

# KEDI 글로벌 생성형 AI 지수 산출방법 (Methodology Book)

April 2024

## 1. 지수 소개

- A. 본 문서는 KEDI Global Generative AI Index (이하 “KEDI 글로벌 생성형 AI 지수”)를 산출하고 관리하는 방법론에 대한 이해를 돕기 위하여 작성되었습니다.
- B. KEDI 글로벌 생성형 AI 지수는 시가총액과 거래대금 기준을 만족하는 NYSE 및 NASDAQ에 상장 종목들 중 Large Language Model (LLM) 기술을 공시보고서와 뉴스 자료에 적용하여 생성형 AI 산업과 관련된 키워드와 유사도가 점수가 높은 종목을 선정합니다. 이후 유사도 점수 가중 방식으로 비중을 구성한 지수입니다.
- C. 키워드를 기반으로 한 LLM적용은 기업의 재무실적과 직접적인 연관성이 높지 않을 수 있는 한계가 존재합니다. 따라서 분기별 실적 설명자료를 토대로 KEDI 지수위원회의 검토를 거쳐 연관성이 낮은 종목은 제외할 수 있으며, 이 경우 차 순위 종목이 최종 포트폴리오로 선정됩니다.
- D. KEDI 글로벌 생성형 AI 지수는 가격지수(Price Return) 입니다.

구분	기준시점	기준지수	발표일	산출시간	리밸런싱
일반전략지수	2018.01.04	1,000pt	2024.02.	8:30AM (KST)	분기별

## 2. 지수 구성 방법

### A. 기초 투자 유니버스 선정

KEDI 글로벌 생성형 AI 지수의 유니버스를 구성하기 위하여, 다음과 같은 요건에 부합하는 기초 투자 유니버스를 선정합니다.

- i. NYSE 및 NASDAQ에 상장된 보통주 또는 ADR
- ii. 정기변경일 기준 시가총액 10억 달러 이상인 기업
- iii. 정기변경일 기준 3개월 일 평균 거래대금이 1백만 달러 이상인 기업
- iv. KEDI-Akros Industry Classification Standard (KAICS) Primary Level 3 기준 다음 분류에 해당하는 기업 중 시가총액 상위 100위에 포함되는 기업
  - 3344: Semiconductor and Other Electronic Component Manufacturing
  - 5132: Software Publishers
  - 5162: Media Streaming Distribution Services, Social Networks, and Other Media Networks and Content Providers
  - 5182: Computing Infrastructure Providers, Data Processing, Web Hosting, and Related Services

\* LLM을 이용한 유사도 평가의 효율성을 위하여 전기전자 하드웨어 및 소프트웨어 기업이 포함된 KAICS Primary Level 1 기준 [33: Manufacturing] 업종과 [51: Information] 업종을 대상으로 키워드 'Generative AI'와 유사도가 높은 종목들이 분포된 Level 3 세부 업종에 해당합니다.

\*\* KEDI Akros Industry Classification Standard (KAICS)는 아크로스테크놀로지스에서 개발한 Akros Industry Classification Standard (AICS)를 기반으로, KEDI 지수위원회의 검토를 거쳐 조정된 분류 체계입니다. AICS는 미국 NAICS(North American Industry Classification System)의 산업분류 체계를 기반으로 기업 개요와 매출구성을 대규모 언어모델(LLM)을 활용하여 종목별로 구분한 업종 분류입니다

**B. 키워드 선정**

공신력 있는 분석 기관 Wells Fargo에서 출간된 "Special Report: The ascent of generative AI" 와 McKinsey&Company에서 출간된 "The economic potential of generative AI" 자료를 활용하여 생성형AI 산업을 반영하는 키워드를 선정하였습니다. 키워드는 산업의 성장 및 변화에 따라 혹은 국가기관 및 연구소에서 인정하는 출처 자료의 변경에 따라 KEDI 지수위원회의 자문을 통하여 변경될 수 있습니다.

키워드	"Special Report: The ascent of generative AI" 자료 기준	"The economic potential of generative AI" 자료 기준
생성형 AI (Generative AI)	pg 1 "The recent emergence of generative AI capabilities has generated considerable investor enthusiasm surrounding the new disruptive technology" pg 14 "We expect that software companies will make new investments that move quickly to include generative AI in their product sets."	pg 4 "ChatGPT, GitHub Copilot, Stable Diffusion, and other generative AI tools that have captured current public attention are the result of significant levels of investment in recent years that have helped advance machine learning and deep learning." pg 8 "We have used two complementary lenses to determine where generative AI with its current capabilities could deliver the biggest value and how big that value could be."

**C. 데이터 수집**

모든 데이터는 정기변경일을 기준으로 공시보고서와 뉴스 자료를 수집합니다. 특히 사업보고서의 경우 비중결정일을 기준으로 NYSE 및 NASDAQ에 상장된 보통주에 대해서는 미국 증권거래위원회에 공시되어 있는 최신 날짜의 분기 보고서 (10-Q) 자료를 사용합니다. 단, 어떠한 사유로 분기 보고서 (10-Q) 자료가 존재하지 않는 경우, 연간 보고서 (10-K)로 대체하여 사용합니다. NYSE 및 NASDAQ에 상장된 ADR의 경우 연간 보고서 (20-F)를 사용합니다.

**D. Large Language Model (이하 "LLM") 기법을 통한 필터링**

**i. 1차 Large Language Model (이하 "LLM")을 이용한 필터링**

기초 투자 유니버스에 해당하는 기업들에 대해서 국가기관 자료, 사업보고서, 검색 및 뉴스 자료를 기반으로, 모든 기업들이 키워드에 대해 얼마나 높은 유사도를 가지는지 LLM을 통하여 다음과 같이 측정합니다 (**절대적 기준의 유사도 평가**):

$$L(K_j, P_k) = \begin{cases} 1.0, & \text{기업정보와 키워드의 유사도가 "매우 높음"} \\ 0.75, & \text{기업정보와 키워드의 유사도가 "높음"} \\ 0.5, & \text{기업정보와 키워드의 유사도가 "중간"} \\ 0.25, & \text{기업정보와 키워드의 유사도가 "낮음"} \\ 0.0, & \text{기업정보와 키워드의 유사도가 "매우 낮음"} \end{cases}$$

L:	LLM 기반 절대적 유사도 점수	개별 기업과 개별 키워드 간의 LLM 기반 유사도 점수
K <sub>j</sub> :	키워드	개별 키워드
P <sub>k</sub> :	기업 정보	개별 기업에 대한 자료, 보고서, 검색 및 뉴스 정보

**ii. 2차 Large Language Model (이하 "LLM")을 이용한 필터링**

1차 LLM 필터링을 통과한 기업들에 대해서 국가기관 자료, 사업보고서, 검색 및 뉴스 자료를 기반으로, 키워드에 대해 어떠한 기업이 다른 기업 대비 높은 유사도를 가지는지 LLM을 통하여 다음과 같이 측정합니다 (**상대적 기준의 유사도 평가**):

$$M(K_i, P_j, P_k) = \begin{cases} 1.0, & \text{기업j와 기업k를 비교했을 때 기업j의 유사도가 기업k에 비해 "높음"} \\ 0.0, & \text{기업j와 기업k를 비교했을 때 기업j의 유사도가 기업k에 비해 "낮음"} \end{cases}$$

M:	LLM 기반 상대적 유사도 점수	기업 대 기업 LLM 기반 키워드 유사도 점수
K <sub>i</sub> :	키워드	개별 키워드
P <sub>j</sub> :	기업 정보	기업에 대한 자료, 보고서, 검색 및 뉴스 정보
P <sub>k</sub> :	기업 정보	기업k에 대한 자료, 보고서, 검색 및 뉴스 정보

**iii. 1차 LLM 필터링 및 2차 LLM 필터링 적용 방법**

기초 투자 유니버스에 해당하는 모든 종목들에 대해서 1차 LLM 기법을 통하여 절대적 기준의 유사도 점수를 측정합니다. 이때, 절대적 기준의 유사도 점수가 0.5점 보다 낮은 점수로 측정된 기업들은 해당 기업 정보와 유사도가 "낮음" 혹은 "매우 낮음"으로 간주되어 투자 유니버스에서 제외됩니다.

기초 투자 유니버스에 해당하며 1차 LLM 필터링을 거쳐 절대적 기준의 유사도 점수를 통과한 기업들에 대해서는 2차 LLM 기법을 통하여 상대적 기준의 유사도 점수를 측정합니다. 이때, 1차 LLM 필터링을 통과한 모든 기업들에 대하여, 각 기업들이 첫 번째부터 마지막 순위에 이르기까지 모든 가능한 조합에 대한 상대적 비교 평가가 이루어집니다. 상대적 기준의 유사도 점수를 **최종 유사도 점수**로 간주합니다.

**E. 개별 종목의 편입 대상 선정 기준**

앞서 산출된 최종 유사도 점수를 기준으로 최종 편입 대상을 선정합니다.

**편입 대상 종목 선정**

기초 투자 유니버스에 해당하는 기업들에 대해서, 키워드로 산출된 최종 유사도 점수가 높은 순서로 상위 30종목을 선정합니다.

다만, 키워드 기반의 LLM 기술로 선정된 기업이 해당 산업과 재무적 연관성이 낮을 수 있습니다. KEDI 지수위원회는 최종 의사결정을 가진 기구로써 공시보고서와 분기별 실적 자료 등을 기반으로 연관성이 낮다고 판단되는 종목은 제외할 수 있으며, 이 경우 차 순위 종목이 최종 포트폴리오로 선정됩니다.

**F. 개별 종목의 편입 비중 산정 방식**

KEDI 글로벌 생성형 AI 지수는 다음과 같은 기준으로 편입 비중을 결정합니다.

키워드	총 종목 수	총 포트폴리오 비중	편입 비중 산정 방식
생성형 AI (Generative AI)	30	100%	최종 유사도 점수 기준 상위 30종목 선정 <u>유동시가총액</u> 을 기준으로 비중 할당 (최대비중: 15%)

단, 정기 변경 시 특정 종목의 비중이 표에 기재된 최대 비중을 초과할 경우, 해당 종목의 비중을 최대 비중으로 제한하며, 해당 종목의 초과분을 차 순위 종목들에 대하여 유동시가총액을 안분하여 배분합니다.

### 3. 지수 산출 방법

#### A. 지수 값의 산출

KEDI 글로벌 생성형 AI 지수는 다음 수식을 이용하여 유동 시가총액을 가중한 지수입니다. 지수 산출 기준일은 2018년 1월 4일이며, 기준지수는 1,000으로 설정하였습니다.

$$I(t = 0) = 1000$$

$$I(t) = \frac{MV_t}{MV_{t-1}} \times I(t - 1) = \frac{\sum_i P_{i,t} \times Q_{i,t}}{\sum_i P_{i,t-1} \times Q_{i,t-1}} \times I(t - 1)$$

t	특정 시점 t
I(t=0):	지수 I의 기준 값 (1000)
I(t):	t 시점 지수 I의 산출 값
I(t-1):	t-1 시점 지수 I의 산출 값
P <sub>i,t</sub> :	t 시점 종목 i의 종가
Q <sub>i,t</sub> :	t 시점 종목 i의 보유량
MV <sub>i,t</sub> :	t 시점 보유 종목들에 대한 지수 내 시가총액의 총 합

#### 세부내용

위 수식을 이용하여 t 시점에 대한 지수 I의 값을 산출할 수 있습니다. 지수 I의 값에 대한 산출은 장 종료 시점(t-1)과 다음 영업일 개장 시점(t) 사이에 발생하는 종목의 시장가치 변화를 활용하여 이루어집니다. 따라서 해당 수식은 주식 분할 등 시장 중립적인 이벤트에 영향을 받지 않습니다.

### 4. 지수 관리 및 유지

#### A. 정기변경일

- i. 모든 영업일은 미국 주식시장 영업일을 기준으로 합니다.
- ii. 비중결정일 (Determination Date): 분기별 마지막 영업일 D를 기준으로 종목을 선정하며, 지수의 비중을 확정합니다.
- iii. 수행일 (Implementation Date): D+3 영업일 장 종료 시점에 정기 변경을 수행합니다.

#### B. 수시변경

합병/분할, 상장폐지 등 특정 기업 이벤트 발생에 대한 수시변경은 일반적으로 「기업 이벤트 적용 방법론」을 따릅니다. 영업 손실 등 운용상의 중대한 사유가 있을 경우 KEDI지수위원회 검토를 거쳐 편입 및 차 순위 종목을 편입할 수 있습니다.

개별 종목의 편입 비중이 T-5 영업일부터 T 영업일까지 연속적으로 30%를 초과하는 경우,

T 영업일과 T+2 영업일을 각각 비중결정일과 수행일로 설정하여, 수행일의 미국 주식시장 종료 시점 및 지수 산출 이후에 2. 지수 구성 방법 F. 개별 종목의 편입 비중 산정 방식에서 제시하는 비중으로 제한합니다. 나머지 종목들도 직전 정기 변경일 비중으로 조정합니다.

---

#### <책임의 한계와 법적고지>

이 문서를 포함한 KEDI(한국경제신문지수)의 지수 컨텐츠는 (주)한국경제신문과 서면동의 없이 복제, 전송 또는 배포할 수 없습니다. KEDI의 지수 컨텐츠는 정보 제공이 목적이며 산출 및 공표와 관련한 정확성 및 완전성을 보증하지 않습니다. (주)한국경제신문은 KEDI 지수를 기반으로 한 제 3 자의 투자 상품에 대한 투자 의견을 표명하지 않으며, 지수 이용자와 제 3 자간에 서비스를 매개하여 발생한 분쟁에 관여할 법적 의무가 없습니다. 또한 투자행위 등으로 발생한 손실 등에 대하여 책임을 부담하지 않습니다.