

이슈보고서 23-13

ISSUE
REPORT

금융산업에서의 인공지능(AI) 활용 방안에 따른 리스크 요인 분석

노성호

금융산업에서의 인공지능(AI) 활용 방안에 따른 리스크 요인 분석

저자 노성호*

최근 가속화되고 있는 인공지능(Artificial Intelligence: AI) 개발 경쟁과 이에 따른 기술의 발전은 금융산업 참여자들의 관심을 불러일으키고 있다. 그러나 높은 관심도의 이면에는 낮은 기술적 이해도와 이에 따른 시장위험의 증가라는 우려 또한 공존하고 있다. 이와 같은 문제의식에 기반하여 본고에서는 구체적인 예시에 기반하여 AI의 작동원리를 소개하고 그에 대응하는 리스크 요인을 분석하였다.

근래에 주목받고 있는 AI의 특징은 머신러닝(machine learning)에 기반한 규칙의 귀납적인 학습(learning)이라고 할 수 있다. 기술적인 면에서 기계의 학습 과정은 두 가지 핵심 요소인 데이터(data)와 알고리즘(algorithm)으로 나누어 볼 수 있다. 특히 현대적인 AI의 급속한 성장은 데이터의 외연적 확장과 알고리즘의 발달에 힘입어 이루어졌는데, 이는 각 요소에 대응되는 근원적 리스크 요인으로서 데이터의 정보적 연관성과 결과의 해석 가능성의 저하라는 부정적 측면을 동시에 시사하고 있다. 이를 구체적으로 금융시장의 참여자별로 살펴보면, 과거 데이터의 편향성으로 인한 예측 오류, 생성된 결과물의 자의적 해석, 과소 추정 또는 거짓 양성반응으로 인한 효율성 감소 등을 예상할 수 있다.

AI의 활용에 따른 근원적 리스크 요인과 그로 인해 예상되는 피해를 최소화하기 위하여 인지, 감시, 책임의 세 단계 정책적 대응 방안을 제시하였다. 우선, 머신러닝 기반 AI의 핵심 요소인 학습 데이터와 알고리즘에 대한 공시를 통하여 사용자의 기술적 이해도를 높일 필요가 있다. 더불어 상시적인 감시체계를 통해 데이터 수집 과정과 학습 알고리즘의 복잡성 및 불투명성으로 인한 피해를 최소화할 필요가 있다. 마지막으로 AI의 결과물에 대한 법적 책임과 규제 체계의 확립은 잠재적 위험을 내재화하여 책임 있는 AI의 개발 및 활용에 기여할 수 있을 것이다.

* 본고의 견해와 주장은 필자 개인의 것이며, 자본시장연구원의 공식적인 견해가 아님을 밝힙니다.

I. 서론

최근 ChatGPT의 인기와 함께 표면화된 대형 자연어처리 모형(Large Language Model: LLM) 개발 경쟁이 이슈화되고 있다. 이와 같은 인공지능(Artificial Intelligence: AI)기술의 발전은 금융산업의 각 분야에서 이를 활용하여 업무의 효율성을 제고하려는 노력을 더욱 가속화하리라 예상된다. 이러한 배경을 바탕으로 본 연구에서는 머신러닝 기반(machine learning-based) AI가 금융산업의 구조와 운용 방식을 어떻게 변화시킬 것이며 이에 수반하는 리스크 요인은 어떤 것이 있는지 금융산업의 참여자 별로 구체적인 예시를 통해 살펴보고자 한다.

최근 금융산업의 여러 분야에서 AI의 활용 방안에 관한 연구가 폭넓게 이루어지고 있는 것은 AI에 대한 높은 관심도를 반영하는 것으로 보인다.⁰¹ World Economic Forum에서 조사한 바에 따르면 금융산업 참여자의 77%가 AI를 금융산업의 핵심 동력(business driver)이 될 것으로 전망하였다(World Economic Forum, 2020). 이는 비단 관심에 그치지 않고 금융산업 참여자 중 75% 이상이 AI를 활용하기 위한 고성능 컴퓨팅 또는 머신러닝 기술에 투자하고 있음이 확인되었다(NVIDIA, 2022). 이와 같은 조사 결과는 AI에 기반한 금융혁신이 지난 수년간 지속해 이루어져 왔으며 현재에도 활발히 진행 중임을 상기시키고 있다.

급격하게 증가한 AI에 대한 관심도의 이면에는 낮은 기술적 이해도와 이에 따른 시장위험의 증가라는 우려 또한 공존하고 있다. CFA Institute는 투자운용업계에서 AI의 활용 현황을 조사한 보고서에서 AI와 그 기반이 되는 머신러닝에 관한 교육에 참여한 인원이 20%가 되지 않음을 지적하였다(CFA Institute, 2019). 이는 실무에서 사용되고 있는 다양한 AI의 기반 기술인 통계적 학습(statistical learning) 이론이 비교적 최근에서야 계산 알고리즘의 발전과 가용한 데이터 범주의 확대에 힘입어 확산할 수 있었던 점과 연관된 것으로 볼 수 있다.

이와 같은 문제의식에 기반하여 본고에서는 근래에 관심이 높아지고 있는 AI의 개념과 작동 원리에 대하여 구체적인 예시에 기반한 논의로 사용자의 이해를 돕고자 한다. 또한 금융산업의

01 AI에 대한 높은 기대와 관심은 비슷한 시기에 진행된 국내외 다양한 연구보고서에서 일관적으로 확인되었다. FINRA(2020)는 미국 내 브로커-딜러(broker-dealer)를 대상으로 한 설문에서 AI가 이미 여러 업무에 광범위하게 활용되고 있음을 확인하였으며, 이는 해외 여러 자산운용사의 AI 활용사례에 대한 조사에서도 확인된 바 있다(금융투자협회, 2020; 우리금융경영연구소, 2022). 국내에서는 이성복(2021)이 AI가 국내 핀테크(FinTech)에 의한 금융혁신에 중요한 동력이 되었음을 밝힌 바 있다. 은행업을 중심으로 머신러닝에 기반한 업무 프로세스 개선 노력이 확산했으며(서정호, 2022a, 2022b) AI를 활용한 가치 창출이 지속해 증가할 것을 예측하고 있다(홍동숙, 2022).

제 분야에서 AI의 적용 예시를 바탕으로 그 효과와 리스크 요인에 대하여 분석하고자 한다.

본고의 구성은 아래와 같다. II장에서 최근 활용도가 증가하고 있는 머신러닝 기반 AI의 핵심 요소와 그에 대응하는 리스크 요인을 설명하고자 한다. III장에서는 구체적인 AI 활용사례를 금융 산업의 참여자별로 세분하여 제시하고 운용 원리와 리스크 요인을 분석하고자 한다. 마지막으로 IV장에서는 앞서 논의된 내용을 바탕으로 시사점과 정책과제를 제안하며 논의를 마치고자 한다.

II. 인공지능의 개념과 리스크 요인

1. 머신러닝 기반 AI의 구성요소

II 장에서는 금융산업을 포함한 여러 분야에서 활용되는 AI의 특징을 분석하고 관련된 리스크 요인을 활용 예시에 따라 살펴보고자 한다. 최근 주목받고 있는 머신러닝 기반 AI의 특징적 요소로는 주어진 자료에 기반하여 기계가 규칙을 스스로 ‘학습’하고 그 결과에 따라 행동하도록 하는 구조를 들 수 있다. 이는 마치 인간이 경험을 바탕으로 귀납적 추론을 통해 최적의 행동을 유추해내는 방식을 기계적으로 구현한 것과 유사한 작동원리를 지닌다.⁰²

‘머신러닝 기반’이라는 이름에서도 알 수 있듯이, AI의 작동원리를 나타내는 핵심 단어는 ‘학습(learning)’이라고 할 수 있다. 여기서 학습은 ‘경험을 통해 얻어지는 행동의 지속적 변화’라는 사전적 의미를 기계적 요소로 대응시킨 개념이라고 볼 수 있다. 행동의 변화는 기계적인 관점에서 의사결정 규칙(decision rule)으로 해석되어 특정한 명령을 인간의 추가적인 지시가 없이도 스스로 수행할 수 있는 근거를 제공한다.

학습 과정의 기계적인 구현 방식이 드러나는 대표적인 예로 최근 그 활용 가능성이 활발하게 논의되고 있는 생성형 AI(Generative AI)와 그 기반이 되는 강화학습(reinforcement learning)을 들 수 있다. 강화학습 과정에서 AI는 데이터(data)의 형태로 주어진 ‘현상’을 분석하고 이에 따라 특정한 행동을 취하면, 미리 정해진 보상을 얻게 된다. 학습 알고리즘(algorithm)에 따라 이를

02 인간의 논리적 사고방식을 크게 연역과 귀납으로 본다면 각각에 대응하는 AI의 분류를 정의할 수 있다. 이에 대해서는 부록에서 자세히 설명하였다.

반복적으로 수행하면서 보상을 최대화하는 행동 규칙을 생성하는 것으로 AI의 개발 목표(일차적인)가 달성되는 것이다. 이 과정을 수행하는 데 있어 핵심이 되는 두 가지 요소인 데이터와 알고리즘에 대하여 아래에서 보다 자세히 살펴보겠다.

가. 데이터의 구분과 활용

AI가 규칙을 만드는 과정에서 필요한 첫 번째 요소인 데이터는 인간의 학습 과정에서 ‘경험’에 해당한다고 볼 수 있다. 규칙을 스스로 만들어내야 하는 특성상 머신러닝 기반 AI의 의사결정 과정은 입력되는 데이터에 크게 의존하게 된다. 따라서 모형의 개발과정에서 제공된 데이터로부터 추론해낼 수 없는 규칙에 대해서는 올바른 답을 기대하기 어렵다고 볼 수 있다.

AI의 개발에 사용되는 데이터는 형태와 목적에 따라서 아래 <표 II-1>과 같이 네 가지로 분류하여 설명하고자 한다.

<표 II-1> 학습 데이터의 분류에 따른 예시

		목적성	
		실험(experimental)	관찰(observational)
기록 형태	정형 (structured)	실험 결과	경제통계 지표, 시장 가격
	비정형 (unstructured)	설문조사 결과	뉴스 보도자료, 기업 공시자료

주: 특정한 목적을 지니고 수집된 데이터를 실험(experimental), 그렇지 않은 경우를 관찰(observational)이라고 한다. 한편 일관된 기록 형태를 정의할 수 있는 데이터를 정형(structured), 그렇지 않은 데이터를 비정형(unstructured)이라고 한다.

데이터의 형태를 구분하는 첫 번째 기준으로 데이터 수집의 목적성을 들 수 있다. 데이터의 수집 단계에서부터 연구 또는 사업상 특정한 목적을 달성하기 위하여 계획적으로 수집된 경우를 실험(experimental) 데이터라고 하자. 이를테면 소비자의 선호를 파악하기 위해 무작위로 두 가지의 광고 시안을 송출하고 그 반응을 수집한 결과는 정해진 연구목적에 달성하기 위해 체계적으로 수집되었다는 점에서 실험 데이터로 볼 수 있다. 이와 반대되는 개념으로 다른 필요로 수집되었거나 사전적으로 정의된 목적이 없이 관측된 자료를 수집한 것을 관찰(observational) 데이터라고 하겠다. 일정한 주기로 관측되는 이자율, 주가 지표, 국내 총생산 지표 등은 특수한 연구목적에 달성하기 위한 것이 아닌 점에서 관찰 데이터에 포함된다.

두 번째 기준으로, 자료가 기록된 형태에 근거하여 정형(structured)과 비정형(unstructured) 데이터로 나누어 볼 수 있다. 정형 데이터는 주로 정량화, 수치화된 자료로서 개별 관측치 사이에 일관된 기록 방식을 정의할 수 있는 경우를 말한다. 앞서 설명한 예시에서 광고를 실제로 시청한 시간을 기록한 자료, 또는 분기별 국내 총생산 지표 등은 모두 일관된 기준에 근거하여 정량화되어 기록하였다는 점에서 정형 데이터의 범주에 들어가는 것으로 볼 수 있다. 반대로 비정형 데이터는 일관된 기록 방식을 정의하기 어려운 경우를 말한다. 예를 들어 새로운 금융상품에 대한 소비자의 의견을 주관식 문답으로 수집하였을 경우, 개개인의 답변은 문자화된 정보로서 직관적인 비교가 어려울 것으로 예상할 수 있다. 나아가 기업 공시자료, 뉴스 보도 등은 특정한 연구 또는 사업 목적을 달성하기 위해 만들어진 자료가 아니라는 점에서 비정형 관찰 데이터로 분류할 수 있다.

머신러닝 기반 AI의 특징으로는 앞서 소개된 네 가지 자료 유형에 구애받지 않고 최대한 많은 자료를 활용하여 모형을 구축하는 것이라고 할 수 있다.⁰³ 특히, 최근 급격하게 이루어진 AI의 발전은 비교적 이해와 활용이 쉬운 실험, 정형 데이터에서 벗어나 점차 관찰, 비정형 데이터를 적극적으로 활용하여 가용한 데이터가 양적으로 크게 증가한 것과 관련이 있다고 하겠다(Kolanovic & Krishnamachari, 2017).⁰⁴

나. 알고리즘의 종류

두 번째 AI의 주요 구성요소로 알고리즘(algorithm)을 들 수 있다. 이는 주어진 자료를 해석하고, 내재한 패턴을 분석하여 이를 바탕으로 결정한 행동에 따른 결과를 정량적으로 평가하는 과정을 의미한다. AI의 핵심 개념을 인간의 행동 양식을 기계적으로 모방하는 것이라고 할 때, 알고리즘은 마치 인간이 스스로 선택의 결과를 돌이켜보고 이를 다음 결정에 반영하는 순차적인 과정과 유사성을 발견할 수 있다.

정량화에 따른 이점은 성과에 대한 객관적인 비교를 가능하게 함에 있다. 더불어 결과의 적합성을 판별하고, AI가 선택한 행동을 스스로 수정하여 복잡한 문제에 직면하였을 경우에도 제한된 시간 안에 가장 효율적인 답을 찾을 수 있도록 한다.

03 이 점에서 근래에 많은 논의가 있었던 빅데이터(big data)의 정의 및 활용 방안과 접점이 있다고 하겠다.

04 근래에는 가용한 데이터의 양적 증대에서 나아가 서로 다른 형태의 데이터를 복합적으로 사용하여 학습시키는 멀티모달 러닝(multimodal learning)에 기반한 AI에 관한 연구가 활발히 진행되고 있다(Ramachandram & Taylor, 2017).

<표 II-2> 머신러닝 알고리즘의 종류

		예시
지도학습 (supervised learning)	분류(classification)	Logistic, SVM, Random Forest
	회귀분석(regression)	LASSO, Ridge, Elastic Net
비지도학습 (unsupervised learning)	군집화(clustering)	K-means, KNN
	요인분석(factor analysis)	PCA, NMF, LDA
혼합형	준지도학습(semi-supervised learning)	
	딥러닝(deep learning)	MLP, CNN
	강화학습(reinforcement learning)	LSTM

주: 약어들은 각각 SVM: Support Vector Machine, KNN: K-Nearest Neighbors, PCA: Principal Component Analysis, NMF: Non-negative Matrix Factorization, LDA: Latent Dirichlet Analysis, MLP: Multi-Layer Perceptrons, CNN: Convolutional Neural Network, LSTM: Long Short-Term Memory를 의미한다. 준지도학습의 예시는 위의 지도학습과 비지도학습의 예시로 제시된 알고리즘의 변형 또는 종합적 사용으로 볼 수 있다.

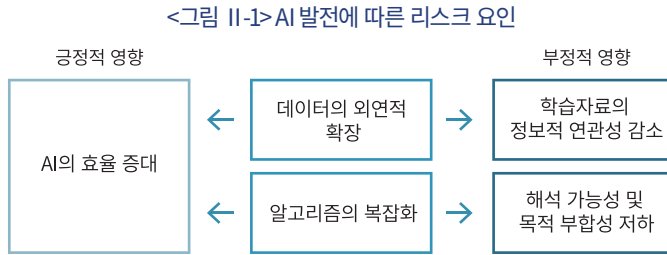
<표 II-2>에서는 대표적인 알고리즘을 형태에 따라 분류하였다. 학습에 사용되는 데이터에 결과값이 포함된 경우는 지도학습(supervised learning)의 범주에 해당하는 알고리즘을 사용할 수 있다. 예를 들어 수익률의 시계열 자료를 활용할 경우, 과거 포트폴리오 구성(원인)과 그에 따른 수익률 실현(결과)이 모두 학습자료에 포함되므로 이를 통해 원인과 결과의 연관관계를 학습할 수 있다. 반면에 비지도학습(unsupervised learning)은 결과가 학습자료에 포함되지 않는 경우를 의미한다. 예를 들어 수많은 종목들의 수익률 시계열 자료만을 가지고 종목별 연관성을 추론하는 문제를 생각해 볼 수 있다. 마지막으로 지도학습과 비지도학습의 요소가 동시에 내재하는 준지도학습(semi-supervised learning), 딥러닝(deep learning), 강화학습(reinforcement learning) 등의 알고리즘은 상대적으로 최근에 개발 및 연구가 활발하게 진행되고 있다.⁰⁵

2. 시의 구성요소에 따른 리스크 요인

금융산업에서 시를 활용하는 과정에서 앞서 살펴본 머신러닝 기반 시의 두 가지 핵심 구성요소가 가지는 리스크 요인은 시의 발전과정과 밀접한 관련이 있는 것으로 보인다. 시의 급속한 성장이

05 준지도학습, 딥러닝, 강화학습 등은 알고리즘의 각 단계에서 지도학습적인 요소와 비지도학습적인 요소가 존재한다는 점에서 혼합형으로 분류하였다. 예를 들어, 딥러닝은 입력된 변수들을 가공하여 연관성이 높은 변수를 생성하고(비지도학습) 이를 활용하여 예측오차를 줄이는(지도학습) 과정으로 구성되었다는 점에서 두 가지 방식이 혼합된 형태로 볼 수 있다.

데이터의 외연적 확장과 학습 알고리즘의 발달에 기인한다고 할 때(Kautz, 2022), 전자는 새로운 데이터를 활용함에 따른 리스크, 후자는 복잡한 알고리즘의 사용에 따른 리스크를 필연적으로 수반한다. 이를 요약하면 <그림 II-1>과 같이 묘사할 수 있다.



주: 최근 머신러닝 기반 AI의 효율을 비약적으로 증대한 두 가지 요인은 각각 잠재적 리스크 요인을 동반하고 있다. 위 그림은 '데이터의 외연적 확장'과 '알고리즘의 복잡화'가 각각 초래할 수 있는 부정적 요인을 나타내었다.

가용한 데이터의 외연적 확장은 AI의 발전을 유도함과 동시에 위험 요소 또한 내포하고 있다. 관찰 데이터의 경우 해결하고자 하는 문제와 정보적 연관성을 보장할 수 없는 관계로 데이터의 유용성이 크게 낮아질 수 있다. 일례로 금융상품에 가입한 소비자의 의견을 조사한 자료는 잠재적 소비자의 입장에서 해당 상품이 어떠한 개선을 필요로 하는지에 대한 정보를 포함하고 있지 않다. 더불어 비정형 데이터의 경우 이를 일관된 형태로 가공하는 과정에서 정보의 변형 또는 손실이 발생할 위험을 내포하고 있다. 고객의 만족도를 분석하는 데 흔히 사용되는 감정분석(sentiment analysis) 모형의 경우 의견란에 적은 문장에서 긍정 또는 부정의 의미를 내포하는 단어의 빈도를 분석하는 과정에서 단어 전후의 문맥을 적절히 반영하지 못할 우려가 있다.⁰⁶

더불어 AI의 개발 과정에서 알고리즘과 관련된 리스크 요인은 주로 학습 과정의 불투명성 또는 복잡성에서 기인한다. 머신러닝 기반 AI는 필연적으로 대량의 데이터를 바탕으로 많은 모수(parameter)를 학습하여 규칙을 형성하는 반면, 계산된 규칙의 효용성을 판별하는 정량적 기준인 목적함수(objective function)는 대체로 비교의 용이성을 위해 일차원적인 특징을 지닌다. 이는 현실의 다면적인 가치를 지나치게 단순화할 위험을 내포하고 있다. 더불어 개발과정에서 참여하는 의사결정 주체의 불안전성 또한 알고리즘의 올바른 선택과 적용을 어렵게 만드는 요인으로 볼 수 있다. 이에 Korinek & Balwit(2022) 등은 AI의 영향을 평가할 때 소셜미디어를 통한

⁰⁶ Loughran & McDonald(2011)는 기업 재무에 관한 공시정보에서 liability라는 단어의 문맥적 의미를 예시로 들었다. 이는 일반적인 해석인 '책임', '문제점'과 같은 부정적인 의미보다 회계 용어로서 '부채'로 해석되는 것이 옳다. 전자의 경우 부정적인 의미이지만 후자의 경우 중립적이거나 "부채가 적정선이다"라는 문맥에서는 오히려 긍정적인 의미를 내포할 수도 있다.

소비자 최적화 광고의 사례와 같이 알고리즘에 기반한 의사결정의 여파가 사회 전반에 영향을 미치는 외부성(externality)이 고려되어야 함을 주장하였다.

III. 금융산업에서의 AI 활용

본 장에서는 금융산업의 참여자를 크게 소비자, 생산자, 중개 기관, 정부로 나누어 각각 대표적으로 활용 가능한 AI의 예시를 살펴보고 그 원리와 리스크 요인을 앞 장에서 정의한 개념을 바탕으로 살펴보고자 한다. 더불어 비정형 문자기반 정보의 요약 및 학습 알고리즘의 대표적인 예시로서 LDA 모형을 소개하고 이를 바탕으로 경기 동향을 분석하는 예시를 통해 앞서 논의한 리스크 요인이 실무적인 차원에서 어떠한 함의를 지니는지 논의하고자 한다.

1. 금융산업 참여자별 AI 활용 예

금융산업은 AI가 활용되고 있는 여러 분야 중에서도 투자 규모와 성장세 면에서 괄목할 만한 성과를 내고 있다. OECD의 보고서에 따르면 2012년부터 2020년 사이 AI 스타트업(start-up)에 대한 금융 및 보험서비스업의 투자액은 약 50억 달러로 상위 5개 산업군에 속하는 비중을 차지한다(OECD, 2021).⁰⁷ 이는 국내에서도 크게 다르지 않아 2026년까지 연평균 38.6% 성장할 것이라는 예측까지 나오고 있다(홍동숙, 2022).

AI에 대한 높은 관심이 공통분모라면 AI 도입에 따른 기대 또는 목적은 금융산업의 참여자 별로 다를 수 있다. <표 III-1>에서는 참여자를 소비자, 생산자, 중개 기관⁰⁸, 그리고 정부로 나누어 AI 활용에 따른 대표적인 기대 효과, 사례, 리스크 요인을 정리하였다.

07 다만, 개별 기업의 AI와 관련된 R&D 투자 비중은 다른 분야에 비해 상대적으로 크지 않은 것으로 보인다. World Economic Forum의 조사에 따르면 설문에 참여한 기관 중 60%가 전체 R&D 투자액 대비 10% 이하만 AI와 관련된 연구에 할애한다고 답하였다(World Economic Forum, 2020).

08 금융 서비스의 생산자로 볼 수도 있으나 본고에서는 금융업의 특성상 거래 중개업의 역할과 비중이 큰 것을 고려하여 분리하여 논의한다.

<표 III-1> 금융시장 참여자별 AI 활용 사례 및 리스크 요인

참여자	대표 사례	기대 효과	리스크 요인	
소비자	포트폴리오 운용 및 주문 집행 자동화	수익성 제고 및 신규 수익 창출	데이터	과거 데이터에 대한 편향성
			알고리즘	복잡성으로 인한 해석의 어려움
생산자	자연어 처리를 활용한 고객 응대 자동화	소비자 경험 개선, 운용 비용 절감 등	데이터	자료의 선택 편향(selection bias)
			알고리즘	자의적 해석의 가능성
중개 기관	빅데이터를 활용한 이상 거래 탐지	준법 감시 및 규제 대응, 신용평가 방식 개선	데이터	학습 데이터의 불투명성
			알고리즘	해석이 어려운 패턴 예측
정부	자연어 처리를 활용한 불완전 거래 등 금융사기 감지	시장 감시 기능의 효율성 및 적시성 제고	데이터	전처리 과정에서의 정보 손실
			알고리즘	과소 추정 또는 거짓 양성

주: 금융시장의 참여자를 소비자, 생산자, 중개 기관, 정부로 분류하였을 때, 각각의 대표적인 AI 활용 사례와 이를 통한 기대효과, 그리고 잠재적인 리스크 요인을 데이터와 알고리즘 측면에서 정리하였다.

가. 소비자

투자자 또는 자산운용사 입장에서 AI는 자산 수익률 예측과 그에 기반한 포트폴리오 재구성을 자동화하는 기회를 제공한다. 이는 기존의 퀀트(quant), 즉 정량적 자산운용 전략의 연장선상에서 이를 양적, 질적으로 발전된 데이터와 알고리즘을 활용해 효율성을 극대화 한 것이라고 볼 수 있다. 많은 자산운용사들은 AI 기반 기술을 응용해서 시장 분석 및 새로운 투자 기회의 발굴 과정을 개선할 수 있었다고 평가하고 있다(FINRA, 2020).

금융투자업에서 AI를 이용한 혁신은 기회인 동시에 근원적인 리스크를 수반하고 있다. 우선, 데이터 측면에서 수익성 예측 모형은 필연적으로 과거 수익률 데이터에 기반하여 학습을 진행하게 되고, 이는 예측의 결과가 과거에 편향적으로 분포할 수 있다는 점을 의미한다. 따라서 팬데믹(pandemic)과 같이 가용한 데이터의 범주 내에 존재하지 않는 사건이 미래에 발생할 경우 기존 예측치의 정확성을 담보하기 어렵다. 더불어 복잡한 예측 알고리즘은 생성된 결과를 인간 개발자와 이용자가 직관적으로 이해하기 어렵게 하는 요인으로 작용한다. 이는 앞서 설명한 예측 오류의 가능성과 더불어 AI에 기반한 의사결정과 그에 따른 결과에 대한 책임소재가 불확실할 수 있음을 의미한다.

나. 생산자

금융 서비스를 제공하는 생산자 입장에서 AI는 고객에게 맞춤형 상품 또는 서비스를 제공할 수 있는 운용 역량을 강화하는 기회를 제공할 수 있다. 예를 들어 최근 활용도가 증가하고 있는 로보어드바이저를 이용함에 있어 고객의 성향에 관한 정보를 내포하는 다양한 자료를 종합하여 선호하는 포트폴리오를 효율적으로 구성할 수 있다.

이와 같이 고객 경험 개선을 위하여 AI를 활용하는 과정에서 발생할 수 있는 리스크 요인은 크게 두 가지로 나누어 생각해 볼 수 있다. 우선 학습의 관점에서 고객의 성향을 분석하는 기반이 되는 데이터가 선택적 편향성(selection bias)을 가질 수 있음을 고려해야 한다. 의견을 수집하는 방식에 따라 특정 세대, 지역, 학력 군의 고객이 배제될 수 있을 뿐만 아니라 애초에 의견을 제시할 의향이 있는 고객의 성향만이 학습에 가용한 데이터로서 수집되었다는 점에서 가용한 데이터가 충분히 대표성을 지니는지를 확인할 필요가 있다. 더불어 학습 알고리즘의 복잡성은 고객 입장에서 제시된 답변에 대한 근거를 직관적으로 파악하기 어렵게 한다. 이는 단기적으로는 AI에 기반한 고객 관리 방식에 대한 불신을 초래하고 장기적으로는 책임소재에 대한 의문을 제기하는 원인이 될 수 있다.

다. 중개 기관

금융 중개 기관의 입장에서 AI는 실시간으로 일어나는 대량의 거래를 감시하고 특이점을 적시에 탐지하는 기술적인 지원을 제공할 수 있다. 흔히 복잡한 규제에 대응하는 혁신 기술로서 레그테크(RegTech)로 불리는 AI 기반 이상 거래 탐지 모형은 이미 고객 확인(customer identification), 자금 세탁 방지(Anti-Money Laundering: AML) 등의 분야에서 연구 및 활용이 활발하게 이루어져 왔다(FINRA, 2018).

AI에 기반한 신기술이 중개 기관의 내부통제를 강화하는 데 활용될 수 있는 한편 운용상 불투명성에 대한 우려 또한 제기할 수 있다. 불투명성은 구체적으로 데이터와 알고리즘 관점에서 나누어 볼 수 있다. 우선, 감시 모형의 특성상 공개된 자료가 아닌 비공개 내부 자료를 사용하여 모형을 학습시키게 되며 이는 외부인으로 하여금 적합성에 대한 검증을 어렵게 하는 요인으로 작용하게 된다. 더불어 복잡한 알고리즘은 결과에 대한 해석을 어렵게 하여 실제로 AI에 기반한 판단이 규정상 모든 검증 절차를 거친 것인지에 대한 의문을 제기할 수 있다.

라. 정부

방대한 양의 데이터를 분석하고 내재하는 패턴을 학습할 수 있는 AI는 정부의 시장 감시 및 규제 기능을 효율화하는데 기여할 수 있다. 이처럼 신기술에 기반한 시장 감독을 의미하는 셉테크(SupTech)는 이미 국내에서도 금융감독원 등에서 도입을 적극 검토하는 것으로 알려져 있다.⁰⁹ 그 대표적인 활용 예시로는 자연어 처리(natural language processing)기술을 응용하여 뉴스, SNS 등에서 실시간으로 발생하는 대량의 문자 정보를 분석하여 불완전 거래 등 금융사기의 가능성을 추정하는 모형을 들 수 있다.

위의 예시와 같은 금융규제 측면에서 AI의 활용이 내포하고 있는 근원적인 리스크 요인 또한 두 가지 측면으로 나누어 볼 수 있다. 우선, 학습에 사용되는 데이터인 문자 기반 정보를 처리하는 과정에서 정보의 손실이 발생할 수 있다. 사용된 단어를 정량화 또는 인코딩(encoding)하는 방식에 따라 문맥적 정보를 소실하고 의미를 단순화할 위험이 있는데 이는 다음 장에서 예시를 바탕으로 보다 구체적으로 설명하고자 한다. 더불어 모형의 예측 결과가 과소추정 또는 거짓 음성(false negative)을 양산할 가능성이 존재한다. 매일 일어나는 수많은 거래에 대비하여 금융사기 사건의 빈도는 낮을 수 밖에 없는데 이는 불균형 표본의 문제(sample imbalance problem)를 야기할 수 있다(Kuhn & Johnson, 2013). 다시 말해 추정하고자 하는 사건이 전체 표본에서 차지하는 비중이 과도하게 적을 경우 지도학습 알고리즘의 학습 효율성이 저하되어 이상 거래일 확률이 과소추정되고 결과적으로 실제 이상거래를 조기에 탐지해내지 못하는 통계적 오류¹⁰가 다수 발생할 수 있다.

2. 사례: LDA를 이용한 경제 전망 분석

이 장에서는 AI를 활용한 정보 추출 및 경기 예측 모형의 사례를 살펴보고 앞서 설명한 리스크 요인에 비추어 시사점을 제시하고자 한다. 경기 동향에 관한 다양한 정보 중 하나의 예시로서 미국 연방준비제도(Federal Reserve System)에서 매년 8회 개최하는 공개시장위원회(Federal Open Market Committee: FOMC) 회의록(minute)을 확률적으로 분석하는 AI 모형을 생각해

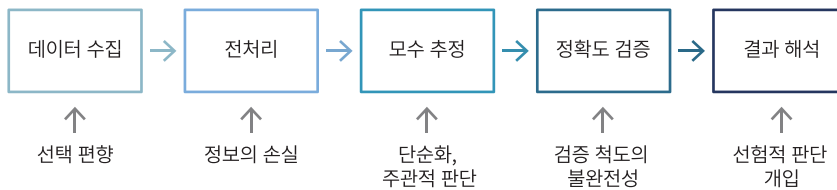
09 금융감독원(2020. 4. 8)

10 실제 이상 거래가 맞음에도 추정값이 아님으로 판별한다는 점에서 거짓 음성(false negative) 또는 II 종 오류에 해당한다고 볼 수 있다.

볼 수 있다. FOMC 회의록은 미국과 국제 경기 동향에 대한 면밀한 연구자료를 바탕으로 작성되었으며 특히 기준금리 결정 과정에 대하여 가장 직접적인 정보를 포함하고 있다(Romer & Romer, 2004). 이에 Edison & Marquez(1998), Meade & Thornton(2012), Hansen, McMahon, & Prat(2018) 등 여러 연구에서 FOMC 회의록의 내용 및 문체를 분석하여 경기 예측의 관점에서 함의를 도출하고자 하였다.

문자에 기반한 정보를 분석하는 대표적인 머신러닝 모형으로는 Latent Dirichlet Analysis(이하 LDA)를 들 수 있다. LDA는 Blei, Ng & Jordan(2003)이 제안한 문자 정보를 확률모형으로 분석하는 머신러닝 기법의 일종으로, 최근 다양한 분야에서 대량의 문서 정보를 요약하는 데 활용되고 있다.

<그림 III-1> LDA 모형의 학습 과정



주: 위의 그림은 LDA 모형의 학습 과정을 단순화하여 표현한 것이다. 각 단계별로 예상 가능한 리스크 요인은 아래에 서술하였다.

<그림 III-1>은 LDA 모형의 학습 과정을 단순화하여 표현한 것이다. LDA뿐만 아닌 모든 머신러닝 기반 모형의 우선 과제는 학습에 필요한 데이터의 수집이라고 볼 수 있다. 본 예시에서는 2009년부터 2022년 사이에 있었던 112회의 FOMC 회의록을 학습에 사용될 데이터로서 수집하였다.¹¹ 해당 기간 동안의 모든 회의록을 사용한 점에서 선택 편향(selection bias)이 발생할 여지가 없다고 볼 수도 있으나 FOMC 회의록의 본질적인 목적이 목표금리 설정을 위한 근거의 서술이라고 할 때, 경기 변동에 영향을 미칠 수 있으나 통화정책의 결정에 의미 있는 영향을 미치지 않는 요소들은 반영되지 않았음을 예상할 수 있다.

수집된 데이터는 문자에 기반한 정보인 만큼 이를 정량화하는 전처리(pre-processing) 및 인코딩(encoding) 과정이 필수적이다. 문서 또는 문장을 수치화하는 과정에서 문맥에 따른 정보가 일부 손실되는 리스크 요인이 존재하는데, 이에 대해서는 다음 절에서 예시를 통해 자세히 설명하고자 한다. 전처리과정 이후 수치화된 데이터를 바탕으로 모형의 모수(parameter)를 학습하게 된다.¹² 모수의 추정 알고리즘은 단어의 분포를 단순화하며 특히 연구자의 판단에 따라

11 FOMC 회의록은 <https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm>에서 확인할 수 있다.

12 보다 자세한 모수의 학습 방식은 부록에서 설명하기로 한다.

결과가 민감하게 변화할 수 있다는 점에서 주의를 기울여야 할 부분이 있다.

계산된 모형을 바탕으로 개발자는 모형의 정확도를 검증하는 단계를 거친다. 이는 일반적으로 학습 데이터 일부를 임의로 제외한 상태에서 모수를 추론하고 이와 같이 구축된 모형이 제외된 데이터와도 일정 수준 이상의 적합도를 나타내는지를 확인하는 과정으로 진행된다. 이때, 적합도를 나타내는 다양한 측도가 있으며 각각은 적합성의 다양한 측면 중 일부만을 측정한다는 점에서 어떤 측도를 선택하느냐에 따라서 최적 모형의 형태가 일관되지 않고 변할 수 있다. 마지막으로 선택된 모형의 분석 결과를 해석하는 과정에서 연구자의 선험적 지식이 관여하게 되는데, 이는 아래 예시를 통해 구체적으로 설명하고자 한다.

가. 데이터의 전처리(pre-processing) 과정

여기서는 학습에 사용된 데이터로서 2009년부터 2022년 사이에 있었던 112회의 FOMC 회의록에서 단어별 빈도를 추출하는 과정을 소개하고자 한다. 흔히 데이터의 전처리(pre-processing)라고 불리는 이 과정은 앞서 설명한 행렬식을 추정하는 데 있어 필수적인 단계라고 할 수 있다.

<표 III-2> 텍스트 전처리 과정 예시

[...] Recent data suggested that real GDP growth in the second half of 2022 was stronger than previously expected, but economic growth was still forecast to slow markedly in 2023 from its second-half pace [...]

recent	data	suggested	real	...
1	1	1	1	...

주 : 위 문장은 2022년 12월 진행된 FOMC 회의 이후 공개된 회의록에서 Staff Overview of Economic Outlook 부분에서 인용하였다. 파란색은 불용어를 나타내며 이는 전처리 과정에서 제외된다.

자료: Federal Reserve System(2022. 12. 14)

<표 III-2>는 문서 데이터의 전처리 과정의 예시를 나타낸다. 우선, 특정 문장이 주어졌을 때, 이를 단어 또는 형태소로 분해한다. 이때, 연결사(예: that), 전치사(예: the) 등은 여러 문장에서 반복적으로 사용되어 특징적인 내용을 파악하는데 있어 혼란을 야기할 수 있어 제외하게 된다.¹³ 이를 모든 문서에 대해서 반복적으로 적용하면 하나의 행렬이 완성되는데 여기서 각 열은 하나의

13 이와 같은 단어를 흔히 불용어(stopword)라고 한다. 불용어는 문맥에 따라서 달라질 수 있는데, 예를 들어 FOMC 회의록에서는 committee와 같은 단어 또한 반복적으로 등장함을 확인할 수 있다. 이와 같은 예시는 일반적인 경우 불용어가 아니지만 FOMC 회의록 분석이라는 연구목적상 불용어로 간주하고 제거하였다.

회의록을 의미하고, 각 행에는 해당 회의록에서 등장한 단어별 빈도가 기록되게 된다. 이와 같은 단어들을 제외하고 단어별 등장 빈도를 적는 방식을 Bag of Words(BoW)라고 한다.

이와 같은 전처리 과정은 문자 기반 정보를 분석하기 위해서 이를 정량화 하는 필수적인 단계라고 하겠다. 그러나 위의 예시에서 볼 수 있듯이 문장을 작은 단위(주로 단어)로 분해하고 정량화 하는 과정에서 필연적으로 정보의 손실이 발생하게 된다. 예를 들어, BoW 방법은 특정 단어가 문장 안에서 몇 번 등장했는지를 파악할 뿐 단어의 전후 문맥을 고려하지는 않게 됨을 알 수 있다.

나. FOMC 회의록의 경제 전망 텍스트 분석 결과

앞서 설명한 LDA 알고리즘을 바탕으로 2009년부터 2022년 사이에 있었던 112회의 FOMC 회의록을 LDA 모형으로 분석하여 5가지 주제를 추출하였다. 주제별로 등장 빈도가 높은 단어를 정리하면 아래의 <표 III-3>과 같다.

<표 III-3> 주제별 고빈도 단어 목록

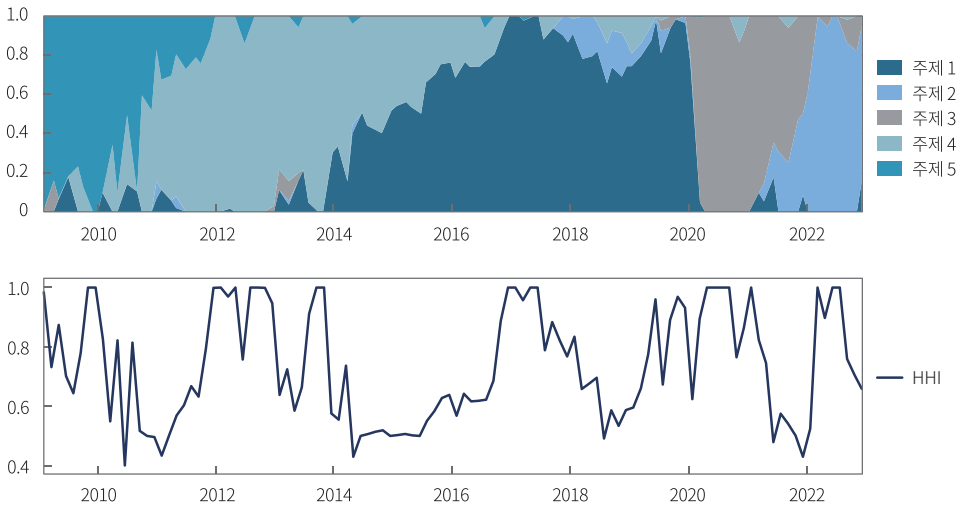
주제 1	주제 2	주제 3	주제 4	주제 5
lower	range	commented	indicated	however
couple	upward	credit	current	continue
investment	remarked	guidance	increase(s)	somewhat
economy	tightening	virus	somewhat	resource
projected	wage	recovery	projection	loans
range	projection	accomodative	housing	housing
appropriate	high(er)	current	slack	next
little	elevated	fiscal	risks	employment
downside	pressures	households	rates	price
developments	strong	businesses	markets	demand
forecast	appropriate	employment	fiscal	appeared
measures	demand	risks	anticipated	slack
term	risk(s)	purchases	purchase(s)	firms
longerrun		support		markets
funds		pandemic		credit

주 : 각 주제별로 추정된 사용 빈도가 높은 단어 순으로 정렬하였으며 같은 어근의 형태가 다른 경우(예: 복수, 비교격)는 하나의 단어로 취급하였다.

자료: 저자 계산

LDA 모형의 학습 결과는 <표 III-3>과 같이 주제별로 자주 사용된 단어를 추출할 뿐, 이를 종합하여 적절한 해석을 붙이는 것은 연구자의 몫이다. 예시로, 주제 1의 경우 투자(investment), 전망(projected), 장기(longerrun) 등의 단어가 자주 등장함을 근거로 장기적이고 완만한 성장과정을 예상하는 것으로 볼 수 있다. 한편 주제 2의 경우 긴축(tightening)에 이어 상향(upward, elevated), (경제적)압박(pressures), 강세(strong) 등의 형용사가 자주 사용되며 동시에 임금(wage), 수요(demand)가 자주 등장하는 것을 근거로 고임금과 수요에 기반한 경기회복과 이에 따른 긴축적 통화정책의 가능성에 관한 논의를 생각해 볼 수 있다. 주제 3의 경우 바이러스(virus), 팬데믹(pandemic) 등의 단어와 회복(recovery), 지원(support) 등의 단어가 함께 사용됨을 근거로 2020년 이후 코로나바이러스로 인한 경기 침체와 부양 방안에 대한 논의임을 비교적 쉽게 파악할 수 있다.

<그림 III-2> FOMC 회의록 주제 분포 및 집중도



주 : 상단은 시기별로 회의록에서 추정된 주제별 확률의 변화를 나타내었으며 하단은 주제들 사이의 분포를 바탕으로 추정한 집중도를 나타낸다.

자료: 저자 계산

<그림 III-2>에서 위 패널은 추정된 5개의 주제의 확률분포가 시계열적으로 변화하여 온 양상을 나타낸다. 글로벌 금융위기 직후에는 주제 5에 대한 논의가 주로 이루어졌으나 2020년 이후 최근 3년간에는 주제 3에서 주제 2로 이동하고 있다. 이를 통해 FOMC에서 팬데믹(pandemic)으로 인한 경기 침체와 이후 회복세를 글로벌 금융위기 직후와 다른 양상으로 인식하고 있음을 추론해 볼 수 있다. 더불어 아래 패널에서는 FOMC 시기별로 특정 주제에 대한 집중도를 허핀달-허쉬만

지수(Herfindal-Hershman Index: HHI)로 표현하였다.¹⁴ HHI가 낮은 시기는 대체로 전망이 일정하지 않거나 다른 주제로 전환되는 시기임을 파악할 수 있다.

다. 결과의 해석 및 시사점

이와 같이 시에 기반한 텍스트 분석모형은 대량의 문자 정보를 빠르고 직관적으로 이해할 수 있게 가공한다는 장점을 가지고 있다. 하지만 그 기반이 되는 데이터와 모형의 잘못된 해석에 따른 위험성을 앞서 설명한 시의 근원적 리스크 요인에 근거하여 살펴볼 필요가 있다.

우선, 문자 기반 데이터의 처리 과정에서 발생하는 정보의 손실이 필연적으로 발생할 수 있다. LDA는 문장 또는 문서를 단어의 집합으로 분류하는 과정에서 특정 단어가 사용된 전후 문맥적 정보를 고려하지 않고 있다. <표 III-3>에서 확인할 수 있듯이 FOMC 회의록의 특성상 주요 단어가 여러 주제에서 반복적으로 사용되어 주제 사이의 명확한 구분을 어렵게 하는 요인으로 작용한다. 이는 경우에 따라 모형의 계산을 어렵게 하는 요인으로 작용한다.

더불어 연구자의 자의적 판단이 개입한다는 점 또한 상기할 필요가 있다. 앞서 살펴본 바와 같이 모형의 결과로 요약된 ‘주제’는 단어의 사용 빈도에 따른 집합에 불과하며 이에 의미를 부여하는 과정에서 연구자의 판단(주관적)이 관여하게 된다. 이처럼 시가 생성한 결과는 많은 경우 배경지식을 가진 인간의 해석이 필요하며, 따라서 시 개발자와 금융, 경제 전문가 사이의 원활한 소통이 전제되어야 그 활용성이 극대화될 수 있을 것이다.

IV. 정책적 함의 및 결론

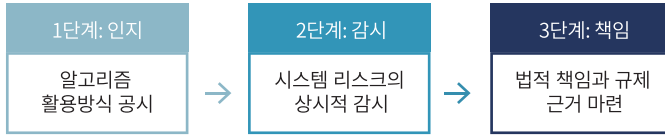
1. 정책적 함의

앞서 설명한 머신러닝 기반 시의 핵심 구성요소와 예시를 바탕으로 이 장에서는 상응하는 리스크 요인을 줄일 수 있는 정책적 함의를 모색해 보고자 한다. 아래 <그림 IV-1>에서 제시한 세 가지

14 흔히 독점성의 여부를 판별하는 기준이 되는 HHI는 수식으로 표현하면 다음과 같다. 문서 d_i 에서 추정된 주제 τ_k 의 빈도를 $P(\tau_k|d_i)$ 라고 할 때, HHI는 이들의 제곱합인 $HHI_i = \sum_{k=1}^5 P(\tau_k|d_i)^2$ 으로 나타낼 수 있다. 따라서 HHI가 1에 가까울수록 하나의 주제가 다른 주제들에 비하여 높은 확률로 집중되어 있음을 의미한다.

항목은 AI로 인한 리스크의 발생 원인을 사전에 방지하고 사후적으로 피해를 보완하는 일종의 3 단계 방어체계에 해당한다고 볼 수 있다.

<그림 IV-1> AI 리스크 요인 억제를 위한 방안



첫 번째 단계로 기술에 대한 사용자의 이해도를 높이기 위하여 AI의 기반 데이터와 알고리즘에 대한 공시제도를 강화하는 것을 고려할 수 있다. 앞서 설명한 바와 같이 AI의 학습에 사용되는 데이터와 알고리즘은 사용자 입장에서 결과만을 가지고 그 구체적인 내용을 파악하기 어려운 것이 사실이다. 이와 같은 불투명성에 근거한 리스크를 최소화하기 위해서는 비전문가 입장에서도 직관적으로 이해할 수 있는 양식을 기반으로 한 공시제도가 전제되어야 할 것이다. 최근 유럽의회에서 재정 준비단계에 들어간 AI 규제법 중 특히 범용 AI(General-purpose AI)의 투명성에 관한 논의¹⁵는 AI의 불투명성과 그로 인한 위험성에 대한 공감대가 형성되고 있다는 점을 시사하고 있다.

적용된 AI에 대한 이해를 바탕으로 두 번째 단계에서는 시스템 리스크의 상시적인 감시(monitoring) 방안을 도입할 필요가 있다. 앞서 설명한 리스크 요인은 머신러닝 기반 AI의 근원적인 요소로서 공시제도와 같은 소극적인 규제로는 통제 및 제거가 불가능할 수 있다. 금융안전위원회(Financial Stability Board: FSB)는 AI와 머신러닝이 금융기관 간 그리고 시장 간의 연결성을 강화하여 시스템 리스크를 증폭시킬 수 있다고 지적한 바 있다(FSB, 2017). 특히, AI의 활용이 사회 전반에 외생적(externality) 영향을 끼친다는 점에서 리스크 요인의 상시적인 감시는 시장원리보다는 거시적인 관점에서 규제를 통해 이루어질 필요가 있다.

마지막으로, AI 활용에 의한 결과에 대한 법적 책임(accountability 또는 liability)과 이에 따른 규제 체계 확립을 논의할 필요가 있다. 올바른 이해와 감시가 전제되었다 하여도 의도치 않은 손해의 발생 확률을 완벽하게 통제하기는 어려울 것이다. 발생한 손해에 대하여 그 근원을 정확히 파악하고 책임소재를 분명하게 하는 것은 AI의 개발자와 사용자가 위험을 내재화(internalize)하여 개발 및 활용을 보다 신중하게 하는 효과를 기대할 수 있을 것이다.

15 European Parliament(2023. 5. 11)

2. 결론

본고에서는 금융산업의 제 분야에서 AI 활용 사례와 리스크 요인을 구체적인 예시를 통하여 분석하였다. 우선 AI에 기반한 혁신의 본질과 파급효과를 올바르게 이해하기 위하여 기술적인 면에서 AI의 핵심 요소인 데이터와 알고리즘의 개념을 소개하였다. 이를 바탕으로 최근 급격하게 이루어진 AI의 발전은 가용한 데이터의 외연 확대에 힘입어 이루어졌음을 확인하였다. 반면 AI 알고리즘의 대형화, 복잡화는 대체로 운영방식의 불투명화를 수반하여 금융산업 내의 리스크 요인으로 작용할 수 있다는 점을 지적하고자 하였다.

새로운 기술의 도입은 항상 기회와 위기라는 양면적 속성을 내포하고 있다. 근래에 널리 활용되고 있는 AI 관련 기술 또한 금융산업의 운영 효율 제고와 금융위기 확산이라는 양면을 동시에 지니고 있다. 이에 AI 활용의 순기능을 극대화하면서 동시에 위험을 최소화하기 위하여 금융산업 참여자 사이의 소통과 노력이 어느 때보다 필요하다고 하겠다.

참고문헌

- 금융감독원, 2020. 4. 8, 인공지능(AI)과 빅데이터를 활용한 금융감독 디지털 전환 추진, 보도자료.
- 금융투자협회, 2020, 『글로벌 자산운용산업의 인공지능(AI) 기반 혁신 동향 및 사례』.
- 서정호, 2022a, 국내은행의 인공지능 도입현황과 경영과제, 『KIF 금융포커스』.
- 서정호, 2022b, 금융업의 인공지능 활용과 정책과제, 『KIF 정책분석보고서』, 한국금융연구원.
- 우리금융경영연구소, 2022, 글로벌 자산운용사의 AI, 빅데이터 활용 사례와 시사점, 『Global Investment Focus』.
- 이성복, 2021, 『핀테크에 의한 금융혁신 양상과 시사점』, 자본시장연구원 연구보고서 21-03.
- 홍동숙, 2022, 금융 AI 시장 전망과 활용 현황: 은행권을 중심으로, 『CIS 이슈리포트』 2022 1호, 한국신용정보원.
- Blei, D. M., Ng, A. Y., Jordan, M. I., 2003, Latent dirichlet allocation, *Journal of Machine Learning Research* 3(Jan), 993-1022.
- CFA Institute, 2019, *AI Pioneers in Investment Management*.
- Edison, H. J., Marquez, J., 1998, US Monetary Policy and Econometric Modeling: Tales from the FOMC Transcripts 1984-1991, *Economic Modelling* 15, 411-428.
- European Parliament, 2023. 5. 11, AI Act: a step closer to the first rules on Artificial Intelligence. <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>
- Federal Reserve System, 2022. 12. 14, Minutes of the Federal Open Market Committee.
- Financial Stability Board(FSB), 2017. 11. 1, Artificial intelligence and machine learning in financial services: Market developments and financial stability implications.
- FINRA, 2018. 9, *Technology Based Innovations for Regulatory Compliance (“RegTech”) in the Securities Industry*.
- FINRA, 2020, *Artificial Intelligence (AI) in the Securities Industry*.

- Hansen, S., McMahon, M., Prat, M., 2018, Transparency and Deliberation within the FOMC: A Computational Linguistics Approach, *The Quarterly Journal of Economics* 133(2), 801-870.
- Kautz, H., 2022, The third AI summer: AAAI Robert S. Englemore memorial lecture, *AI Magazine* 43(1), 105-125.
- Kolanovic, M., Krishnamachari, R. T., 2017, Big Data and AI Strategies: Machine Learning and Alternative Data Approach to Investing, J.P.Morgan Global Quantitative & Derivatives Strategy.
- Korinek, A., Balwit, A., 2022, Aligned with Whom? Direct and Social Goals for AI Systems, NBER Working Paper Series 30017.
- Kuhn, M., Johnson, K., 2013, Applied predictive modeling, New York: Springer.
- Loughran, T., McDonald, B., 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *The Journal of Finance* 66(1), 35-65.
- Meade, E. E., Thornton, D., 2012, The Phillips Curve and US Monetary Policy: What the FOMC Transcripts Tell US, *Oxford Economic Papers* 64, 197-216.
- NVIDIA, 2022, State of AI in Financial Services: 2022 Trends, Survey Report.
- OECD, 2021. 9., Venture Capital Investments in Artificial Intelligence, *OECD Digital Economy Papers* No. 319.
- Ramachandram, D., Taylor, G. W., 2017, Deep multimodal learning: A survey on recent advances and trends, *IEEE signal processing magazine* 34(6), 96-108.
- Romer, C. D., Romer, D. H., 2004, A new measure of monetary shocks: Derivation and implications, *American Economic Review* 94(4), 1055-1084.
- World Economic Forum, 2020, *Transforming Paradigms: A Global AI in Financial Services Survey*.

<부록 1> AI의 분류

이 장에서는 AI의 특징을 그 발전과정과 분류 체계에 근거하여 살펴보고자 한다. AI의 궁극적인 목표가 인간의 사고방식을 모방하여 문제를 해결하는 기계를 구현하는 것이라고 한다면, AI의 분류 또한 인간의 사고체계에 기반하여 정의할 수 있을 것이다.

인간의 논리적 사고방식을 크게 연역적 추론과 귀납적 추론으로 나누어 이에 AI의 추론 형식을 대응시킬 수 있다. 이 경우, 규칙 기반(rule-based) AI는 연역적 추론과 유사한 형태를 의미한다고 볼 수 있다. 이는 개발자가 직접 정해진 규칙(rule)을 기계가 실행할 수 있는 알고리즘(algorithm)으로 치환하여 이에 따라 업무를 자동화하는 형태를 의미한다. 이와 같은 AI는 정해진 규칙이 있을 때 개발이 상대적으로 쉬운 특징으로 인해 그 역사가 오래된 것으로 알려져 있다.¹⁶

이와 대비되는 형태로서 머신러닝 기반(Machine Learning-based) AI는 주어진 자료에 기반하여 기계가 규칙을 스스로 ‘학습’하고 그 결과에 따라 행동하도록 하는 구조를 말한다. 이는 행동 규칙을 외부에서 주어진 것이 아닌 관찰된 현상, 즉 데이터(data)를 통해서 스스로 생성한다는 점에서 귀납적 추론 방식과 유사한 것으로 볼 수 있다. 인간 개발자가 모든 경우의 수를 파악하고 의사결정 과정을 세밀하게 지시할 필요 없이 기계가 스스로 규칙을 정립할 수 있다는 점에서 머신러닝 기반 AI는 규칙 기반 AI에 비해 그 범용성이 높다고 볼 수 있다. 동시에 이와 같은 복잡한 학습 알고리즘을 실현하기까지 오랜 연구개발 과정이 소요되어 비교적 최근에 들어와서야 상용화될 수 있었다.¹⁷ 이를 정리하면 <부록 표 1>과 같다.

16 규칙 기반 AI의 일종인 전문가 시스템(expert system)의 첫 개발 및 활용은 1968년에 개발된 유기화합물을 분석하는 프로그램인 Dendral로 알려져 있으며 비슷한 시기에 다양한 분야에서 연구 또는 상업적 목적으로 전문가 시스템 AI의 개발이 활발하게 이루어졌다(Kautz, 2022).

17 다만 학습에 기반한 정보처리와 자동화 노력은 비즈니스 수요에 따라서 산발적으로 진행되어 왔다. 일례로 Walmart는 1950년 헬리콥터를 이용해 주차장의 활용도를 관찰 및 분석, 이를 통해 소비자의 동향을 파악하여 부동산 투자 효율 제고를 시도하였다(Kolanovic & Krishnamachari, 2017). 이는 최근 위성사진을 활용해 데이터 수집부터 인코딩(encoding)과 분석까지 수행하는 자동화 알고리즘을 이용할 수 있게 되어 비로소 현대적 의미의 AI 기반 비즈니스 모형으로 진화한 것으로 볼 수 있다.

<부록 표 1> 추론방식에 따른 AI의 분류

	규칙 기반 AI	머신러닝 기반 AI
추론 방식	연역적 추론	귀납적 추론
적용 범위	제한적	다목적 활용 가능
복잡도	상대적으로 낮음	매우 높음
활용 예시	전문가 시스템(expert system)	생성형 AI(Generative AI)

주: 편의상 귀납적 추론의 요소를 내재하는 모든 AI 형태를 머신러닝 기반 AI로 분류하였으나 생성형 AI의 경우 귀납과 연역적 추론이 순차적으로 반복되는 구조로도 해석할 수 있다.

생성형 AI를 포함하여 최근 그 활용방안에 대하여 많은 논의가 이루어지고 있는 강화학습(reinforcement learning)과 같이 복합적인 형태의 AI의 경우 스스로 학습한 규칙에 근거하여 행동하고 이를 수정하는 과정을 반복적으로 수행한다는 점에서 귀납과 연역이 공존하는 것으로도 볼 수 있다(Kautz, 2022). 다만, 논의를 단순화하기 위하여 강화학습의 경우에도 귀납적 추론의 요소를 포함한다는 점에서 머신러닝 기반 AI로 분류할 수 있다. 이러한 관점에서 볼 때, 근래에 널리 응용되고 있는 AI 기술을 이해하는 데 있어 머신러닝의 원리와 그에 상응하는 리스크 요인에 대한 이해가 필수적이라고 하겠다.

<부록 2> LDA 모형의 모수(parameter) 학습

LDA 모형의 기본 구조는 하나의 문장, 문단, 또는 문서를 형태소(token) 또는 단어의 집합으로 분해하고, 이를 다시 소수의 ‘주제(topic)’로 분류하여 주제별로 특정 단어가 사용될 확률을 추정하는 것이다.

이를 수식으로 나타내면 다음과 같다. 가용한 학습 데이터로 D 개의 문서 d_1, d_2, \dots, d_D 가 있다고 하자. 각 문서 d_i 는 T 개의 ‘주제’ 중 하나에 대하여 작성된 것이다. 이때, 문서 d_i 가 특정 주제 τ 에 대하여 논의하고 있을 확률은 $P(\tau|d_i)$ 로 나타낼 수 있다. 이제 특정 주제 τ 에 대하여 문서를 작성하였을 경우 단어 w 가 사용될 확률을 $P(w|\tau)$ 라고 나타내기로 한다. 이제 베이즈 정리(Bayes Rule)를 사용하여 문서 d_i 에서 단어 w 가 사용될 확률을 다음과 같이 나타낼 수 있다.

$$P(w|d_i) = \sum_{\tau=1}^T P(w|\tau)P(\tau|d_i) \quad (1)$$

이러한 관계식을 바탕으로 D 개의 문서로 구성된 학습 데이터 전체에서 사용된 모든 단어 w_1, w_2, \dots, w_V 의 분포를 행렬식을 이용하여 표현할 수 있다. 우선 벡터 θ_i 와 ϕ_k 를 다음과 같이 정의하기로 한다.

$$\begin{aligned} \theta_i &= (P(\tau_1|d_i), P(\tau_2|d_i), \dots, P(\tau_T|d_i)) \\ \phi_k &= (P(w_1|\tau_k), P(w_2|\tau_k), \dots, P(w_V|\tau_k))' \end{aligned}$$

여기서 θ_i 는 문서 d_i 가 특정 주제를 논의하고 있을 확률의 분포를 나타낸 $1 \times T$ 행벡터, ϕ_k 는 주제 τ_k 가 각 단어를 사용하는 빈도를 나타낸 $V \times 1$ 열벡터임을 알 수 있다. 이제 관계식 (1)을 각 문서 d_i 에 포함된 모든 단어에 대하여 적용하고 이를 행렬구조로 나타내면 데이터 전체에 분포하는 단어의 빈도를 다음과 같이 풀이할 수 있다.

$$\Phi\Theta = (\phi_1, \phi_2, \dots, \phi_T)(\theta_1, \theta_2, \dots, \theta_D)'$$

위의 행렬식에서 좌변은 $P(w_j|d_i)$ 를 (j, i) 번째 원소로 가지는 $V \times D$ 행렬이 된다. 따라서 이는 주어진 학습 데이터 D 에 분포하는 단어의 빈도를 구조화한 행렬¹⁸임을 알 수 있다. 이 때, LDA 모형은 Θ 와 Φ 가 각각 독립적인 디리클레(Dirichlet) 분포를 따른다고 가정한다. 즉, $\theta_i \sim Dir(\alpha)$, $\phi_k \sim Dir(\beta)$ 일 경우, 주어진 자료를 활용하여 α 와 β 를 추정하면 각 문서가 어떠한 주제를 얼마의 확률로 논의하고 있는지를 계산할 수 있다.

18 이를 문서 단어 행렬(Document-Term Matrix: DTM)이라고 한다.