

김정민

국가데이터연구원
데이터과학연구팀

'보는 통계'에서 '읽는 통계'로의 전환

- 나우캐스트 지표 생성형 AI 서비스 제공 방안 연구 -

- (연구 목적) 본 연구는 나우캐스트 지표의 가독성 향상을 위해 생성형 AI 기반의 통계 나레이션 모델을 시범 구현하고 실무 활용성을 평가하였다.
- (핵심 결과) LangGraph 기반의 자가 품질 점검 프로세스를 통계 나레이션 모델에 추가할 시, 신뢰할 수 있는 수준의 나레이션 도출이 가능함을 확인하였다.
- (향후 활용) 본 연구에서 제안한 LangGraph 기반 통계 나레이션 모델 구조는 나우캐스트 지표 외 다양한 통계서비스에 적용 시 참고하기에 용이하다.

연구배경 및 목적

- (연구 배경) 국가데이터처는 민간 자료를 활용한 속보성(주간, 월간) 지표인 나우캐스트 지표를 생산해 국가통계포털을 통해 서비스 중
- (문제 인식) 그간 자료 정제, 행정자료와의 연계 등 제공 지표의 정합성 확보에 집중해 온 관계로, 통계서비스 측면의 개선 검토는 상대적으로 부족
- (연구 필요성) 나우캐스트 시스템 고도화 및 이용자 편의성 향상을 위하여, 생성형 AI 기반의 현행 서비스 개선 방안을 모색해 볼 필요
- (연구 목표) 생성형 AI 기반의 통계 나레이션 모델을 시범 개발 후 나우캐스트 지표에 적용해 봄으로써 생성형 AI의 통계서비스 활용 가치 진단

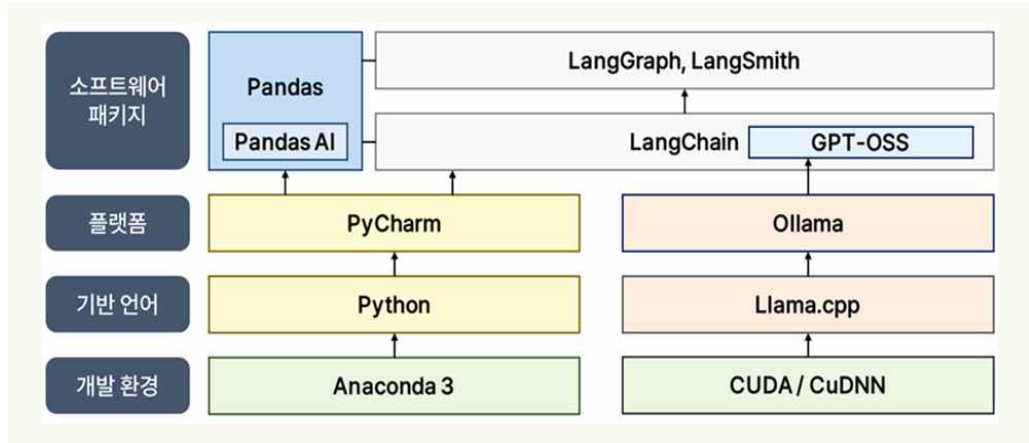
데이터 및 방법

- (활용 데이터) 나우캐스트 지표 9종 중, 신용카드 이용금액, 온라인채용 모집인원수, 모바일 인구이동 지표의 원천 자료 전수를 주요 실험 대상으로 선정
 - (구현 방법론) 통계 나레이션 시나리오를 6단계*로 구분하고, 단계별로 독립적인 시스템 프롬프트를 할당 후 답변 결과를 연계하는 방식을 채택
- * 지표 개요 → 최근 변화 추이 → 최대/최솟값 → 기초 통계량 → 종합 결론 → 유의사항

데이터 및
방법
(계속)

- (활용 패키지) LangChain, LangGraph, Pandas 등 다수

[그림1] 통계 나레이션 AI 모델 구현을 위한 주요 기술 스택



- (AI 모델 선정) 8종의 오픈소스 AI 모델을 정확성, 일관성, 지시 이행의 3가지 기준으로 정성 평가함으로써 GPT-OSS 20B*를 최적합 모델로 판단

* OpenAI가 '25년 8월 공개한 오픈소스 AI 모델로서, 파라미터 규모에 따라 20B와 120B 모델이 존재하며, MOE 기술을 채택해 적은 컴퓨팅 자원으로도 활용 가능

<표1> 오픈소스 생성형 AI 모델 8종의 성능 검토 결과

No.	개발사	모델명	규모(B)	성능 평가		
				정확성	일관성	지시이행
1	네이버	HyperClova X SEED	1.5	○	○	○
				○	×	×
				○	×	×
2	알리바바	Qwen3	32	○	△	×
				○	△	×
				○	△	×
3	알리바바	QwQ	32	○	○	×
				○	○	×
				○	○	×
4	LG AI 연구원	EXAONE 3.5	32	△	△	△
				△	△	△
				△	△	△
5	구글	Gemma3	27	△	△	○
				△	△	○
				△	△	○
6	메타	Llama 4 Scout	109	○	△	△
				○	△	△
				○	△	△
7	환광퀀트	DeepSeek-R1	32	△	○	×
				△	○	×
				△	○	×
8	OpenAI	GPT-OSS	20	○	△	○
				○	△	○
				○	△	○

주요 결과

- (1차 구현) LangChain 기반 1차 구현 결과, 대체로 의도한 바와 일치하는 답변 생성을 확인하였으나, 기초 통계량 분석 과정에서 간헐적 오류 관측*

* 오류의 원인은 환각현상(hallucination)으로, 생성형 AI 모델의 고유 특성

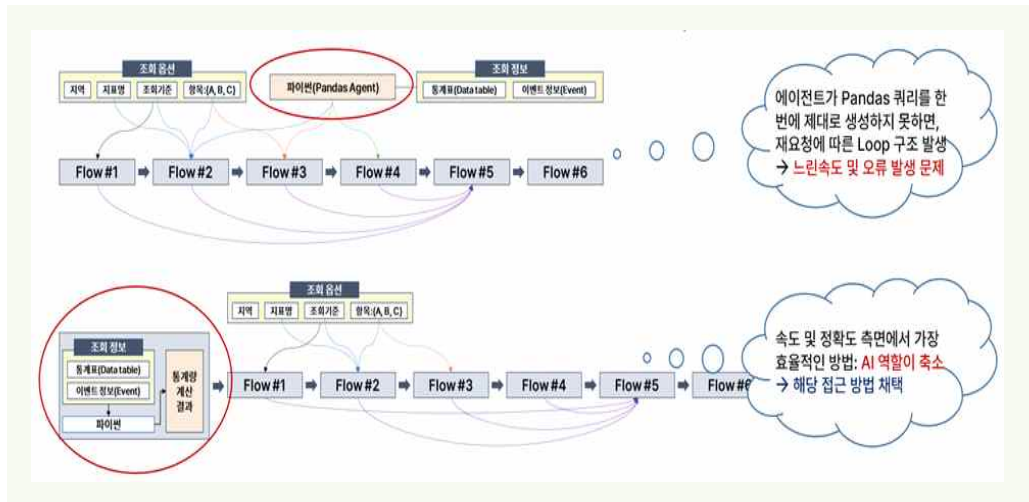
<표2> AI 나레이션 생성 결과의 통계 분석 오류 건수

(독립 시행 1,000회 기준)

모델명	정상 출력 건수	비정상 출력 건수	오류율(%)
GPT-OSS 20B	977	23	2.3
Gemma3 27B	951	49	3.9

- (오류 회피 전략) AI의 통계량 분석 오류를 회피하기 위하여 AI 에이전트를 활용하는 방법과 AI 동작 과정에서 일반 함수를 활용하는 방식을 검토

[그림2] AI 모델의 통계량 분석 오류 회피를 위한 방법 검토 결과



- (오류 회피 실증) 검토한 방법*을 실증해 본 결과, 일반 함수를 활용하는 방식이 추론 속도 및 오류 발생 건수가 가장 낮아 활용에 적합한 것으로 확인

<표3> 모델의 통계량 산출 방식별 모델 평균 실행 시간

(독립 시행 100회 기준)

모델 구분	AI가 직접 분석	AI 에이전트 활용	일반 함수로 산출
전체 실행	2분 50초	4분 47초	1분 21초
오류 발생 건수	5회	17회	2회

- (한계 인식) 1차 구현 방식은 지표의 종류에 따라 통계 나레이션 모델을 추가 개발해야 하는 비효율*과 답변 품질 고도화를 위한 대응책의 한계**가 뚜렷

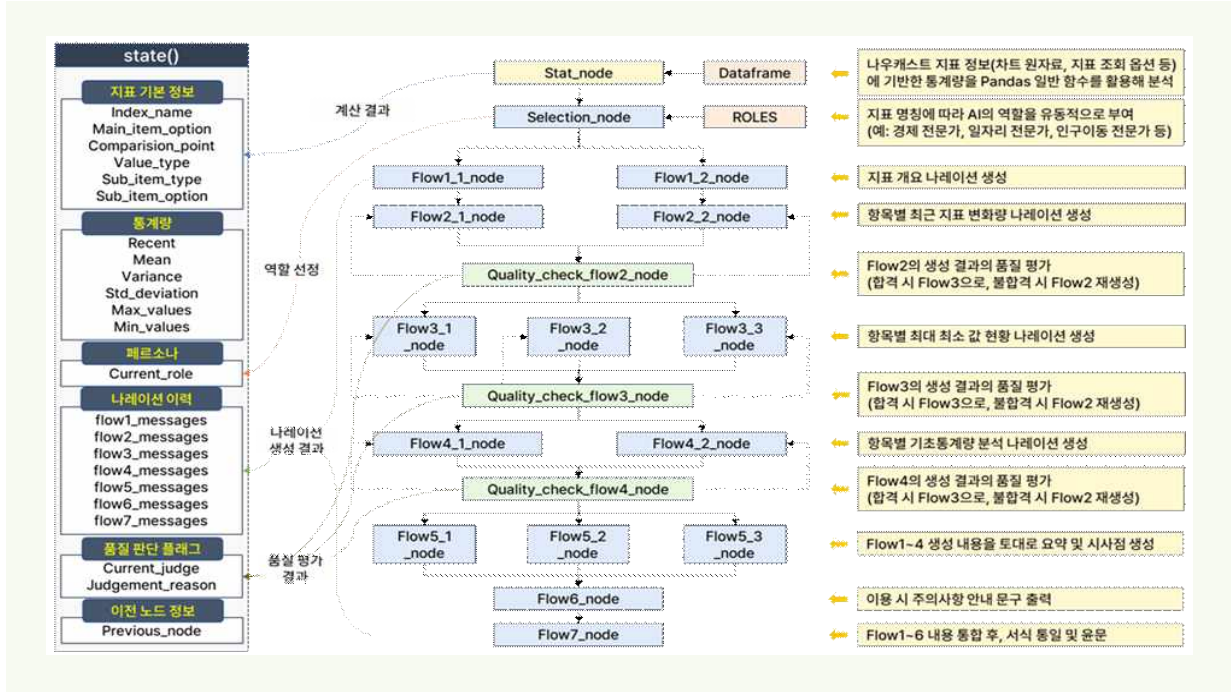
* LangChain은 다수의 AI 답변을 하나로 통합할 수 있는 이점이 있으나, 사전 정의된 일방향 동작만을 허용하므로 지표가 변경되면 모델도 새로 정의해야 하는 문제

** AI의 통계량 분석 오류를 해결하는 방식으로, AI가 오류가 없는 답변을 생성하도록 유도하는 방안 외 후처리(오류 발생 시 수정) 고려가 불가능

주요 결과
(계속)

- (2차 구현: 개선 모델) LangGraph 패키지를 도입하여, 조건부 분기와 순환 구조를 갖춘 그래프 형태의 통계 나레이션 모델로 재구현

[그림3] LangGraph 기반 통계 나레이션 모델 구조 설계 결과



- (2차 모델의 개선점) 나우캐스트 지표 종수를 확장하는데 있어 진입장벽이 낮고, 답변 품질에 대한 자가 점검 기능을 지원

- ① 스테이트(state) 관리를 통해 모든 변수의 체계적 추적 및 제어가 편리
- ② 페르소나 자동 선정 기능으로 지표 유형별 맞춤 나레이션 지원
- ③ 품질 평가 노드를 통해 답변의 자가 검증 및 재생성 순환 구조 구현
- ④ 윤문을 위한 AI 동작을 추가해 최종 나레이션의 서식 일관성 확보

- (자가 품질 평가*의 효과성) 통계량 오류 발생 시 답변 재생성을 지시한 결과, 최종 답변의 오류율을 0%에 수렴하도록 제어**할 수 있음을 확인

* LLM-as-a-Judge로 명명하는 기법으로서, 최근 다양한 논문에서 효과가 입증

** 단, 자가 품질 평가 노드를 모델에 추가하는 경우 품질 평가를 위한 AI 추론이 필요하므로, 최종 답변이 도출되는 데까지 걸리는 시간이 증가

시사점

- (서비스 활용성) LangGraph 기반의 품질 자가 점검 프로세스가 모델에 포함되어 있음을 전제로, 서비스에 적용할 수 있을 것으로 기대
- (실시간 서비스) 현재 기술 수준에서 실시간 나레이션 생성 서비스는 품질 보장을 위한 추론 속도 지연으로 인해 추천되지 않음

시사점 (계속)

- (AI 에이전트) 에이전트 기술은 나레이션 모델에 부적합하나, 향후 SPSS, SAS 등 통계 분석 패키지가 에이전트 기능을 지원할 시, 활용 여지 존재
- (연구 성과 활용) 본 연구의 모델 설계 구조는 나우캐스트 지표 외에도 다양한 통계서비스의 해설문 자동 생성에 확장 적용될 수 있음

한계 및 향후 연구

- (이벤트 정보) 당초 본 연구는 나우캐스트 지표상 제시된 이벤트 정보를 통계 나레이션에 연계하고자 하였으나, 등록된 이벤트 정보 규모* 문제로 한계
 - * 향후 언론 기사 빅데이터를 활용한 이벤트 정보 확충을 검토해 볼 필요
- (통계 벤치마크*) AI 모델의 데이터 테이블 해석 역량을 정량적으로 평가할 수 있는 벤치마크가 부재하므로, 관련 벤치마크 개발 검토 필요
 - * AI 모델의 성능을 진단하기 위해 개발된 도구로서, 측정하고자 하는 성능과 관련된 비공개 문항으로 구성
- (고도화 방안) 지표별 차별화된 시스템 프롬프트 정교화가 검토될 시, 연구를 위해 시범 구현된 결과 대비 높은 생성 품질을 기대할 수 있음
- (정부망 내 AI 도구 활용) 정부망의 보안 정책으로 인해 허깅페이스, VS-code 등 오픈소스 플랫폼의 접근이 제한되는 문제가 있어, 해소 방안 검토 필요

참고 문헌

- 김정민(2024), 국가통계 분야 인공지능 활용 체계 연구, 국가데이터연구원
- 김정민(2025), 통계 나레이션을 위한 생성형AI 개발: Nowcast 지표를 대상으로, 2025 국가통계연구 심포지엄
- Statistics Canada (2025), Project Examples of AI for Analytical Product Preparation and Validation`
- Sui, Y. et al. (2024), Table meets LLM: Can large language models understand structured table data?, WSDM 2024
- Zhang, X. et al. (2025), TableLLM: Enabling tabular data manipulation by LLMs in real office usage scenarios, ACL 2025
- Zhao, Y. et al. (2023), Investigating table-to-text generation capabilities of LLMs in real-world information seeking scenarios, EMNLP 2023

⇒ (연구보고서 원본) 국가데이터연구원홈페이지 > 연구간행물>연구보고서 > 나우캐스트 지표 생성형 AI 서비스 제공 방안 연구

출처를 밝히지 않고 국가데이터연구원 데이터 리서치 브리프의 내용을 무단전재 또는 복제하는 것을 금합니다.
본 리서치 브리프의 내용은 필자들의 개인적 의견이며, 국가데이터연구원의 공식적인 의견이 아님을 밝힙니다.

발 행 일 2026년 4월 30일

발 행 처 국가데이터처 국가데이터연구원

발 행 인 김 진

주 소 35220 대전광역시 서구 한밭대로 713 통계센터 6~8층

전화번호 042-366-7118

홈페이지 mods.go.kr/dsri