 운영, 모니터링)에 걸쳐 구체적으로 정의 (위험관리) AI 서비스 자체 평가, 관리정책 마련
데이터의 정확성, 안전성 확보	<ul style="list-style-type: none"> AI 학습 데이터의 품질 검증, 개선 및 최신성 유지 개인정보 보호(오남용 방지, 불필요한 개인신용정보 처리 최소화 등)
서비스의 투명성, 공정성 담보	<ul style="list-style-type: none"> 서비스 특성에 맞는 위험 통제 소비자 차별 방지 등 서비스 특성별 공정성 기준을 설정, 평가
금융소비자 권리의 보장	<ul style="list-style-type: none"> 금융소비자 앞 AI 이용 사전 고지, 소비자의 권리 및 이의신청, 민원제기 등 권리구제 방안 등을 알기 쉽게 안내

2021년 7월에 발표된 ‘금융분야 AI 가이드라인’의 주요 내용은 다음과 같다.

- 1) 3중 내부통제 장치 마련 : AI 서비스의 책임 있는 운영을 위해, AI 활용에 관한 윤리 원칙과 기준을 수립한다. 또한 AI 시스템 위험 평가 및 관리를 위한 AI 전담 조직을 구성하며, AI 활용으로 인한 잠재적 위험을 관리하는 데 필요한 위험관리 정책을 수립한다.
- 2) 학습데이터 관리 및 개인정보보호 : 올바른 AI 학습을 위하여 학습 데이터의 출처, 품질, 편향성 등에 대한 조사·검증을 강화하고, 민감한 개인정보 활용 시 비식별 조치 등의 충분한 안전조치와 개인정보 활용 필요성 등을 평가하여 개인신용정보에 대한 오·남용을 방지한다.
- 3) 시스템 위험관리 및 공정성 제고 : AI 활용 결과 불합리한 소비자 차별 등이 발생하지 않도록 서비스 특성별로 위험 요인을 통제하고, AI 시스템 공정성 평가지표를 선정 및 측정하여 서비스의 공정성을 개선한다.
- 4) 소비자의 권리 보장 : 소비자에 AI 활용 사실을 사전에 알려 AI 서비스에 대한 충분한 설명을 제공하고, 소비자가 자신의 권리를 충분히 행사할 수 있도록 보장한다.

〈표 2〉 금융분야 AI활용 활성화 및 신뢰확보 방안(2022)

핵심 가치	내용
양질의 빅데이터 확보 지원	<ul style="list-style-type: none"> • 가명정보 재사용을 허용하는 「금융 AI 데이터 라이브러리」 구축 • 협업을 통한 데이터 공동 확보 • 데이터 전문기관 추가 지정
AI활성화를 위한 제도 점검	<ul style="list-style-type: none"> • AI개발, 활용 안내서 발간 • 설명 가능한 AI 요건 마련 • 망분리 및 클라우드 규제 개선
신뢰받는 AI 활용환경 구축	<ul style="list-style-type: none"> • 금융 AI 테스트베드 구축 • AI 기반 신용평가모형 검증체계 마련 • AI 보안성 검증체계 구축

2022년 8월에 발표된 「금융권 인공지능 활용 활성화 및 신뢰 확보 방안」의 주요 내용은 다음과 같다.

- 1) 「금융 AI 데이터 라이브러리」 구축 : 가명 정보는 이용 목적을 달성한 후 파기해야 하므로 대량의 가명 정보 데이터셋을 구축하고 운영하는 데 어려움이 있다. 이를 해결하기 위해 규제 샌드박스 등을 통해 가명 정보 재사용을 허용하는 AI 데이터 라이브러리를 구축한다.
- 2) 「AI 개발 · 활용 안내서」 발간 : 실제 AI 서비스 도입 시 참고 가능한 기능 · 서비스별 안내서가 필요하다는 의견에 따라, 신용평가 · 여신 심사, 로보어드바이저, 챗봇, 맞춤형 추천, 이상 거래탐지 등 5대 서비스별 AI 개발 · 활용 안내서를 마련한다.
- 3) 망분리 규제 개선 : 원활한 AI 개발 및 활용을 위해서는 외부 API, 클라우드 활용이 필요하므로, 가명 정보를 활용하는 AI 개발 · 테스트 서버에 대해서는 규제 샌드박스 등을 통한 물리적 망분리 예외 허용을 추진한다.
- 4) 금융 AI 테스트베드 구축 : 금융 AI의 정확성 및 신뢰성 확인을 위해서는 다양한 데이터를 통한 검증 과정이 중요하므로, 금융분야 인프라 기관을 통해 다양한 금융 분야 AI 테스트가 가능한 검증 데이터셋 및 테스트 환경을 구축한다.
- 5) AI 기반 신용평가 모형 및 AI 보안성 검증체계 구축 : 금융회사와 핀테크 · 플랫폼 업체들의 신용평가 모형 개발 및 운용 시도가 확대됨에 따라, AI 기반 신용평가 검증체계를 마련한다. 또한, AI 활용 시 개인정보 유출, 알고리즘 오작동 등의 다양한 보안 위험 요소가 존재하므로 이를 위해 AI 보안성 검증체계를 마련한다 [2].

이처럼 한국은 AI 기술의 안전하고 윤리적인 활용을 위해 다양한 가이드라인을 발간하였고, 그 중 대표적으로 ‘2021년 금융분야 AI 가이드라인’과 ‘2022년 8월 금융권 인공지능 활용 활성화 및 신뢰 확보 방안’에 대해 설명하였다. 그러나 이러한 가이드라인에는 한계점이 존재하는 것이 사실이다. 이는 AI 보안성 확보를 위한 규제 및 원칙을 다루고 있지만, 실무에 적용할 수 있는 세부적인 안내 또한 필요하다. 특히, 금융분야는 고객의 재산과 직접적인 관련이 있기에 AI 보안성 확보를 위한 실무 위주의 안내가 더욱 중요하다 [8].

(2) 금융분야 가이드라인 분석

‘2023년 금융분야 인공지능(AI) 보안 가이드라인’은 한국 금융위원회가 발표한 가이드라인이다. 이 가이드라인은 금융권에서 AI를 안전하고 신뢰성 있게 활용하기 위한 다양한 지침을 제공하고 있다. 주요 내용에는 AI 모델 개발 과정에서 고려해야 할 보안 요소와 AI 챗봇 서비스에 대한 보안 체크리스트가 포함된다. 주요 내용은 다음과 같다.

1) AI 모델 개발 단계별 주요 보안 고려 사항

첫째, 학습 데이터 수집 단계에서 오염된 데이터를 학습하여 발생할 수 있는 보안 문제와 성능 저하를 방지하는 것이 중요하다. 이를 위해 신뢰성 높은 출처에서 데이터를 수집하고, 데이터 출처와 수집 시점을 명확히 관리하여 데이터 관련 공격이나 장애 발생 시 그 원인을 파악할 수 있는 데이터 관리 체계를 구축하는 것이 필수적이다.

둘째, 학습 데이터 전처리 단계에서는 수집된 데이터를 학습에 적합한 형태로 가공하여 AI 모델의 품질과 보안성을 높일 필요가 있다. 이상치를 확인하고 처리하며, 적대적 예제 생성 및 학습을 통해 AI 모델에 대한 적대적 공격을 예방해야 한다.

셋째, AI 모델 설계 및 학습 단계에서는 AI 모델이 적대적 공격에 쉽게 노출되지 않도록 설계하는 것이 중요하다. 이를 위해 예상 가능한 단순한 설계를 피하고, AI 모델의 구조를 변형하는 보안 기법을 적용하여 잠재적 공격자가 모델의 정보를 유추하기 어렵게 해야 한다.

넷째, AI 모델 검증 및 평가 단계에서는 학습을 완료한 AI 모델이 잠재적 공격이나 개인정보 유출로부터 안전한지 보안성을 검증하는 과정이 필요하다. 이를 위해 선제적인 적대적 공격을 수행하여 AI 모델의 탐지 및 방어 능력을 확인하고, AI 모델

의 입력과 출력 횟수를 제한하여 공격자가 모델의 정보를 수집하기 어렵게 해야 하며, AI 모델을 통해 개인정보가 타인에게 노출되지 않도록 주의해야 한다.

2) AI 챗봇 서비스에 대한 보안성 체크리스트

AI 챗봇 서비스는 사용자 입력을 통해 작동하므로, 입력된 데이터가 보안 위협이 되지 않도록 하는 것이 중요하다. 특히나 개인정보를 처리하지 않는 챗봇의 경우, 이용자에게 입력창에 개인정보를 입력하지 않도록 사전에 필수적으로 명시해서 실무자에게 알려야 한다. 이는 개인정보 유출을 방지하고, 챗봇이 악의적으로 사용되는 것을 예방한다. 금융위원회에서는 실무자들이 쉽게 활용할 수 있도록 구체적인 보안 요구 사항에 대한 AI 챗봇 서비스 보안성 체크리스트를 제공하고 있다. 본 목차에서는 주요 고려 사항이라 생각되는 체크리스트 질문을 선정하여 분석해 보았다.

〈표 3〉 AI 챗봇서비스 보안성 체크리스트 문항

분류	소분류	연번	보안성 체크리스트
공통	입력제한	6	개인정보를 처리하지 않는 챗봇의 경우 입력창에 개인정보를 입력하지 않도록 이용자에게 사전에 안내하는가?
	중요 정보 보안	8	챗봇 서비스 관련 중요정보*에 대해 적절한 보호대책 (계정/ 권한 관리, 접근통제, 암호화 등)이 적용되어 있는가? *개인정보, AI 모델파일, 학습데이터(학습용, 검증용 등), 챗봇관리시스템, 챗봇인프라 관련 로그, 발화문 원문 저장 DB 등
	학습데이터 관리	20	학습데이터 출처에 대한 신뢰성 평가 기준을 수립하고 이행하였는가?
		23	AI 모델과 학습데이터 등의 주요파일 위변조 시 탐지방안 혹은 무결성 검증 방안이 있는가?
선택	개인정보 활용	25	(챗봇 서비스 내 개인정보 활용 시) 챗봇서비스에 개인정보 또는 민감정보를 활용하는 경우에는 안전성 확보에 필요한 조치를 이행하고 이를 정기적으로 점검 하는가?

1) (연번 6번) 개인정보 입력 금지 명시: 개인정보 유출 방지와 챗봇의 악의적 사용을 예방하기 위해 개인정보 입력 금지 고지가 중요하다. 사용자에게 입력창에서 개

인정보를 입력하지 않도록 명확히 고지하는 것은 실수로 민감한 정보를 입력하는 것을 방지해준다. 예를 들어, 챗봇이 제공하는 서비스의 성격상 개인정보를 다루지 않아도, 사용자가 실수로 민감한 정보를 입력할 가능성이 존재할 수 있기 때문에 개인정보 입력 금지 고지를 통해 사용자의 주의를 환기시키는 것이 중요하다.

2) (연번 8번) 입력 데이터의 보안 처리: 사용자 입력 데이터는 암호화되어 저장되고, 접근 권한이 제한된 방식으로 관리되어야 한다. 이는 데이터 유출 및 무단 접근을 방지하기 위한 필수적인 조치다. 예를 들어, 사용자 입력 데이터가 암호화되지 않고 저장되면, 데이터베이스가 침해당했을 때 사용자의 개인정보가 쉽게 유출될 수 있다. 따라서 모든 데이터는 전송 시 암호화되어야 하며, 저장 시에도 강력한 암호화 알고리즘을 사용해야 한다.

3) (연번 20번) 데이터 출처와 신뢰성 관리: AI 모델의 학습 데이터는 신뢰할 수 있는 출처에서 수집되어야 하며, 데이터의 무결성과 신뢰성을 보장하기 위해 철저히 관리되어야 한다. 신뢰할 수 없는 출처의 데이터는 AI 모델에 부정적인 영향을 미칠 수 있으며, 데이터 오염이나 편향된 학습 결과를 초래할 수 있다. 따라서 데이터 출처와 수집 시점, 메타 정보를 철저히 관리하고, 정기적인 데이터 검증을 통해 학습 데이터의 품질을 유지해야 한다.

4) (연번 23번) 이상치 및 데이터 오염 처리: 이상치 탐지 및 처리 시스템은 데이터의 무결성을 유지하고, AI 모델의 신뢰성을 높이는 데 필수적이다. 데이터 오염은 AI 모델의 성능을 저하할 수 있는 위협이므로, AI 시스템이 의도하지 않은 방식으로 작동하게 만들 수 있다. 따라서 이상치를 탐지하고 처리하는 시스템을 구축하여 데이터의 무결성을 유지하고, 악의적으로 변조된 데이터를 사전에 제거해야 한다. 이는 AI 모델의 안정성과 신뢰성을 유지하는 데 중요한 역할을 한다.

5) (연번 25) AI 모델의 보안성 검증: AI 모델의 보안성 검증은 정기적으로 수행되어야 한다. 이는 AI 시스템이 잠재적 공격에 대해 얼마나 기여하는지 평가하는 중요한 역할을 한다. 적대적 공격 시뮬레이션을 통해 모델의 취약점을 파악하고, 필요한 보안 조치를 취해야 한다. 이러한 검증 과정은 AI 모델이 실제 환경에서 안전하게 운영될 수 있도록 보장하며, 모델의 신뢰성을 높인다. 또한, AI 모델의 입출력 값을 제한하여 잠재적 공격자가 모델의 정보를 수집하기 어렵게 하는 등의 보안 기법을 적용할 필요가 있다.

앞서 분석한 주요 항목들은 챗봇 서비스의 보안성을 확보하는 데 중요한 역할을 한다. 보안성 확보는 금융권에서 AI 기술을 활용하는 데 있어 필수적이며 이를 통해

금융기관은 고객의 신뢰를 확보하고, 법적 규제를 준수하며, AI 기술의 안전하고 효과적인 활용을 촉진할 것이다. 앞으로도 이러한 보안 체크리스트를 지속적으로 업데이트하고 개선하여 신뢰성 있는 서비스를 제공하는 것이 중요하다 [8].

4. 생성형 AI 국외 규제 동향

(1) EU

2024년 3월 13일, EU(유럽 연합)은 최초로 인공지능 법(AI Act, Artificial Intelligence Act)을 가결하였다. 이는 AI에 대한 포괄적인 규제 프레임워크 수립을 의미하며, AI 시스템을 위험 수준에 따라 분류하고 고위험 애플리케이션에 대한 엄격한 요구 사항을 부과한다. AI 시스템을 활용 데이터 민감도와 방식에 따라 4개의 리스크 계층으로 분류하여 위험도가 높은 ‘금지’ 와 ‘고위험’ 으로 분류된 AI에 대해 엄격한 조치를 적용한다. 여기서 생체인식 분류, 사회적 행동이나 개인적 특성을 기반으로 개개인을 평가 · 분류하는 AI를 ‘금지’ 로 분류하고, 고용 및 HR 시스템에 사용되는 등의 AI 시스템은 ‘고위험’ 으로 분류한다. EU AI act 핵심 요소 중 하나는 ‘고위험’ AI 시스템의 식별과 감독 강화이다. 이를 통해 AI 신뢰성 보장과 기본권 보호, 디지털 단일 시장 내에서 혁신의 촉진을 목표로 한다. 또한, AI 시스템의 위험 수준에 따라 의무를 부과한다. 특히, 생체 인식 분류와 같이 개인의 안전과 민주주의와 같은 기본 권리를 침해할 수 있는 AI 시스템을 금지한다. 더 나아가 범용 AI 시스템에 대한 투명성 요구 사항을 설정하고, 혁신 및 중소기업(SMEs) 을 지원하려는 조치도 도입했다. 이 법안은 AI 기술의 혁신을 촉진하면서도 투명성, 안전성, 비차별성을 보장하는 조항과 데이터 거버넌스 및 책임감에 대한 의무를 도입하여 AI 규제에 대한 글로벌 기준을 설정했다. 유럽연합 내 AI 기술의 규제와 표준화가 이루어짐에 따라 글로벌 AI 시장에서도 유사한 규제 프레임워크가 도입될 가능성이 커지게 되어 AI 기술의 국제적 협력과 통합이 더욱 가속화될 것이다 [11].

(2) 미국

2023년 3월, 美는연방거래위원회(FTC), 평등고용기회위원회(EEOC), 소비자금융보호국(CFPB)과 법무부(DJ)는 각 기관의 유관 영역에서 ‘AI를 포함한 소프트웨어 및 알고리즘 프로세스’ 에 대해 관리 · 감독 권한이 있다는 공동 성명을 발표하였다.

2023년 7월 21일, 미국 백악관에서 AI 기업을 대상으로 한 ‘AI 안전 서약’ 을 발표하였다. 서약은 기업들이 생성형 AI 콘텐츠에 워터마크를 포함하는 등 AI를 책임

있고 안전하게 사용할 것을 약속하는 8개의 조항으로, 제품을 대중에게 소개하기 전 안전성 보장, 보안을 최우선으로 하는 시스템을 구축, 워터마크와 같이 AI로 생성된 창작물임을 알 수 있도록 신뢰를 보장한 콘텐츠 개발 및 배포 등의 내용이 포함되었다. 업계 리더들(오픈AI, 구글, 마이크로소프트, 아마존, 메타, 엔트로픽, 인플렉션 AI 등 대형언어모델(LLM)을 보유한 주요 기업들)이 이러한 서약을 자발적으로 AI 개발 및 배포에서 안전, 보안, 투명성을 우선시하도록 장려하였다. 2023년 10월 30일, 조 바이든 정부는 미국 최초의 ‘AI 행정 명령’을 발표했다. 이는 AI 개발과 활용에 대한 대통령령으로 총 8가지 원칙을 발표했다. 이와 함께 미국 행정부는 국내에서 AI 관련 의제를 추진하는 동시에 해외 동맹국 및 파트너들과 함께 AI 개발 및 사용을 관리하기 위한 국제적 프레임 워크 수립을 위해 협력할 것임을 발표했다 [11, 12]. 2024년 5월 16일, 미국은 AI 기술을 안전하고 신뢰할 수 있는 개발 및 사용을 위한 새로운 원칙을 발표했다. 발표한 AI 원칙은 다음과 같다 [13].

- 1) 근로자 역량 강화 : 근로자와 그 대표자, 특히 소외된 지역 사회의 근로자는 직장에서 사용할 AI 시스템의 설계, 개발, 테스트, 학습, 사용 및 감독에 대한 정보를 얻고 진정한 의견을 제시할 수 있어야 한다.
- 2) AI 거버넌스 및 인간 감독 확립 : 조직은 직장에서 사용할 AI 시스템에 대한 명확한 거버넌스 시스템, 절차, 인간 감독 및 평가 프로세스를 갖추어야 한다.
- 3) AI 사용의 투명성 보장 : 고용주는 직장에서 사용되는 AI 시스템에 대해 근로자와 구직자에게 투명하게 공개해야 한다.
- 4) 노동 및 고용 권리 보호 : AI 시스템은 근로자의 단결권, 건강 및 안전 권리, 임금 및 근로 시간 권리, 차별 금지 및 보복 금지 보호를 침해하거나 훼손해서는 안 된다.
- 5) AI를 사용하여 근로자 지원: AI 시스템은 근로자를 지원하고 보완하며, 지원하고 일자리의 질을 개선해야 한다.
- 6) AI의 영향을 받는 근로자 지원 : 고용주는 AI와 관련된 직무 전환 중에 근로자를 지원하거나 숙련도를 향상해야 한다.
- 7) 근로자 데이터의 책임 있는 사용 보장: AI 시스템에 의해 수집, 사용 또는 생성된 근로자 데이터는 범위와 위치가 제한되어야 하며, 합법적인 비즈니스 목표를 지원하는 데에만 사용되어야 하며, 책임감 있게 보호 및 처리되어야 한다.

또한 마이크로소프트와 인디드는 이 원칙을 자사에 적용하기로 이미 합의한 상태

이다. 이러한 AI 원칙들을 통해 AI가 근로자를 보호하고 지원하는 환경을 조성하고 근로자의 복지와 권리를 중심으로, 윤리적으로 AI를 사용하는 데 중요한 역할을 할 것으로 전망된다.

(3) 중국

중국 정부는 AI를 국가적 우선순위로 지정하고, 최근까지도 “중국제조 2025”와 같은 전략적 계획을 통해 AI 기술 개발을 지원하고 있다. AI 활용을 적극적으로 추진하기 위해서 규제와 정책을 점진적으로 발전시키고 있으며, 연구와 개발에 막대한 자금을 투자하며 최근 몇 년 동안 인공지능 분야에서 급속한 성장을 이루었다. 주요 AI 관련 법률로는 데이터 보안법(DSL)과 개인정보 보호법(PIPL)이 있으며, 이 법률은 AI의 안전한 사용을 보장하는 데 중요한 역할을 한다. 또한 생성형 AI 기저에 있는 디지털 데이터를 규제하고 데이터와 개인정보는 생성형 AI의 학습 데이터로 활용되도록 한다 [14]. DSL은 데이터의 수집, 전송, 저장에 관한 전반적인 규제를 제공하며, 데이터의 중요도에 따라 데이터를 등급화하고, 각 등급에 맞는 보안 조치를 요구한다. 고위험으로 분류된 데이터는 더욱 엄격한 보안 조치가 적용된다. PIPL은 2021년에 시행된 첫 포괄적인 개인정보 보호 법률로, 2023년부터 도입된 새로운 규정에 대한 주요 내용은 다음과 같다.

- 1) 동의 요구: 개인 정보를 수집하거나 처리하기 전에 데이터 주체의 자발적이고 명확하며 정보에 입각한 동의를 받아야 한다. 생체 정보, 종교적 신념, 건강, 재정, 위치 정보 등과 같은 민감한 개인 정보에 대해서는 추가적인 보호 조치가 필요하다.
- 2) 데이터 현지화 및 삭제 요구: 특정 기준을 충족하는 경우 데이터 현지화 요구 사항이 적용되며, 데이터 보호 책임자를 지정해야 한다. 수집 목적이 달성되거나 서비스 제공이 종료되면 데이터를 삭제해야 한다.
- 3) 국경 간 데이터 전송 제한: 국경 간 데이터 전송 시 데이터 주체의 동의를 얻고, 데이터 수신자가 PIPL에서 요구하는 데이터 보호 기준을 충족해야 하며, 중요한 데이터를 해외로 전송할 때 사이버 공간 관리국(CAC)의 보안 검토를 완료해야 한다.

이처럼 2023년 중국에서는 개인정보 보호 분야에서 강화된 행정법 집행이 이루어졌다. 법률 분야인 개인정보보호법뿐만 아니라 생성형 AI의 특징인 데이터, 알고리즘, 콘텐츠 생성 등 각각의 기술에 초점을 맞춘 규정을 제정했다. 또한 2024년에는

‘AI 플러스’ 이니셔티브를 통한 산업 체계가 자리 잡으며, AI+교육, AI+산업, AI+문화 등으로 AI의 활용 범위가 꾸준히 확대되고 있다. 이에 따라 해당 분야에 새로운 관리규범을 마련하고 있는데, 대표적인 현행 규정은 인터넷 정보 서비스 알고리즘 추천 관리규정과 인터넷 정보 서비스 딥페이크 관리규정으로 내용은 다음과 같다.

- 1) 인터넷 정보 서비스 알고리즘 추천 관리규정: 알고리즘의 작동 원리를 사용자에게 명확하게 설명해야 하고, 사용자는 데이터의 수집, 사용, 보관에 있어 엄격한 규정을 준수해야 한다. 알고리즘이 일으킬 수 있는 위험은 사전에 평가하고, 이에 대한 대응책을 마련하며 알고리즘은 인종, 성별, 종교 등의 이유로 차별을 해서는 안 된다.
- 2) 인터넷 정보 서비스 딥페이크 관리규정: 딥페이크 콘텐츠는 명확하게 표시되어야 하며, 사용자에게 실제가 아님을 명확히 알려야 한다. 딥페이크 기술을 사용하는 플랫폼은 콘텐츠의 진위를 확인하고, 허위 정보의 유포를 방지할 책임이 있으며, 악의적으로 딥페이크 기술을 사용하여 타인을 속이거나 피해를 주는 경우 법적 제재를 받을 수 있다.

이와 같이, 향후 중국은 AI가 다양한 분야와 접목되어 활용될 때도 관련 산업을 위해 정책과 규정을 분야별로 제정될 것으로 예상된다 [14, 15, 16, 17, 18].

(4) 영국

2023년 3월, 과학혁신기술부(DIST)는 ‘AI 규제에 대한 혁신적 접근방식(A pro-innovation approach to AI regulation)’라는 AI 백서를 발표하였다. 해당 백서에서는 AI 주도 국가로서 자국의 입지 강화, AI의 성장과 번영, AI에 대한 대중의 신뢰성 증진을 목표로 한다. 이를 위해 유연한 규제 체계를 제시하고 있다. AI에 대한 고정 규칙이나 법안 도입을 최소화하고, 기존 관련 기관이 부문별·상황별 지침을 채택한다. 또한, AI 규제를 위한 다섯 가지 주요 원칙을 제시하고 있다.

- 1) 안전 · 보안 및 견고성(safety, security and robustness): 규제 기관은 AI 시스템의 안전 관련 위험을 이해하고 전달해야 한다. 또한 AI 개발자와 배포자는 안전 위험 평가를 하고 발견된 위험 완화 조치를 취해야 하며, AI의 수명주기 전반에 걸쳐 사이버 보안 위험을 완화하고 회복력을 구축해야 한다.
- 2) 투명성 및 설명 가능성(transparency and explainability): 투명성과 설명 가능성은 AI에 대한 신뢰를 쌓고, AI의 채택 및 혁신을 증진하는 데 중요한 요소이다. 이를

위해 규제 기관은 AI 개발자가 AI 시스템 작동 방식의 정보를 명확하게 제공하고, AI 배포자가 최종 사용자에게 이를 이해하기 쉽게 설명하도록 한다. 이는 최종 사용자가 충분한 정보를 통한 자신의 권리를 행사하고, 제품을 구성하는 합법적·불법적 사용에 대해 이해하는 것을 목표로 한다.

- 3) 공정성(fairness): 규제 기관은 자신의 규제 범위 내 AI 시스템 결과에 적용되는 공정성에 대한 기술을 계속 개발하고, 기존의 기술을 안내하며, 여러 규제 범위에 걸쳐 공정성의 기술을 조율하고 공동의 도구 및 지침을 개발해야한다. 예시로 2010 평등법이나 영국 GDPR과 같은 영국의 기존 법률을 준수하는 방식으로 AI를 사용해야하며, 개인을 차별하거나 불공정한 상업적 결과를 만들면 안된다 [19].
- 4) 책임 및 거버넌스 (accountability and governance): 투명성과 설명 가능성은 통해 AI가 사용되는 방식에 대하여 적절한 감독이 필요하다. 또한, 결과에 대한 명확한 책임을 보장하기 위한 책임 의식을 강화해야한다.
- 5) 경쟁 가능성 및 보상(contestability and redress): 규제 기관은 AI 개발자 및 배포자와 사용자들이 AI가 내린 유해한 결과에 이의를 제기할 수 있는 경로를 명확히 해야하며, 구제 방법을 인지할 수 있도록 한다.

이 프레임워크는 규제기관이 AI 도입과 관련된 문제를 효과적으로 다루고 있다고 판단하는 경우, 특정 기술에 얹매이지 않도록 중립적인 접근 방식을 채택할 수 있게 한다. 이를 통해 AI 개발자, 배포자 및 최종 사용자들이 규제 기관의 대응 방식을 이해할 수 있도록 한다. 이는 AI 기술의 혁신을 주도하고 영국에서 AI 활용을 촉진하기 위해 필수적인 명확성과 신뢰성을 보장하는 데 있어 핵심적이다 [20, 21, 22].

(5) 캐나다

캐나다 정부는 AI의 잠재적 이점을 강조하면서도 사용 전 신중한 위험 평가와 주요 이해관계자 참여를 권장한다. 이는 법률, 개인 정보 보호, 보안, 지식재산권 및 인권 분야의 연방 법률 및 정책 준수하기 때문이다.

이 가이드라인에 따르면, 생성형 AI 도구 사용 시 먼저 저위험 사용(예: 이메일 작성, 초안서 편집)으로 시작하여 실험을 진행해야 한다. 고위험 사용(예: 챗봇 배포, 클라이언트 정보 요약 생성)을 고려하기 전에 충분한 위험 평가와 완화 조치를 시행해야 한다.

연방 기관은 항상 모범 사례와 위험 완화 조치를 각 용도를 맞게 조정해야 하며, 고위험 사용을 고려하기 전에 저위험 사용을 실험해야 한다. 또한 공무원들은 윤리적 의사결정 가이드를 참조하여 생성형 AI 도구 여부를 결정해야 한다. 생성형 AI는 직원을 대체하는 것이 아니라 돋는 용도로 평가되어야 하며, 윤리적이고 FASTER 원칙에 맞도록 사용해야 한다 [21]. 기관은 이러한 도구의 책임 있는 사용을 탐색하고 이를 적극적으로 추진해야 한다. 다음은 FASTER 원칙이다.

- 1) 공정성(Fair) : 도구의 콘텐츠가 편견을 포함하지 않도록 하고 인권 및 접근성을 준수한다.
- 2) 책임 (Accountable): 생성된 콘텐츠의 정확성, 합법성, 윤리성을 보장하고, 모니터링 메커니즘 구축한다.
- 3) 보안(Secure) : 정보 보안과 개인 정보 보호를 철저히 관리한다.
- 4) 투명성(Transparent) : AI 사용 사실을 명확히 알리고, 관련 정책과 교육 데이터를 공개한다.
- 5) 교육(Educated) : 도구의 장점과 한계를 교육하고, 효과적인 사용법을 학습한다.
- 6) 관련성(Relevant) : 도구가 사용자와 조직의 요구를 충족하고 환경적 영향을 고려하는지 확인한다.

직원들이 AI 도구를 효과적이고 책임감 있게 사용할 수 있도록 교육하고 편향된 콘텐츠 감지 등 관련 지식을 향상하게 시키는 지원을 제공해야 한다. 또한 정부 정보 보호 및 보안 요구 사항을 충족하는 안전한 AI를 제공하고, 관련 정책을 준수하도록 해야 한다. 기관은 감독 및 성능 관리 프로세스를 통해 도구의 영향을 모니터링하고, 법률 및 정책 준수를 확인하며, 이해관계자와 협의하여 고위험 배포를 준비해야 한다. 관리자는 도구 사용에 대한 현실적인 기대치를 가져야 하며, 교육기관은 위험과 기회를 평가하여 지침을 개발해야 한다 [23].

(6) 인도

인도는 AI 기술의 발전을 촉진하면서도 금융권에서의 AI 사용에 따른 위험을 관리하기 위해 규제를 점진적으로 도입하고 있다. 인도 정부에서는 개인 데이터 보호에 대해 강화된 규제를 적용하고 있는데, 이를 위해 2023년 새로운 디지털 개인정보 보호법인 Digital Personal Data Protection Act(DPDP Act)를 제정했다. DPDP Act는 인도의 데이터 보호와 관련된 주요 규제로 AI 시스템이 개인정보를 어떻게 수집,

저장, 처리하는지를 규제함으로써 데이터 처리자의 책임을 강화하는 데 중점을 두고 있다. 이 법안이 포함하고 있는 주요 요소는 다음과 같다.

- 1) 데이터 주체의 권리 보호: 데이터 주체는 자신의 개인정보가 어떻게 사용되는지에 대한 명확한 정보를 받을 권리가 있으며, 자신의 데이터를 열람, 수정, 삭제할 권리를 가진다. 특히, 데이터 주체는 개인정보 처리에 대한 동의를 철회할 권리가 있으며, 이에 따라 데이터 처리자는 해당 데이터를 삭제해야 한다.
- 2) 데이터 처리자의 책임 강화 : 데이터 처리자는 데이터 주체의 동의를 받아 데이터를 수집해야 하며, 데이터를 안전하게 저장하고 처리해야 한다. 데이터 유출 시 즉각적으로 데이터 보호 위원회(Data Protection Board of India)에 보고해야 하며, 데이터 보호 책임자(Data Protection Officer)를 임명하여 데이터 처리와 관련된 모든 활동을 관리해야 한다. 또한, 데이터 처리자는 데이터 보호 영향 평가(Data Protection Impact Assessment)를 수행하고 정기적인 보안 감사(Security Audit)를 통해 보안성을 유지해야 한다.
- 3) 데이터 관리와 보안: DPDP Act는 데이터 무결성과 보안을 유지하기 위한 기술적, 관리적 조치를 취할 것을 요구한다. 데이터 접근 권한을 제한하고, 정기적인 보안 점검을 통해 해킹이나 데이터 유출을 방지해야 한다. 또한, 모든 데이터 유출 사건은 즉시 데이터 보호 위원회와 관련된 데이터 주체에게 보고되어야 하며, 이는 데이터 보호의 투명성을 높이는 데 기여한다.
- 4) 국제 데이터 전송 : DPDP Act는 데이터의 국경 간 전송에 대한 규제를 포함하고 있다.
- 5) 데이터 전송은 특정 조건을 충족할 때만 허용되며, 데이터 보호 표준을 준수해야 한다. 이는 국제적으로 데이터가 안전하게 전송되고 보호될 수 있도록 보장하는 중요한 조치이다.
- 6) 아동 데이터 보호 : DPDP Act는 아동의 개인정보 보호를 강화하기 위해 부모나 보호자의 동의를 요구한다. 또한, 아동 데이터를 대상으로 한 추적, 행동 모니터링, 맞춤형 광고 등을 금지하며, 아동 데이터를 처리하는 경우 특별한 보호 조치를 취해야 한다.

DPDP Act는 AI 기술이 금융 산업에서 안전하게 활용될 수 있도록 하는 데 중요한 역할을 하지만 아직 규제 개선이 진행되고 있는 만큼 한계점도 존재한다. 대표적으로, 모든 개인 데이터 유출은 데이터 보호 위원회에 보고해야 하며 데이터 주체에게도 통지해야 한다는 점에서 정보 과부하와 불필요한 경고를 초래할 수 있다. 이

러한 광범위한 정부의 권한은 개인정보 보호의 투명성을 저해할 수 있으며 특정 상황에서 권한이 남용될 우려도 있다. 따라서, DPDP Act의 실효성을 높이기 위해서 법안의 세부 규정과 정부의 권한 사용 방식에 대한 지속적인 감시와 개선이 필요하다 [24, 25, 26, 27].

(7) 싱가포르

싱가포르는 2018년 금융분야 인공지능 및 데이터분석(AI and Data Analysis, AIDA) 활용에 관한 14개 FEAT 원칙을 제시한 데 이어 2023년 2월부터 베리타스(Veritas) 컨소시엄을 통해 AIDA 라이프사이클과 관련된 FEAT 원칙 준수를 위한 평가방법론 및 전 단계(End-to-End) 체크리스트 등을 제시한다. 다음은 싱가포르 통화청(MAS)의 FEAT 원칙이다.

- 1) 공정성(Fairness): AIDA 기반 결정은 특정 개인이나 그룹에 대한 정당성을 확보해야 하며, 개인 속성을 입력 요소로 사용하는 것은 정당화된다. 또한, 의도하지 않은 편향을 최소화하기 위해 AIDA 모델을 정기적으로 리뷰·검증해야 하며 모델이 의도된 대로 작동하도록 보장해야 한다.
- 2) 윤리성(Ethics) : AIDA는 회사의 윤리적 표준, 가치 및 행동강령과 일치해야 하고, 인간중심의 윤리적 기준을 준수해야 한다.
- 3) 책임감(Accountability): 내부적 책임으로, AIDA 기반 결정에서 AIDA의 사용은 감독 당국의 승인이 필요하다. AIDA를 사용하는 기업은 내부 개발 또는 외부에서 수급된 AIDA 모델에 대해서도 책임을 져야 하며, AIDA 사용에 대해 경영진과 이사회 의 적극적 인식이 필요하다. 외부적 책임으로, 데이터 주체의 질문, 이의신청, 검토 등을 위한 채널 및 데이터가 제공되어야 하고, 데이터 주체에 의해 검증된 데이터 및 보완 데이터는 AIDA 기반 결정 검토 시 고려된다.
- 4) 투명성(Transparency) : 신뢰성을 높이기 위해 AIDA 사용은 데이터 주체에게 적극적으로 공개된다. 데이터 주체는 AIDA에 사용된 데이터와 그 영향에 대한 명확한 설명을 제공받으며 요청 시 결과에 대한 명확한 설명을 제공받아야 한다.

<표 4> 싱가포르 통화청(MAS) 주도 FEAT 평가방법론

FEAT 평가방법론	일반	Fairness 체크리스트 활용	Ethics & Accountablility 체크리스트 활용	Transparency 체크리스트 활용
1) 실천 원칙	○	○	○	○
2) 시스템 컨텍스트 정의와 설계	○	○	○	○
3) 입력 데이터 준비	○	○		
4) 구축 및 유효성	○	○		○
5) 배포(Display) 및 모니터링	○	○	○	

다음은 싱가포르 통화청(MAS) 주도 FEAT 평가방법론에 관한 내용이다. 금융 서비스 기관(FSI)은 FEAT 원칙을 실천하고 이러한 원칙이 제대로 준수되고 있는지 전제적으로 평가하기 위해, AIDA 시스템 개발 라이프사이클 전반에 걸쳐 FEAT 평가방법론을 포함해야 한다. FEAT 평가방법론은 FEAT 각 원칙(공정성, 윤리성 및 책임감, 투명성)을 고려한 체크리스트 질문 세트를 포함하여 5개의 부분으로 구성되어 있다. 체크리스트 대부분의 질문은 FEAT 각 원칙에 맞추어 설계되며, 일부 질문은 AIDA 시스템을 더 꼭넓게 평가하는 데 도움이 되도록 구성된다 [28, 29, 30, 31, 32].

(8) 일본

2024년 4월, 일본은 AI의 사회적 적용과 거버넌스를 위한 비즈니스 운영자들이 협력하여 활동할 수 있도록 ‘AI 비즈니스 가이드라인(AI Guidelines for Business Ver 1.0)’을 제정하였다. AI 비즈니스 가이드라인은 인간의 존엄성을 존중하며 능력을 발휘할 수 있는 도구로 AI 활용 추구, 다양한 배경과 가치를 가진 사람들이 각자의 행복을 추구할 수 있는 사회를 만들기 위해 AI 기술 사용, 지속 가능한 사회를 기본 철학으로 한다. 그리고 이를 실현하기 위해 인간 중심, 안전, 공정성, 프라이버시 보호, 보안 확보, 투명성, 책임성, 교육/정보 이해, 공정한 경쟁 보장, 혁신의 공통적 지도 원칙을 제시한다. 이러한 원리를 바탕으로 AI 개발자, AI 제공자, AI 비즈니스 사용자별로 구체적인 지침을 제시한다.

1) AI 개발자

데이터 전처리 및 학습 시, 적절한 데이터 학습을 위해 개인정보나 지식재산권 보호가 필요한 데이터를 법규에 따라 처리하고, 데이터 관리 및 접근 통제 기능을 도입한다. 또한, 데이터 품질을 적절히 관리하고, 데이터 편향을 인지하여 AI 모델을 훈련시킨다.

AI 개발 시에는 인간의 생명과 안전을 고려한 정책과 보안 조치를 설정하고, AI 모델의 알고리즘 편향을 고려한다. 개발 후에는 AI 시스템에 대한 새로운 보안 위협에 대응하기 위해 최신 보안 동향을 고려하고, 관련 이해관계자들에게 AI 시스템 정보와 예상 위협 및 대응책을 제공한다.

2) AI 제공자

AI 시스템 구현 시, AI가 인간의 생명, 신체, 재산, 정신 및 환경에 해를 끼치지 않도록 하고 AI 개발자가 설정한 범위 내에서 AI를 사용한다. 데이터를 제공할 때 공정성을 보장하고, AI 모델의 합리성을 정기적으로 평가하여 편향성을 모니터링한다. 프라이버시 보호와 보안을 위한 메커니즘을 도입하며, 의사 결정에 영향을 미치는 시스템에 관한 내용을 문서화한다.

AI 시스템이나 서비스 제공 후에는 AI 시스템이나 서비스가 목적에 맞게 사용되고 있는지 주기적으로 검증하고, 프라이버시 침해에 대한 대응책을 마련한다. 또한, 최신 위험 동향을 식별하여 취약점 공격에 대비하고, AI 비즈니스 사용자에게 데이터 정확성과 최신성, In-context learning 과정에서의 부적절한 학습, 개인정보 입력 문제에 대해 주의를 환기시킨다.

3) AI 비즈니스 사용자

AI 시스템 및 서비스 사용 시, 정확성과 최신성이 보장된 데이터를 입력한다. 입력 데이터의 편향을 고려하여 공정성이 보장된 데이터를 사용하고, 프롬프트에서의 편향에도 주의한다. AI 시스템에 부적절한 개인정보 입력을 피하고, AI 제공자가 지정한 보안 지침을 준수한다. 관련 이해관계자에게 AI 사용 지침, 데이터 제공 방식, 프라이버시 정책 등을 알리고, AI 출력 결과를 평가 자료로 사용할 때는 정확성, 공정성, 투명성을 보장하며 자동화 편향을 고려한 합리적인 판단을 내린다. 평가 대상자가 설명을 요청할 때 이를 수용하여 설명을 제공한다 [33, 34].

5. EU, 미국, 중국의 AI 정책의 시사점

EU의 AI 법안은 AI 시스템을 관련 위험에 따라 체계적으로 분류하고 윤리적 규범을 준수하도록 엄격한 규제를 시행하고 있다. EU는 강력한 AI 규제를 통해 기술의 안정성과 윤리성을 강조한다. EU AI Act는 AI 시스템을 활용 데이터 민감도와 방식에 따라 4개의 리스크 계층으로 분류하여 위험도가 높은 AI에 대해 엄격한 조치를 적용한다. EU AI Act의 핵심 요소 중 하나는 ‘고위험’ AI 시스템의 식별과 감독 강화이다. 이를 통해 AI 신뢰성 보장과 기본권 보호, 디지털 단일 시장 내에서 혁신의 촉진을 목표로 한다.

하지만 이에 대해 우려하는 목소리도 적지 않다. 범용 AI를 개발하거나 서비스하는 기업에 대해 투명성 의무를 주고, 범용 AI 중에서도 시스템적으로 위험이 있는 ‘강력한’ 시스템으로 분류되면 시스템적 위험 평가·완화, 사고 보고 등 의무가 추가로 부여되기 때문이다. 이는 AI를 개발 및 서비스하는 기업에 부담으로 다가올 수 있으며, 장기적으로 AI 개발이 위축될 수 있다는 예상도 있다 [35].

현재 AI 거대 강국인 중국과 미국은 서로 다른 AI 거버넌스에 대한 태도를 보이며 정책 로드맵이 갈라졌다. 이는 두 국가의 정치·경제·사회적 체제의 차이에서 발생한다고 볼 수 있다. 두 나라 모두 세계적으로 가장 발전된 AI 역량과 산업을 가지고 있고, 미국과 중국의 AI 규제 정책에는 주목할 차이점이 분명히 존재하며, 두 나라의 경쟁은 글로벌 차원에서 중요한 의미가 있다.

우선 정부 역할의 관점에서 살펴보면, 중국 정부는 AI 규제 정책을 적극적으로 주도하고 있다. 중국은 최근 디지털 분야 정책이나 법률 제정에 대해 유례없는 속도 전을 펼치고 있다. 법률제정이 결코 빠른 국가가 아님에 불구하고, 유독 디지털 분야는 예외인 것이다 [36]. AI 기술의 성장을 국가 우선순위로 정할 만큼 관련 AI 발전 전략 수립과 그에 대응하는 법·제도를 마련하는 등 주도적으로 정책을 이끌고 있다.

반면, 미국 정부는 상대적으로 규제에 소극적인 태도를 보인다. 그동안 개별 기관의 자율적인 규제 체계 수립을 지원하는 수준이었으나, 최근 연방 정부 차원에서 포괄적인 AI 규제 정책을 마련하고 있다. 이는 기본적으로 친기술적 규제를 마련하는 성향이 있기 때문이다. 하지만 AI 사용으로 인한 부작용이나 위험에 대해서는 꾸준하게 규제의 범위나 강도를 확장하고 있다.

또한, 규제 목적의 차이도 주목할 만하다. 중국은 AI 기술의 안전한 사용과 사회적 영향력 관리에 중점을 두고 있다. 중국 정부는 AI의 잠재적 위협을 ‘여론 선동’의 위험과 영향력으로 보는 경향이 있다. AI 기술이 허위 정보를 생성하고 이를 확산

시킬 수 있는 잠재력을 경계하는 것이다. 이에 따라, 여론 조작이나 허위 정보의 유포를 방지하기 위해 알고리즘 투명성 강화, 사용자 데이터 통제 권한 확대, 불법 및 불량 콘텐츠 확산 방지의 목적을 가진 규정을 강화하고 있다. 이는 AI 기술이 가져다주는 사회적 영향력 관리에 초점을 두고 부작용을 최소화하는 것에 주력한다고 볼 수 있다. 반면, 미국의 경우에는 중국과 달리 AI의 혁신을 촉진하면서도 책임성과 윤리성을 강화하는 방향으로 규제를 마련하고 있다.

미국 정부는 AI 기술의 활용을 통해 경제적 이익과 국가 경쟁력 제고에 더 큰 관심을 보인다. 역사적으로 산업 자율 규제와 유해 응용 프로그램에 대해 사후 개입을 해왔던 미국이기에, AI의 성장과 위험을 신속하게 입법화하지 못하고 있어 입법 과정과 기술의 발전 속도 차이에 따른 격차가 생기게 된다. AI 연구 및 개발에는 상당한 투자가 이루어졌지만, 거버넌스 측면에서는 분산되어 있고 통일된 AI 전략이 부족하다. 이런 방식에는 미국 정부가 AI 거버넌스에 뒤처지고 있다는 지적도 있다.

환경적인 관점에서도 차이가 있다. 미국은 국가 의회에서의 움직임은 적지만, 주 의회에서의 AI 관련 법안 제정이 활발하다. 국가 단위의 AI 규제에 있어서 연방과 주 정부 간 조율이 불가피하고 그렇기에 2024년 1월 말 기준 16개 주에서는 이미 AI 관련 법안을 제정했다. 특히 주목해야 할 주 중 하나는 캘리포니아이다. 세계적인 IT 기업 및 기술은 기술 혁신과 창의성을 중시하는 미국의 문화적 특성을 따라 캘리포니아 실리콘밸리를 중심으로 밀집해 있다. AI 산업의 경제적 가치는 매우 크며, 실리콘밸리는 미국 AI 기술 발전의 중심지 역할을 하고 있다. 하지만 주마다 다양한 규제는 규정 준수의 복잡성 및 이를 따르기 위한 비용으로 인해 잠재적인 비즈니스 위축을 초래할 수 있다. 미국 전역의 주법이 복잡하게 얹혀 있는 상황은 AI를 구현하는 데 규정 준수 및 윤리적 고려 사항을 우선시하기에 기술 개발이 늦어질 수 있다는 관점이 존재한다. 따라서 AI 사용에 대한 일관된 프레임워크를 제공하기 위한 국가 표준 또는 법률의 필요성이 대두된다. 종합적으로 지역적 편중을 해소하고 균형 있는 발전을 위한 주 정부의 정책적 노력이 필요할 것이다 [37].

반대로 중국은 중앙 집권적 체제를 가지고 있어 중앙 정부 주도로 일관된 규제 정책을 추진할 수 있다. 더불어 국가 주도 경제 체제의 성격을 띠고 있어 정부의 개입 및 통제가 가능하다. 그러므로 거대한 인구와 넓은 시장에서 발생하는 방대한 양의 데이터를 수집하기에 적합한 환경을 갖추고 있다. 이러한 환경은 다양한 형태의 데이터셋을 구성하는 데 중요한 역할을 한다. 지난해 8월 15일, 중국은 생성 AI 기술 개발에 대한 제한을 부과하는 선구적인 법을 제정하여 주요 중국 기술 기업 8

곳이 대화형 AI 서비스를 배포할 수 있는 CAC의 승인을 받을 수 있게 되었다 [38]. 많은 부분을 오픈소스로 공개하는 것을 허용하여 전 세계 개발자들이 이를 활용할 수 있도록 허용한 것이다. 이는 중국의 사상과 기술이 전 세계로 확산되는 파장을 불러올 수 있다는 시사점을 가진다. 이와 같이, 최근에 중국의 기술 표준이 글로벌 표준으로 자리 잡는 데 크게 기여를 하고 있기 때문에 글로벌 시장에서는 중국이 주축을 흔들 만큼 급격한 발전을 이루고 있다.

미국은 이러한 중국의 AI 기술 발전에 대해 불안감을 느끼며 중국의 AI 개발을 늦추기 위해 기술 규제와 수출 통제 등의 조치를 취하고 있다 [39]. 조치의 하나로 미국은 진보된 AI 모델에 중국 접근 규제하는 방안을 고려 중이다. 미국 정부는 챗GPT와 같은 인공지능(AI) 시스템의 핵심 소프트웨어인 진보된 AI 모델을 독점 또는 비공개 소스화 함으로써 중국의 미국 AI 기술에 대한 접근을 막는 방안을 추진 중인 것으로 알려졌다. 미국 정부는 미국의 적성국들이 방대한 양의 텍스트와 이미지를 마이닝 하는 모델을 사용해 파괴적인 사이버 공격을 수행하거나 강력한 생물학적 무기를 만들 가능성이 있을 것으로 보고 있다 [40]. 이러한 미국 정부의 움직임은 중국 기업의 인공지능 기술이 미국 기업보다 2년 뒤쳐져 있다는 평가가 나오게 한다. 차이 충신 알리바바 회장은 AI 분야에서 중국 기업들이 미국 기업들에 비해 2년 뒤쳐져 있는 것으로 평가했다. 차이 회장은 “그래픽처리장치(GPU)와 같은 첨단 반도체에 대한 중국 기업의 접근을 금지하는 미국의 수출 규제가 알리바바를 포함한 중국 테크 기업들에 분명히 영향을 미쳤다.”고 말했다 [41].

자국의 AI 기술의 혁신과 상용화를 적극적으로 추진하면서 중국의 기술 발전을 견제하는 미국의 전략에도 불구하고, 중국은 견제를 피하며 자국의 AI 기술을 확장하는 추세로 현 AI 규제에 대한 글로벌 동향은 미국과 중국의 경쟁 구도에 도달하게 되었다. 이러한 상황 속 2024년 5월, 중국은 아랍 국가들과 협력을 계속 확대하며 단결을 특징으로 하는 집단 협력의 모델을 창출할 것이라고 덩 리(Deng Li / 邓励) 중국 외교부 부부장이 발표하였다 [39]. 중국 아랍 간의 공동체를 구축하여 서로의 이익을 쟁기면서 미국의 견제를 피하는 방향으로 추진하는 것이다. 중국과 아랍국가들이 협력하여 AI 기술 발전시키는 것은 여러 가지 측면에서 EU와 미국에 영향을 미친다.

EU는 기술의 안전성과 윤리성을 강조하기 위해 강력한 AI 규제를 시행한다. 중국은 AI 기술의 성장을 국가 우선순위로 정할 만큼 관련 AI 발전 전략과 그에 대응하는 법·제도를 마련하여 주도적으로 정책을 추진하고 있다. 미국은 친기술적 규제

를 마련하는 성향으로 그동안 개별 기관의 자율적인 규제 체계 수립을 지원하는 수준이었으나, 최근 연방 정부 차원에서 포괄적인 AI 규제 정책을 마련하고 있다. 우리나라는 현재 규제 및 가이드라인이 아직 구성 중이다. 이러한 점들을 고려할 때, 우리나라는 무분별하게 규제하기보다는 규제기관의 AI 관련 이해도 증진과 역량 강화에 중심을 둔 유연한 대응이 필요할 것이다. 따라서, 다음 장에서 우리나라에 현재 가이드라인을 분석하여 이러한 대응이 어떻게 적용될 수 있을지 살펴보겠다.

II. 결론 및 제언

최근 인공지능의 폭발적인 성장은 다양한 분야에 응용되어 사회 및 산업의 여러 분야에 광범위한 영향을 미치고 있다. 특히, 금융분야에 AI 도입은 금융 거래의 효율성과 업무의 효율성을 높이고, 새로운 서비스를 통해 고객 만족도를 향상하는 등 경제적 이점을 가져오고 있다. 그러나 이러한 변화와 함께 보안 위협 또한 커지고 있다. 따라서 금융기관들은 보안 위협에 대한 적절한 대응 방안과 규제 체계를 마련해야 할 것이다.

본 연구에서는 생성형 AI 기술의 금융권 활용에 따른 주요 보안 위협을 분석하고, 이에 대한 해결 방안을 모색했다. 데이터 유출, 데이터 오염, 환각 등 다양한 보안 위협에 대해 비식별화, 임계치 설정, RAG, HITL 등의 기술적 대응 방안을 제시하였다. 이러한 기술적 대응 방안을 통해 금융권은 생성형 AI 기술의 장점을 극대화하면서도 보안 위협을 효과적으로 관리할 수 있을 것으로 기대된다. 하지만 이를 뒷받침하기 위한 법적 규제와 정책적 지원 또한 필요할 것이다.

본 연구에서는 국내외 AI 가이드라인에 대해 분석하였다. 해외 주요 AI 선도국 중 EU는 AI 시스템을 위험 수준에 따라 분류하고 윤리적인 규범을 준수하는 데 있어 엄격한 규제를 시행한다. 반면에 중국은 AI 기술의 신속한 발전을 위해 규제를 완화하고 있으며, 자국의 기술 표준을 글로벌 표준을 만들고자 한다. 이에 따라 AI 기술의 선두를 달리는 미국은 중국을 견제하며 보안과 윤리 문제에 더 집중하고 있다. 아직 AI 규제가 만들어지고 있는 인도, 싱가포르 등의 나라들도 선도국들의 규제를 참고하여 규제 및 가이드라인을 작성하며, AI 활용에 더욱 적극적인 태도를 보인다. 이를 통해 아직 규제를 준비 중인 한국에 적용할 수 있는 부분을 참고해야 할 것으로 보인다. 따라서 우리는 해외의 AI 규제를 살펴보고, 국제적인 흐름을 파악하여 규제의 혁신과 균형을 이루는 방향으로 나아가야 한다. 이는 우리나라가 글로벌 AI 시장에서 경쟁력을 유지하고 확장하는 데 중요한 초석이 될 것이다. AI 기술 활용의 발전을 위해서는 AI로 인한 모든 부작용을 예측해 무분별하게 규제하기보다 규제기관의 AI 관련 이해도 증진과 역량 강화에 중심을 둔 유연한 대응이 필요할 것이다. 향후 연구로는 빠르게 변화하는 AI 관련 규제 동향을 지속적으로 모니터링해 금융권에서 AI 활용 기술 발전에 따른 새로운 대응 전략에 관해서 연구하고자 한다.

[참고문헌]

- [1] 홍동숙, “금융 AI 시장 전망과 활용 현황 : 은행권을 중심으로 ”, 한국신용정보원 CIS 보고서, no 47, pp. 2-4, 2022.05.16
- [2] 금융위원회 보도자료 “(금융규제혁신) 금융분야 인공지능 활용 활성화 간담회 개최 - 금융분야 인공지능 활용 활성화 및 신뢰확보 방안 발표 -”, 2022.08.04.
- [3] 류정희(애플경제), 뱅크오브아메리카 '에리카'... 세계 금융 챗봇 모델로 주목 [Internet], Available : <https://www.apple-economy.com/news/articleView.html?idxno=59423>, 2024.07.04
- [4] 김지혜(폴리뉴스), NH농협, 최초로 전국에 AI행원 배치…시중은행 AI 서비스 확대 [Internet], Available : <https://www.polinews.co.kr/news/articleView.html?idxno=636544>, 2024.07.04
- [5] 금융보안원, 금융보안원이 전망하는 2024년 디지털금융 및 사이버보안 이슈 [Internet], Available : <https://www.fsec.or.kr/bbs/detail?menuNo=69&bbsNo=11362>, 2024.06.30
- [6] 김선애(데이터넷), 지난해 딥페이크 3000% 증가 . . . 생체인식 사기 40%, 딥페이크 [Internet], Available : <https://www.datanet.co.kr/news/articleView.html?idxno=194489>, 2024.07.07
- [7] 오태준, “AI 확산이 동반하는 리스크에 대한 인식과 대비 - 금융업계 AI 확산으로 인한 리스크와 대응방안”, 우리금융경영연구소, 2024.05.31
- [8] 금융보안원, “금융분야 AI 보안 가이드라인”, pp.42~49, 2023.04.17
- [9] 김영옥(SAMSUNG SDS), AI 시스템을 강화하는 휴먼인더루프(HUMAN IN THE LOOP), [Internet], Available : https://www.samsungsds.com/kr/insights/human_in_the_loop.html, 2024.07.06
- [10] 김수연(동아일보), 산업 육성할 ‘AI 기본법’ 1년넘게 방치하다 폐기, [Internet], Available : <https://www.donga.com/news/Economy/article/all/20240612/125384182/2>, 2024.07.05
- [11] 정보통신산업진흥원, “2024년 AI 규제 정책 전망”, 2024.02
- [12] 김경숙, 홍건식, “바이든 행정부의 첫 인공지능(AI) 행정명령과 시사점”, 국가안보전략연구원, no 480, 2023.11.10.
- [13] 정한영(인공지능신문), 美 조 바이든 행정부, 근로자 복지를 위한 혁신적인 ‘인공지능 원칙’ 발표 [Internet], Available : <https://www.aitimes.kr/news/articleView.html?idxno=31179>, 2024.07.02
- [14] 세계법제정보센터, “중국의 AI 관련 법제 현황”, 2024.04.03.
- [15] Arendse Huld(CHINA BRIEFING), China Cybersecurity and Data Protection Regulations – 2023 Recap and 2024 Outlook, [Internet], Available

:<https://www.china-briefing.com/news/china-data-protection-regulations-2023-2024/>, 2024.07.07

- [16] Alex Roberts(Linklaters), China – At-a-glance summary of the new data transfer regime, [Internet], Available :
<https://www.linklaters.com/en/insights/blogs/digilinks/2023/january/china---at-a-glance-summary-of-the-new-data-transfer-regime>, 2024.07.07
- [17] James Gong, Fengming Jin(Bird&Bird), China Data Protection and Cybersecurity: Annual Review of 2023 and Outlook for 2024 (I), [Internet], Available :
<https://www.twobirds.com/en/insights/2024/china/china-data-protection-and-cybersecurity-annual-review-of-2023-and-outlook-for-2024-1>, 2024.07.03
- [18] Wang Yongzhan Tian Xianjian(人民日报), 用好“人工智能+”赋能产业升级[Internet], 2024-03-07
- [19] GOV.UK, UK unveils world leading approach to innovation in first artificial intelligence white paper to turbocharge growth, 2023.03.29
- [20] GOV.UK, Guidelines for examining patent applications relating to artificial intelligence (AI), 2024.05.07
- [21] GOV.UK, Implementing the UK’s AI regulatory principles: initial guidance for regulators, 2024.02.06
- [22] 국가전략정보포털, “A pro-innovation approach to AI regulation (AI 규제에 대한 혁신적 접근 방식)”, 2023.03.28
- [23] Government of Canada, Guide on the use of generative artificial intelligence, 2024.06.26
- [24] Gayathri Haridas, Sonia Kim Sohee, Atharva Brahmecha, The Key Policy Frameworks Governing AI in India [Internet], Available :
<https://accesspartnership.com/the-key-policy-frameworks-governing-ai-in-india/>, 2024.06.29
- [25] Rahul Kapoor, Shokoh H. Yaghobi, Theresa T. Kalathil, “AI Regulation in India: Current State and Future Perspectives”, Morgan Lewis, PA, 2024.01.26
- [26] Anirudh Burman, “Understanding India’s New Data Protection Law”, Carnegie Endowment for International Peace, Washington, D.C., 2023
- [27] Ikigai Law, Summary Of The Digital Personal Data Protection Bill, mondaq, Bristol, 2023

- [28] Monetary Authority of Singapore(2022.02.04), MAS-led Industry Consortium Publishes Assessment Methodologies for Responsible Use of AI by Financial Institutions, Available :
<https://www.mas.gov.sg/news/media-releases/2022/mas-led-industry-consortium-publishes-assessment-methodologies-for-responsible-use-of-ai-by-financial-institutions>
- [29] 김규림, 최연경, 노승환, ”혁신의 부스터 AI에 풀드는 금융”, 삼정 KPMG 경제연구원, 서울, 2024.5
- [30] Marcus Evans, Wilson Ang, Jeremy Lua 외 2명(NORTHERN ROSE FULBLIGHT), Singapore releases New Guidelines on the Use of Personal Data in AI Systems[Internet], Available :
<https://www.dataprotectionreport.com/2024/03/singapore-releases-new-guidelines-on-the-use-of-personal-data-in-ai-systems/>, 2024.06.28
- [31] Rajesh Sreenivasan, Regina Liew, Steve Tan 외 5명(National University of Singapore, Singapore), Responsible Use of AI – Guidance from a Singapore Regulatory Perspective [Internet], Available :
<https://law.nus.edu.sg/trail/responsible-use-of-ai/>, 2024.07.01
- [32] PDPC Singapore, “Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems” , 2024.05.01
- [33] 권용수, “일본의 생성형AI 관련 가이드라인 속 저작권 쟁점” , 한국저작권위원회, 2023.06
- [34] Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry, “AI Guidelines for Business Ver1.0” , 2024.04.19.
- [35] 정빛나(연합뉴스), EU, 세계 첫 포괄적 AI규제법 2026년 전면 시행 확정 [Internet], Available : <https://www.yna.co.kr/view/AKR20240521164200098>, 2024.06.30
- [36] 윤성혜(프레시안), 법률 제정속도 느린 중국, AI만은 예외다 [Internet], Available : <https://www.pressian.com/pages/articles/2023110918333428680>, 2024.07.05.
- [37] Grant Gross(CIO Korea), 미국 주 의회들의 잇단 제정 움직임 ··· AI 규제 환경, 이미 복잡해졌다 [Internet], Available :
<https://www.ciokorea.com/news/332582#csidx82dd927d9208b488476d8ffe8f0ae43>, 2024.07.09
- [38] Laney Zhang, China: Generative AI Measures Finalized” , Law Library of

Congress [Internet], Available :

<https://www.loc.gov/item/global-legal-monitor/2023-07-18/china-generative-ai-measures-finalized/>, 2024.07.01.

[39] CGTN, China ready to advance cooperation, uphold mutual support with Arab states [Internet], Available :

<https://news.cgtn.com/news/2024-05-27/China-to-advance-cooperation-uphold-mutual-support-with-Arab-states-1tWYFMbkrSM/p.html>, 2024.07.01.

[40] 김정아(한경 글로벌마켓), 美정부, 진보된 AI모델에 중국 접근 규제방안 고려중 [Internet], Available : <https://www.hankyung.com/article/202405085703i>, 2024.07.03.

[41] 김준희(국민일보), 중국 AI 미국보다 2년 뒤쳐져"... 이유는 美 규제 탓? [Internet], Available : <https://www.kmib.co.kr/article/view.asp?arcid=0019971850>, 2024.07.03.

AI 챗봇 서비스 보안성 체크리스트



구분	점검사항		Y	N
1 AI 시스템의 투명성 및 설명 가능성	AI 모델의 결정 과정이 투명하게 공개되고 있으며, 결과를 설명 할 수 있는 메커니즘이 마련되어 있는가?			
2 AI 시스템의 공정성과 비차별성	AI 모델의 공정성과 비차별성을 보장하기 위한 검토와 수정이 이루어지고 있으며, 알고리즘 편향을 방지하기 위한 절차가 마련되어 있는가?			
3 AI 시스템의 보안성 검증	AI 모델의 보안성을 정기적으로 검증하고, 잠재적 공격에 대비 한 시뮬레이션을 수행하고 있으며, 보안 취약점을 파악하고 필요한 조치를 취하고 있는가?			
4 데이터 출처 의 신뢰성	AI 모델의 학습 데이터는 신뢰할 수 있는 출처에서 수집되며, 데이터 출처와 무결성을 보장하기 위한 절차가 마련되어 있는가?			
5 생성 AI 기술 에 대한 규제	생성 AI 기술의 개발 및 배포에 대한 규제 준수 여부를 확인하 고 있으며 생성 AI 기술이 허위 정보나 악의적 목적으로 사용되지 않도록 방지하는 조치가 마련되어 있는가?			
6 AI 시스템의 공정성과 비차별성	AI 모델의 공정성과 비차별성을 보장하기 위한 검토와 수정이 이루어있으며 알고리즘 편향을 방지하기 위한 절차가 마련되어 있는가?			

개선점: 논문의 시사점을 참고하여 중국과 미국의 AI 규제를 반영한 새로운 체크리스트 질문을 추가해보았습니다. 중국과 미국의 AI 규제 접근 방식을 참고하여 우리나라의 AI 보안 가이드라인을 최신 보안 트렌드에 맞게 강화할 수 있습니다. 이는 AI 시스템의 신뢰성과 안전성을 높이는 데 도움이 될 것입니다.