www.ksii.or.kr

2026년
추계학술 발표대회

10월 22일(목) – 10월 24일(토)

세인트존스호텔 (강원도 강릉시)

주최 · 주관

사단법인 한국인터넷정보학회

후 원

㈜나루씨큐리티, ㈜피씨엔, ㈜아이티센엔텍, 한화시스템㈜,
㈜LIG시스템, 케이투웹테크㈜, 강원관광재단, 강릉관광개발공사



www.ksii.or.kr
KSII
사단법인 KOREAN SOCIETY FOR INTERNET INFORMATION 한국인터넷정보학회

논문발표

10월 23일(금) 10:40~12:10 (Room 5 / 바부다 4)

세션 I	좌장: 김영철 교수(홍익대)

AI-Metaverse

I-1 가상 융합 쇼핑몰에서의 테스트 에이전트 기반 자동 테스트 메커니즘
정현욱, 김장환, 김영철(홍익대)

I-2 가상 융합 전시관 내 시각·공간 문맥 활용 기반 고도화 RAG 지향 AI 도슨트 구현 사례
김인용, 김장환, 서채연(홍익대), 전병국(강원대), 김영철(홍익대)

I-3 사용자 의도와 생성형 AI 결과의 일치도 평가를 위한 개념 정의 및 평가방법
정연주, 서채연, 김영철(홍익대)

I-4 사용자 선호도 메트릭스 기반 Text-to-Video 생성 결과 평가 방안
정연주, 서채연, 김영철(홍익대)

I-5 LLM기반 AI챗봇 자가 치유 메커니즘
김성회, 서채연, 김영철(홍익대)

10월 23일(금) 10:40~12:10 (Room 6 / 카리브 2)

[워크숍] XR 콘텐츠 자율변형 프레임워크 기술 세션 3	자체 진행

10월 23일(금) 10:40~12:10 (Room 7 / 카리브 1)

육사 AI R&D센터 특별섹션 세미나 1	좌장: 이준 교수(육사)

사용자 의도와 생성형 AI 결과의 일치도 평가를 위한 개념 정의 및 평가방법

정연주[1], 서채연[2], 김영철[3*]
[1]홍익대학교 스마트도시과학경영대학원
[2,3*]홍익대학교 소프트웨어융합학과
e-mail: hello.yeonjujung@gmail.com, {chaeyun, bob@hongik.ac.kr

Concept Definition and Evaluation Method for Alignment between User Intent and Generative AI Outputs

Yeonju Jung[1], Chaeyun Seo[2], R. Young Chul Kim[3*]

1. 서 론

생성형 AI(Generative AI) 결과를 사용자 관점에서 평가하려면 결과 자체의 품질뿐 아니라 사용자가 원하는 내용과 조건이 결과에 반영되었는지를 확인해야 한다. 그러나 사용자의 의도는 최초 프롬프트에 모두 드러나지 않으며, 생성형 AI와 상호작용하는 과정에서 구체화되거나 변경될 수 있다[1][2].

IntentFlow는 사용자 목표(Goal)와 의도(Intent)를 구분하고, 의도가 상호작용 과정에서 나타나거나 변경될 수 있다고 설명한다[1]. Adaptive Prompt Elicitation은 최초 입력 이후 질문과 사용자 선택을 통해 잠재적인 의도를 구체화한다[2]. 따라서 생성 결과를 평가하기 전에 무엇을 사용자 의도로 볼 것인지 확인할 필요가 있다.

생성 결과 평가에서는 복합 지시를 개별 요구로 분해하는 방법이 제안되고 있다. InfoBench[3], Beyond Facts[4], CoPrompter[5]는 복합 지시나 요구를 세부 단위로 나누어 생성 결과의 충족 여부, 누락 또는 잘못된 해석을 평가한다.

본 연구는 사용자 목표와 사용자 의도를 구분하고, 사용자가 확인한 의도를 생성 결과에서 독립적으로 판정할 수 있는 요구 단위로 구성한 뒤 각 요구 단위와 생성 결과를 비교하는 평가방법을 제안한다. 연구질문은 **"생성형 AI 제품에서 평가 기준이 되는 사용자 의도를 어떻게 정의하고, 이를 어떤 요구 단위로 구성하여 생성 결과와 비교할 수 있는가?"**이다. 2장에서는 관련 연구를 검토하고, 3장에서는 평가방법을 제안하며, 4장에서는 결론과 향후 연구를 기술한다.

2. 관련 연구

관련 연구는 사용자 의도를 확인·구체화하는 연구와 복합 요구를 평가 가능한 단위로 분해하는 연구로 구분할 수 있다. IntentFlow는 사용자 입력을 High-level Goal과 Low-level Intent로 구분하고, 의도가 상호작용 과정에서 나타나거나 변경될 수 있음을 설명한다[1]. Adaptive Prompt Elicitation은 최초 입력 이후 시각적 질문과 사용자 선택을 반복하여 Feature Requirement를 구체화하고, 사용자가 확인된 요구를 직접 수정할 수 있도록 한다[2].

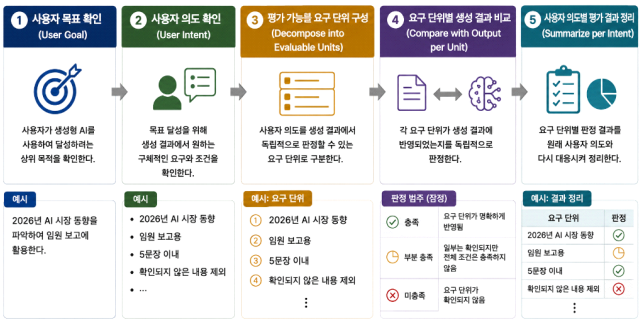
InfoBench는 복합 Instruction을 각각 평가 가능한 Criteria로 분해하고 Decomposed Requirements Following Ratio(DRFR)를 제안하였다[3]. Beyond Facts는 복합 Query를 Intent Constraint로 구분하여 생성 결과의 누락과 잘못된 해석을 평가한다[4]. CoPrompter는 사용자 요구사항을 Atomic Instruction으로 분해하고 Criteria Question으로 변환하여 생성 결과와 비교하며, 사용자가 평가기준을 수정·삭제·추가할 수 있도록 한다[5].

본 연구는 사용자 목표와 사용자 의도를 먼저 구분하고, 사용자가 평가 기준으로 확인한 의도를 요구 단위로 구성한 뒤 동일한 생성 결과와 각각 비교하는 순서로 평가절차를 구성한다.

IntentFlow와 Adaptive Prompt Elicitation은 사용자의 의도를 확인하고 구체화하는 과정에 초점을 두지만, 확인된 의도를 생성 결과의 평가 단위로 사용하는 방법까지 다루지는 않는다[1][2]. 반면 InfoBench, Beyond Facts, CoPrompter는 프롬프트나 질의에 포함된 요구를 세부 단위로 분해하여 생성 결과를 평가한다[3][4][5]. 본 연구는 두 연구 흐름을 구분하여 참고하고, 최초 프롬프트가 아니라 사용자가 확인한 의도를 평가 기준으로 삼는다는 점에서 차이가 있다.

3. 사용자 의도와 생성형 AI 결과의 일치도 평가방법

본 연구에서는 사용자 의도와 생성형 AI 결과의 일치도를 각 요구 단위가 생성 결과에 어느 정도 반영되었는지를 판정하여 확인한다. 이를 위해 평가절차를 다음의 다섯 단계로 구성하였다.



(그림 1) 사용자 의도와 생성형 AI 결과의 일치도 평가 절차

3.1 사용자 목표 확인

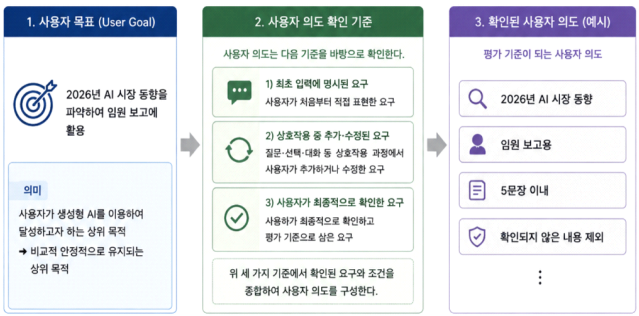
첫 번째 단계에서는 사용자가 생성형 AI를 사용하는 상위 목적을 확인한다. 본 연구에서 '사용자 목표(User Goal)'는 사용자가 생성형 AI 제품을 이용하여 달성하려는 상위 목적을 의미한다. 이 정의는 Goal을 비교적 안정적인 상위 목적, Intent를 목표 달성을 위한 세부 전략·선호·제약조건으로 구분한 IntentFlow의 개념을 참고하였다[1].

예를 들어 사용자가 생성형 AI를 이용하여 2026년 AI 시장 동향을 임원에게 보고하려는 경우, 사용자 목표는 2026년 AI 시장 동향을 파악하여 임원 보고에 활용하는 것으로 정의한다. 이 단계에서는 생성 결과의 세부 조건을 평가하지 않고 사용자가 생성형 AI를 사용하는 목적을 먼저 확인한다.

3.2 사용자 의도 확인

두 번째 단계에서는 사용자 목표를 달성하기 위해 생성 결과에서 원하는 구체적인 요구와 조건을 확인한다. 본 연구에서 **사용자 의도(User Intent)**는 사용자가 목표를 달성하기 위해 생성 결과에서 충족되기를 원하는 요구, 선호 및 조건을 의미한다.

본 연구에서는 최초 입력에 명시된 요구, 상호작용 과정에서 사용자가 추가하거나 수정한 요구, 사용자가 최종적으로 확인한 요구를 기준으로 사용자 의도를 구성한다. IntentFlow에서는 의도가 상호작용 과정에서 나타나거나 변경될 수 있으며[1], Adaptive Prompt Elicitation에서도 최초 입력 이후 사용자 선택을 통해 요구를 점진적으로 구체화한다[2].



(그림 2) 사용자 목표와 사용자 의도 확인 과정 예시

예를 들어 사용자의 목표가 2026년 AI 시장 동향을 파악하여 임원 보고에 활용하는 것이라면, 생성 결과에 대해 2026년 AI 시장 동향, 임원 보고용, 5문장 이내, 확인되지 않은 내용 제외와 같은 구체적인 요구와 조건을 확인할 수 있다. 이 단계에서 확인한 사용자 의도는 이후 생성 결과의 반영 여부를 판정하는 기준으로 사용한다.

3.3 평가 가능한 요구 단위 구성

세 번째 단계에서는 선행연구의 요구 분해 방법을 참고하여, 사용자 의도를 생성 결과에서 각각 독립적으로 판정할 수 있는 '요구 단위'로 구분한다. InfoBench는 복합 Instruction을 각각 평가 가능한 Criteria로 분해하며[3], Beyond Facts는 하나의 Query에 포함된 개념과 행동을 각각 하나의 요구를 나타내는 Intent Constraint로 구성한다[4]. CoPrompter는 복합 요구사항을 Atomic Instruction으로 분해한 뒤 Criteria Question으로 변환한다[5].

예를 들어 2026년 AI 시장 동향을 임원 보고용으로 5문장 이내로 정리하고 확인되지 않은 내용은 제외라는 사용자 의도는 2026년 AI 시장 동향, 임원 보고용, 5문장 이내, 확인되지 않은 내용 제외의 요구 단위로 구분할 수 있다. 요구 단위는 문장을 문법적으로 나눈 결과가 아니라, 하나의 생성 결과에서 각각 반영 여부를 판정할 수 있는 조건으로 구성한다.

3.4 요구 단위별 생성 결과 비교

네 번째 단계에서는 각 요구 단위가 생성 결과에 반영되었는지를 독립적으로 판정한다. 생성 결과 전체에 하나의 평가값을 먼저 부여하지 않고, 동일한 생성 결과를 각각의 요구 단위와 비교한다.

예를 들어 생성 결과가 2026년 AI 시장 동향을 다루고 있고 5문장 이내로 작성되었더라도, 임원에게 필요한 핵심 정보 중심으로 구성되지 않았거나 확인되지 않은 내용이 포함되었다면 각 요구 단위의 평가 결과는 서로 다르게 나타날 수 있다.

본 연구에서는 요구 단위별 비교 결과를 설명하기 위한 잠정적인 판정 범주로 충족, 부분 충족, 미충족을 사용한다. 충족은 요구 단위가 생성 결과에 명확하게 반영된 경우, 부분 충족은 요구한 내용의 일부는 확인되지만 전체 조건을 충족하지 않은 경우, 미충족은 요구 단위가 생성 결과에서 확인되지 않은 경우를 의미한다. 세 범주는 평가절차를 설명하기 위한 잠정적인 구분이며, 세부 판정기준은 향후 연구에서 검증한다.

하나의 사용자 의도가 여러 요구 단위로 구성된 경우 각 요구의 판정 결과는 서로 다를 수 있다. 따라서 일부 요구가 충족되었다는 이유로 전체 사용자 의도가 충족된 것으로 판단하지 않는다. 요구 단위별 결과를 유지하면 생성 결과에서 어떤 요구가 반영되었고 어떤 요구에서 차이가 발생했는지를 구분할 수 있다.

3.5 사용자 의도별 평가 결과 정리

마지막 단계에서는 요구 단위별 판정 결과를 사용자 의도에 따라 정리한다. 이를 통해 사용자 의도에 포함된 요구 가운데 어떤 요구가 충족되었고, 부분적으로 충족되었거나 충족되지 않았는지를 확인한다. 본 연구에서는 요구 단위의 중요도나 가중치를 적용한 종합점수는 산정하지 않는다.

4. 결론 및 향후 연구

본 연구는 생성형 AI 결과를 사용자 관점에서 평가하기 위해 사용자 목표와 사용자 의도를 구분하고, 사용자가 확인한 의도를 독립적으로 판정할 수 있는 요구 단위로 구성하여 생성 결과와 비교하는 방법을 제안하였다. 이 방법은 최초 프롬프트를 사용자 의도로 확정하지 않고, 상호작용 과정에서 확인된 요구를 평가 기준에 포함한다. 또한 생성 결과 전체에 하나의 평가값을 부여하지 않고, 각 요구가 결과에 어떻게 반영되었는지를 개별적으로 확인한다.

본 연구에서는 요구 단위별 결과를 설명하기 위해 충족, 부분 충족, 미충족을 잠정적 판정 범주로 사용하였다. 향후 연구에서는 각 범주의 세부 판정기준과 요구 단위의 중요도 및 필수조건 처리방법을 구체화할 예정이다. 이후 사용자, 제3자 평가자, LLM-as-a-Judge가 동일한 요구 단위와 생성 결과를 평가했을 때 판정 결과가 어느 정도 일치하는지 평가자 간 일치도(Inter-rater Agreement)를 확인하고, 동일한 평가를 반복했을 때 판정 결과가 안정적으로 유지되는지도 검증할 예정이다.

참고문헌

[1] Y. Kim, K. Son, S. Kim, B. Chin, and J. Kim, "IntentFlow: Investigating Fluid Dynamics of Intent Communication in Generative AI," in Proc. of Designing Interactive Systems Conference (DIS '26), 2026.

[2] X. Wen, L. Hegemann, X. Jin, S. Ma, and A. Oulasvirta, "Adaptive Prompt Elicitation for Text-to-Image Generation," in Proc. of 31st International Conference on Intelligent User Interfaces (IUI '26), 2026.

[3] Y. Qin et al., "INFOBENCH: Evaluating Instruction Following Ability in Large Language Models," in Proc. of Findings of the Association for Computational Linguistics: ACL 2024, pp. 13025 - 13048, 2024.

[4] Y. Hao, H. Yu, and J. You, "Beyond Facts: Evaluating Intent Hallucination in Large Language Models," in Proc. of the 63rd Annual Meeting of the Association for Computational Linguistics, pp. 7046 - 7069, 2025.

[5] I. Joshi et al.,"CoPrompter: User-Centric Evaluation of LLM Instruction Alignment for Improved Prompt Engineering," in Proc. of 30th International Conference on Intelligent User Interfaces (IUI'25), 2025.