

KEDI 미국 AI 소프트웨어 지수 산출방법 (Methodology Book)

January 2024

1. 지수 소개

- A. 본 문서는 KEDI US AI Software Index (이하 "KEDI 미국 AI 소프트웨어 지수")를 산출하고 관리하는 방법론에 대한 이해를 돕기 위하여 작성되었습니다.
- B. 본 지수는 시가총액과 거래대금 기준을 만족하는 NYSE 및 NASDAQ에 상장 종목들 중 Large Language Model (LLM) 기술을 공시보고서와 검색 및 뉴스 자료에 적용하여 AI 소프트웨어와 관련된 키워드와 유사도가 점수가 높은 종목을 선정합니다. 이후 유사도 점수가 높은 종목들에 대하여 유동시가총액이 높은 순으로 지수 포트폴리오에 편입하여, 최종적으로 유동시가총액 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. 본 지수는 가격지수(Price Return) 입니다.

구분	기준시점	기준지수	발표일	산출시간	리밸런싱
일반전략지수	2017.12.31	1,000pt	2023. 12. 31.	8:30AM (KST)	분기별

2. 지수 구성 방법

A. 기초 투자 유니버스 선정

KEDI 미국 AI 소프트웨어 지수의 유니버스를 구성하기 위하여, 다음과 같은 요건에 부합하는 기초 투자 유니버스를 선정합니다:

- i. NYSE 및 NASDAQ에 상장된 보통주 또는 ADR
- ii. 정기변경일 기준 시가총액 100억 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:
 - Software Publishers (소프트웨어 제작)
 - Computing Infrastructure Providers, Data Processing, Web Hosting, and Related Services (컴퓨팅 인프라 제공, 데이터 처리, 웹 호스팅 등 기타 서비스)

** KEDI Akros Industry Classification Standard (KAICS)는 아크로스테크놀로지스에서 개발한 Akros Industry Classification Standard (AICS)를 기반으로, KEDI 지수위원회의 자문을 거쳐 조정된 분류 체계입니다

B. 키워드 선정

미국 스탠포드 대학의 연구 기관 *Institute for Human-Centered Artificial Intelligence*에서 매년 출간하는 ["Artificial Intelligence Index Report 2023"](#) 자료를 활용하여 AI 소프트웨어 산업을 반영하는 키워드를 선정하였습니다. KEDI 지수위원회의 자문을 통하여 해당 산업과 관련성이 낮은 키워드는 제외할 수 있으며, 키워드는 산업의 성장 및 변화에 따라 혹은 국가기관 및 연구소에서 인정하는 출처 자료의 변경에 따라 수정될 수 있습니다.

키워드	"Artificial Intelligence Index Report 2023" 자료 기준
AI 소프트웨어 (Artificial Intelligence Software)	<p>pg68: "Some of the most starred GitHub Repositories include libraries like TensorFlow, OpenCV, Keras, and PyTorch, which are widely used by software developers in the AI coding community"</p> <p>pg108: "Machine translation studies how well AI software can translate languages"</p> <p>pg218: "We continue to see opportunities across [the software and analytics] segment as payers, providers, and partners take advantage of our high ROI solutions and realize the benefits of our data, AI models, and workflow capabilities."</p>
AI 서비스 (Artificial Intelligence Service)	<p>pg 227: "Another important class of robots are service robots, which the ISO defines as a robot "that performs useful tasks for humans or equipment excluding industrial automation applications"</p> <p>pg 228: "A higher number of professional service robots installed in the world for several key application areas, including hospitality, medical robotics, professional cleaning, and transportation and logistics"</p>

C. 데이터 수집

모든 데이터는 정기변경일을 기준으로 국가기관 자료, 사업보고서, 검색 및 뉴스 자료를 수집합니다. 특히 사업보고서의 경우 비종결정일을 기준으로 NYSE 및 NASDAQ에 상장된 보통주에 대해서는 미국 증권거래위원회에 공시되어 있는 제일 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 각각의 키워드에 대해 얼마나 높은 유사도를 가지는지 LLM을 통하여 다음과 같이 측정합니다 (절대적 기준의 유사도 점수):

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

L: LLM 기반 절대적 유사도 점수 개별 기업과 개별 키워드 간의 LLM 기반 유사도 점수
 K_j: 키워드 개별 키워드
 P_k: 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**):

$$M(K_i, P_j, P_k) = \begin{cases} 1.0, & \text{기업}j\text{와 기업}k\text{를 비교했을 때 기업}j\text{의 유사도가 기업}k\text{에 비해 "높음"} \\ 0.0, & \text{기업}j\text{와 기업}k\text{를 비교했을 때 기업}j\text{의 유사도가 기업}k\text{에 비해 "낮음"} \end{cases}$$

M: LLM 기반 상대적 유사도 점수 기업 대 기업 LLM 기반 키워드 유사도 점수
 K_i: 키워드 개별 키워드
 P_j: 기업 정보 기업에 대한 자료, 보고서, 검색 및 뉴스 정보
 P_k: 기업 정보 기업에 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.

개별 키워드의 상대적 기준의 유사도 점수의 합산 점수를 최종 유사도 점수로 간주합니다. 이때 최종 유사도 점수는 0.0에서 1.0 사이로 조정합니다.

E. 개별 종목의 최종 편입 대상 선정 기준

앞서 산출된 최종 유사도 점수와 유동시가총액 기준으로 최종 편입 대상을 선정합니다.

키워드	최종 유사도 점수	편입 대상 종목 수	최종 편입 종목 수
AI 소프트웨어 (Artificial Intelligence Software)	개별 키워드에 대한 상대적 기준의 유사도 점수의 합산 점수	30	15
AI 서비스 (Artificial Intelligence Service)			

편입가능 종목 선정

편입 대상 종목은 기초 투자 유니버스에 해당하는 기업들에 대해서, 각 키워드별 산출된 최종 유사도 점수가 높은 순서로 최종 편입 종목 수의 2배수만큼 편입가능 종목 수를 산정합니다.

최종 편입 종목 선정

최종 편입 종목은 편입 대상 종목들 중에서 정기 변경 시점 기준 유동시가총액이 높은 순으로 선정합니다. 총 편입 종목 수는 15종목입니다.

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

F. 개별 종목의 편입 비중 산정 방식

본 지수는 다음과 같은 기준으로 편입 비중을 결정합니다:

키워드	총 종목 수	총 포트폴리오 비중	편입 비중 산정 방식
AI 소프트웨어 (Artificial Intelligence Software)	15	100%	유동시가총액과 최종 유사도 점수의 곱을 기준으로 할당 (최대비중: 20%, 최소비중: 3%)
AI 소프트웨어 (Artificial Intelligence Software)			

단, 정기 변경 시 특정 종목의 비중이 표에 기재된 최대 비중을 초과할 경우, 해당 종목의 비중을 최대 비중으로 제한하며, 해당 종목의 초과분을 차순위 종목들에 동일하게 배분합니다. 특정 종목의 비중이 최소 비중 미만일 경우, 해당 종목의 비중을 최소 비중으로 설정하며, 해당 종목의 증가분만큼 차순위 종목들의 비중을 동일하게 축소합니다.

3. 지수 산출 방법

A. 지수값의 산출

본 지수는 다음 수식을 이용하여 유동시가총액을 가중한 지수입니다. 지수 산출 기준일은 2017년 12월 31일이며, 기준지수는 1,000으로 설정하였습니다.

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

$$I(t) = \frac{MV_t}{MV_{t-1}} \times I(t - 1) = \frac{MV_{t-1} + \Delta_{t,t-1}MV}{MV_{t-1}} \times I(t - 1)$$

$$I(t) = \frac{\sum_i P_{i,t} \times Q_{i,t} \times FF_{i,t}}{\sum_i P_{i,t-1} \times Q_{i,t-1} \times FF_{i,t-1}} \times I(t - 1) = \frac{\sum_i P_{i,t-1} \times Q_{i,t-1} \times FF_{i,t-1} + \Delta_{t,t-1}(P_i \times Q_i \times FF_i)}{\sum_i P_{i,t-1} \times Q_{i,t-1} \times FF_{i,t-1}} \times I(t - 1)$$

$$\therefore I(t) = \left(1 + \frac{\Delta MV}{MV_{t-1}}\right) \times I(t - 1) = \left(1 + \frac{\Delta_{t,t-1}(P_i \times Q_i \times FF_i)}{\sum_i P_{i,t-1} \times Q_{i,t-1} \times FF_{i,t-1}}\right) \times I(t - 1)$$

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 _{i,t} :	t 시점 종목 i의 종가
Q _{i,t} :	t 시점 종목 i의 보유량
FF _{i,t} :	t 시점 종목 i의 유동주식비율 (Free Float Ratio)
MV _{i,t} :	t 시점 보유 종목들에 대한 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 KEDI 지수 위원회의 검토를 거쳐 편출 및 차순위 종목을 편입할 수 있습니다.

<책임의 한계와 법적고지>

이 문서를 포함한 KEDI(한국경제신문지수)의 지수 콘텐츠는 (주)한국경제신문과 서면동의 없이 복제, 전송 또는 배포할 수 없습니다. KEDI의 지수 콘텐츠는 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.