Review article

인공지능 시대의 영문 편집: 인식적 무결성 확보하기

황윤희, Andrew Dombrowski*

컴팩스

Language editing in the artificial intelligence era: achieving epistemic integrity: an expanded Korean translation

Yunhee Whang, Andrew Dombrowski*

Compecs Inc., Seoul, Korea

***Corresponding email:** yunhee@compecs.com

Received: August 18, 2026 **Accepted:** August 21, 2026 **Published:**

대규모 언어 모델(large language models, LLMs)은 과학 논문의 작성과 편집에 점점 더 널리 활용되고 있다. 그러나 LLM의 높은 언어적 유창성은 연구 주장이 언어적으로 제시되는 방식과 이를 실제로 뒷받침하는 근거 사이의 불일치를 가릴 수 있다. 본 튜토리얼에서는 이러한 문제를 '인식적 무결성(epistemic integrity)'의 관점에서 살펴본다. 인식적 무결성이란 과학적 진술이 이를 뒷받침하는 근거의 질, 강도 및 범위에 적절히 부합하는 상태를 의미한다. LLM을 활용한 수정과정에서 동사, 헤지(hedges), 시제, 평가적 수식어, 범위 표지 등의 인식적 표지(epistemic markers)가 부적절하게 변경되면 주장의 확실성, 범위 및 해석 수준이 근거와 어긋날 수 있다. 이러한 인식적 오조율(epistemic miscalibration)은 개별 문장뿐 아니라 논문 전체에 누적되어 결과와 해석의 경계를 흐리거나 섹션 간 인식적 입장의 표류(stance drift)를 초래할 수 있다. 따라서 인공지능(artificial intelligence)를 활용한 학술 글쓰기에서 편집인은 문법과 문체뿐 아니라, 논문 전체에서 주장의 확실성, 범위 및 해석 수준이 연구 근거에 의해 정당화되는지를 검토함으로써 근거와 주장 간의 정합성(alignment)과 인식적 무결성을 확보해야 한다.

Keywords: 대규모 언어 모델; 언어 편집; 인식적 무결성; 인식적 조율; 인식적 표지

서론

배경

연구 진실성(research integrity)은 흔히 연구의 윤리적 수행과 소통에 관한 문제로 이해된다. 즉, 연구자는 날조, 변조, 표절과 같은 연구부정행위를 피해야 한다. 그러나 이러한 도덕적 의무만으로 연구 진실성의 개념이 모두 설명되는 것은 아니다. 연구 진실성에는 인식적 차원(epistemic dimension)도 존재한다. 과학적 주장은 이를 뒷받침하는 근거의 질, 강도 및 범위에 관하여 독자를 오도해서는 안 된다. De Winter와 Kosolosky [1]는 이러한 차원을 '인식적 무결성(epistemic integrity)'으로 개념화하고, 이를 연구 진실성의 도덕적 측면과 구분하여 과학적 실천을 통해 생성되는 진술의 신뢰성과 비기만성(non-deceptiveness)에 연결하였다. 따라서 인식적 무결성은 과학적 진술이 이를 뒷받침하는 근거가 허용하는 수준을 얼마나 충실하게 반영하는가와 관련된다.

과학적 주장은 언어를 통해 독자에게 전달되며, 이는 주로 문자 언어의 형태를 취하지만 반드시 이에 국한되지는 않는다. 따라서 인식적 무결성은 연구자가 무엇을 주장하는가 뿐만 아니라 그 주장을 어떻게 표현하는가 와도 밀접하게 관련된다. 연구자는 다양한 언어 자원을 활용하여 자신의 입장(stance)을 구성하고, 연구결과의 인식적 지위와 근거가 제공하는 뒷받침 수준에 맞추어 주장의 확실성, 범위 및 해석의 수준을 조율한다[2-4].

경험 많은 편집인에게 언어와 근거 간의 조율(calibration)은 새로운 문제가 아니다. 편집인들은 오래전부터 언어적 선택이 근거가 정당화하는 수준의 확실성, 범위 및 해석을 적절히 반영하는지를 평가해 왔다. 그러나 연구설계, 방법론적 한계 또는 기타 관련 근거를 충분히 고려하지 않은 상태에서도 유창한 텍스트를 생성할 수 있는 대규모 언어 모델(large language models, LLMs)이 널리 활용되면서 이러한 판단의 중요성이 더욱 커졌다[5]. LLM이 생성하거나 수정한 문장은 문법적으로 정확하고 자연스러워 보이더라도,

해당 연구의 근거가 정당화하는 '인식적 수준'을 적절히 반영하지 못할 수 있다. 최근 연구들은 LLM이 자신의 응답 정확성에 관한 불확실성을 언어적으로 얼마나 적절히 표현하는지에도 의문을 제기하고 있다[6-8]. 모델이 언어로 표현하는 확실성은 실제 응답의 정확성이나 모델 내부에서 추정되는 불확실성과 일치하지 않을 수 있으며, 일부 연구에서는 이러한 과잉 확신의 정도가 상당한 수준에 이를 수 있음이 보고되었다[9]. 이러한 연구가 학술논문의 근거–주장 관계를 직접 다루는 것은 아니지만, LLM이 생성하는 확신의 언어를 근거에 대한 신뢰할 만한 평가로 간주해서는 안 된다는 점을 시사한다. 학술 글쓰기와 편집의 맥락에서는 이러한 불일치가 근거가 실제로 정당화하는 수준과 언어적으로 표현된 주장의 확실성, 범위 또는 해석 수준 사이의 간극, 즉 '인식적 오조율(epistemic miscalibration)'로 나타날 수 있다.

목적

본 튜토리얼은 인식적 표지(epistemic markers)로 기능하는 언어적 요소들을 살펴보고, LLM이 이러한 요소들을 부적절하게 선택하거나 수정할 때 문장 및 담화 수준에서 인식적 오조율(epistemic miscalibration)이 어떻게 발생할 수 있는지를 설명한다. 본고의 목적은 첫째, 언어적 선택이 주장의 확실성, 범위 및 해석 수준을 어떻게 조절하는지에 대해 저자와 학술지 편집인의 주의를 환기하고, 둘째, 적절한 인식적 조율을 위해서는 언어와 연구 근거, 그리고 학문 분야별 관습을 종합적으로 고려하는 인간의 해석과 판단이 필요함을 강조하며, 셋째, LLM의 도움으로 다듬어진 언어적 유창성이 근거와 주장 사이의 인식적 오조율을 가릴 수 있음을 보여주는 데 있다.

문장 수준의 인식적 표지

저자와 심사자는 인식적 표지(epistemic markers)를 개별적이고 국소적인 언어 선택으로 인식할 수 있다. 그러나 언어 교정의 관점에서 문장 수준의 수정은 문법 오류를

바로잡거나 문체를 다듬는 데 그치지 않고, 언어적으로 표현된 확실성, 범위 및 해석의 수준이 연구 근거에 적절히 부합하는지를 판단하는 작업까지 포함한다. LLM은 이러한 인식적 차원을 근거에 맞게 조율하지 못하고 확실성이나 불확실성을 부적절하게 표현할 수 있다. 또한 인간 피드백에 기반한 학습과정에서는 사용자가 불확실성을 드러내는 응답보다 단정적이고 확신에 찬 응답을 선호하는 경향이 보고되었으며[10,11], 이러한 선호는 LLM이 생성하는 언어의 확실성에도 영향을 미칠 가능성이 있다. 그 결과 필요한 헤지(hedges)가 약화되거나 삭제되고, 근거가 충분히 뒷받침하지 않는 수준의 단정적인 표현이 생성될 수 있다. 이러한 경향은 사용자의 기대나 관점에 지나치게 부응하려는 LLM의 성향과 결합하여[12], 근거에 맞게 불확실성과 주장의 범위를 조율해야 하는 과학적 글쓰기에서 인식적 무결성을 훼손할 수 있다.

본 튜토리얼에서는 분석과 설명의 편의를 위해 인식적 표지를 동사, 헤지 표현, 시제, 평가적 수식어, 범위 표지의 다섯 가지 유형으로 구분하여 살펴본다. 이러한 언어적 요소에 대한 LLM의 수정은 개별적으로는 미미해 보일 수 있지만, 궁극적으로는 주장의 확실성, 적용범위, 근거에 대한 평가 및 해석 수준을 변화시킬 수 있다.

동사의 선택

동사는 미묘한 의미 차이를 지니며, 저자가 연구결과나 해석을 어느 정도의 확실성으로 제시하는지를 드러내는 중요한 언어적 자원이다. 따라서 동사는 연구설계, 결과의 일관성, 방법론적 한계, 그리고 학문 분야별 관습을 고려하여 선택되어야 한다. 그러나 LLM은 표본크기, 실험연구와 관찰연구의 구분, 통제되지 않은 변수 등 적절한 동사 선택에 필요한 연구 맥락을 충분히 고려하지 못할 수 있다. 또한 사용자의 기대나 관점에 지나치게 부응하는 아첨적 동조(sycophancy) 경향은 결과나 해석을 보다 확신에 찬 언어로 표현하는 방향으로 작용할 가능성이 있다[12]. 이 경우 근거가 정당화하는 수준보다 강한 동사가 선택되어 주장의 확실성이 과도하게 높아질 수 있다. 다음은 이러한 문제를 잘 보여주는 단순화된 예시이다.

(1) The results demonstrate (vs. indicate or suggest) an association between X and Y.

(2) The results suggest (vs. demonstrate or indicate) a causal effect of X on Y.

"Demonstrate," "indicate," 그리고 "suggest"는 일반적으로 서로 다른 정도의 인식적 확신을 전달한다. "Demonstrate"는 뒤따르는 주장이 근거에 의해 비교적 강하게 뒷받침되고 있음을 나타내는 반면, "suggest"는 보다 제한적이거나 잠정적인 해석을 제시하는 데 흔히 사용된다. "Indicate"는 문맥에 따라 그 사이의 의미를 나타낼 수 있다. 동사 선택의 적절성은 동사의 강도만으로 판단할 수 없으며, 동사가 결합하는 명제의 성격과 이를 뒷받침하는 연구 근거를 함께 고려해야 한다. 예를 들어, 충분한 방법론적 근거를 갖춘 대규모 무작위 대조시험(randomized controlled trial)에서 특정 효과가 일관되게 확인되었다면 "demonstrate"와 같은 비교적 강한 동사의 사용이 정당화될 수 있다. 반면, 제한된 표본에서 탐색적으로 관찰된 결과라면 "observe"나 "suggest"와 같은 표현이 더 적절할 수 있다. 그러나 "suggest"와 같이 비교적 신중한 동사를 사용하더라도, 그 뒤에 근거가 뒷받침하지 못하는 인과적 주장이 이어진다면 전체 표현은 여전히 부적절할 수 있다.

주의해야 할 것은 이러한 인식적 함의는 목적어와 연구 맥락에 따라서 달라지기도 한다. 예를 들어, 연구자가 제한된 표본에서 얻은 결과를 신중하게 제시하기 위해 "We observed a trend toward improved performance"라고 썼다고 가정했을 때, 이를 LLM이 "The analysis confirmed a trend toward improved performance"로 수정하면 문장이 단순히 더 단정적으로 들리는 데 그치지 않고, 관찰된 결과에 부여되는 인식적 확실성이 높아질 수 있다. 따라서 편집인은 연구설계, 방법론적 한계, 연구범위와 결과의 일관성 등을 고려하여 해당 주장을 근거가 어느 정도까지 뒷받침하는지 판단하고, 이러한 판단이 동사 선택에 적절히 반영되어 있는지를 검토해야 한다. 특히 LLM을 활용한 수정과정에서는 동사의 인식적 강도가 근거 없이 높아지거나 낮아지지 않았는지 확인할

필요가 있다. 언어적으로 매끄러운 문장일수록 이러한 근거–주장 간의 불일치가 쉽게 드러나지 않을 수 있기 때문이다.

헤지 표현

일반적인 헤지 표현(hedges)에는 조동사(may, might, could), 부사(possibly, potentially, perhaps), 동사(suggest, appear, seem) 등이 포함된다. 헤지는 단순히 신중한 태도를 나타내는 언어적 장치가 아니라, 근거가 허용하는 수준에 맞추어 주장의 확실성을 조절하고 불확실성의 정도를 드러내는 기능을 한다[2]. 헤지를 지나치게 사용하면 주장이 근거에 비해 불필요하게 약화될 수 있고, 반대로 필요한 헤지를 사용하지 않으면 주장이 지나치게 단정적으로 제시될 수 있다. 다음은 이러한 문제를 보여주는 예시이다.

(1) This treatment improves insulin sensitivity.

(2) This treatment may improve insulin sensitivity.

(3) This treatment might potentially improve insulin sensitivity.

예를 들어, 표본이 작고 결과의 불확실성이 상당한 경우라면 문장 (2)가 근거에 보다 적절하게 조율된 표현일 수 있다. 그러나 LLM은 헤지를 삭제하여 문장 (1)과 같이 지나치게 단정적인 표현으로 바꾸거나, 반대로 문장 (3)처럼 의미가 중복되는 헤지를 덧붙여 주장을 필요 이상으로 약화시킬 수 있다.

문제는 헤지가 때때로 근거에 대한 인식적 판단의 결과라기보다 신중하거나 정중한 문체를 만드는 장치로 취급될 수 있다는 점이다. 예를 들어, “The results may prove that the method is effective”에서 “may”는 불확실성을 나타내지만, “prove”는 뒤따르는 주장에 매우 높은 수준의 확실성을 부여한다. 이처럼 서로 다른 수준의 인식적 확신을 나타내는 표현이 한 문장 안에 결합되면, 저자가 해당 주장에 어느 정도의 확신을 부여하는지가 불명확해질 수 있다. 과학적 글쓰기에서 헤지는 신중하고 객관적인 문체를 형성하는 데

기여할 수 있지만, 핵심은 주장을 약하게 표현하는 데 있는 것이 아니라 근거가 허용하는 수준의 불확실성을 정확하게 표현하는 데 있다. 따라서 인식적 조율(epistemic calibration)은 헤지의 유무나 빈도 자체가 아니라, 해당 주장에 부여된 불확실성의 정도가 연구 근거에 의해 정당화되는가에 달려 있다. LLM을 활용한 수정에서는 이러한 관계가 충분히 고려되지 않은 채 헤지가 추가되거나 삭제될 수 있으므로, 편집인은 헤지가 문체적 신중함을 위한 장식이 아니라 근거와 주장 사이의 관계를 정확하게 조율하고 있는지를 검토해야 한다.

시제 선택

학술 글쓰기에서 현재시제는 일반적으로 보편적 사실이나 현재 받아들여지는 지식을 제시하는 데 사용되며, 과거시제는 방법이나 결과와 같이 완료된 연구를 보고하는 데 사용된다. 현재완료는 과거의 연구나 사건을 현재의 논의와 연결할 때 주로 사용된다. 그러나 시제 선택은 단순히 사건의 시간적 위치를 나타내는 데 그치지 않고, 연구결과를 어느 범위까지 일반화하여 제시할 것인지와 관련된다는 점에서 인식적 기능도 수행한다. 예를 들어, “A phenomenon was observed”는 특정 연구와 조건에서 관찰된 결과를 보고하는 반면, “A phenomenon is observed”는 해당 현상을 보다 일반적인 경험적 패턴으로 제시하고 있다. 이처럼 특정 맥락에서 과거시제는 결과를 해당 연구의 관찰범위에 한정하는 효과를 갖는 반면, 현재시제는 그 결과를 보다 일반화된 지식이나 해석으로 제시하여 주장의 적용범위를 넓힐 수 있다. 따라서 과거시제에서 현재시제로의 변화는 단순한 문법적 수정이 아니라 연구결과가 어느 범위까지 적용되는지에 대한 해석을 변화시킬 수 있다.

시제 사용에 관한 관습은 학문 분야와 논문의 수사적 맥락에 따라 달라질 수 있으므로, 편집인은 해당 분야의 관습과 문장의 기능을 함께 고려하여 판단해야 한다. 문제는 LLM이 이러한 맥락을 충분히 고려하지 않은 채 문법적 또는 문체적 일관성을 위해 시제를 변경할 때 발생할 수 있다[13]. 과거시제를 현재시제로 바꾸면 연구 특이적인

결과가 지나치게 일반화될 수 있고, 반대로 현재시제를 과거시제로 바꾸면 저자가 의도한 일반화된 해석이나 현재적 타당성이 약화될 수 있다. 요컨대 시제의 불일치는 반드시 바로잡아야 할 문법적 오류가 아니다. 표면적으로 불일치해 보이는 시제의 변화가 실제로는 주장의 범위나 저자의 인식적 입장의 차이를 나타낼 수 있기 때문이다. 따라서 영문 교정자는 시제를 기계적으로 통일하기보다, 각 시제가 수행하는 수사적·인식적 기능을 파악하고 연구 근거가 허용하는 범위 안에서 그 선택의 타당성을 검토해야 한다.

평가적 수식어

연구자는 "clear/clearly, substantial/substantially, strong/strongly, compelling"과 같은 평가적 형용사와 부사를 사용하여 연구결과나 근거의 특정 측면을 강조한다. 이러한 수식어는 주장의 확실성, 효과의 크기, 근거의 강도 등 서로 다른 측면을 선택적으로 부각할 수 있다. 예를 들어 "The treatment clearly reduced postoperative pain"에서 "clearly"는 결과에 대한 높은 확신을 나타내고, "The drug substantially improved glycemic control"에서 "substantially"는 효과의 크기를 강조한다. 또한 "The findings provide compelling evidence that the intervention improves glycemic control"에서 "compelling"은 제시된 근거를 강력한 것으로 평가한다. 이처럼 평가적 수식어는 동일한 연구결과에서도 확실성, 효과의 크기, 근거의 강도 등 서로 다른 측면을 선택적으로 부각한다.

편집하는 과정에서 LLM이 연구 근거와 맥락을 충분히 고려하지 않은 채 이러한 수식어를 삽입하거나 강화하게 되면 문제가 발생할 수 있다. 예를 들어, LLM이 "The treatment reduced postoperative pain"을 단지 더 명확하거나 설득력 있게 들리도록 하기 위해 "The treatment clearly reduced postoperative pain"으로 수정할 수 있다. 그러나 연구 근거가 "clearly"가 나타내는 높은 수준의 확신을 정당화하지 않는다면, 이러한 수정은 단순한 문체 개선이 아니라 결과에 부여되는 확실성을 부적절하게 높이는 변화가 된다. 평가적 수식어는 단순히 주장을 더 강하게 만드는 것이 아니라, 연구결과와 근거의 서로 다른

측면에 대한 저자의 평가를 언어적으로 드러내는 것이다. 따라서 영문 교정자는 평가적 수식어를 단순한 문체적 강조 장치로 보지 않고, 각 수식어가 결과나 근거의 어떤 측면을 부각하며 그러한 평가가 연구 근거에 의해 정당화되는지를 검토해야 한다.

국소적 범위 표지

국소적 범위 표지(local scope markers)는 문장 수준에서 주장이 적용되는 범위를 제한한다. 예로는 "in this study, under the tested conditions, in high-risk patients, in this cohort" 등이 있다. 가령 "The method was effective under the tested conditions," "라는 표현은 연구결과가 특정 맥락에 한정된다는 점을 명시한다. LLM은 적절한 범위 표지를 삭제하거나 필요한 범위 표지를 반영하지 않음으로써 근거가 허용하는 수준을 넘어 주장의 적용범위를 부주의하게 확대하고, 의도하지 않은 과잉 일반화를 초래할 수 있다. 영문 교정자는 필요한 경우, 범위 표지를 복원하거나 추가하여 주장이 근거가 허용하는 경계 안에 머물도록 해야 한다. 범위 표지는 모집단, 연구환경, 중재, 측정조건, 시간범위를 명확히 하며, 이를 삭제하면 특정 조건에 한정된 결과가 보다 일반적인 주장으로 미묘하게 확대될 수 있다.

본 섹션에서는 인간 편집인의 판단이 필요한 인식적 표지를 다섯 가지 유형으로 구분하여 살펴보았다. 그러나 이러한 구분은 분석을 위한 것이며, 실제 학술 글쓰기에서 각 표지가 독립적으로 기능하지 않고 서로 결합하여 주장의 인식적 성격을 형성한다. 따라서 실제 편집과정에서는 개별 표지를 분리하여 평가하기보다 이들의 상호작용과 복합적 효과를 함께 고려해야 한다. 특히 LLM을 활용한 논문에서는 동사, 헤지, 시제, 평가적 수식어 등 여러 언어적 장치가 동시에 변경되면서 저자가 의도한 인식적 입장이 과장되거나 축소될 수 있다. 따라서 영문 교정자는 개별 표현의 적절성뿐 아니라 이들이 결합하여 형성하는 전체적인 인식적 입장(epistemic stance)이 근거의 강도와 범위에 부합하는지를 판단하고, 연구 주장의 정확성과 인식적 무결성이 유지되도록 해야 한다.

담화 수준의 인식적 표지

LLM이 생성한 텍스트를 문장 수준에서 검토하다 보면 동사의 미세한 강화, 겉보기에는 사소한 헤지 삭제, 미묘한 시제 변화 등이 발견될 수 있다. 영문 교정자는 연구 근거에 비해 주장을 지나치게 강화하거나 약화시키지 않는지 검토하고, 필요한 경우 근거에 맞게 다시 조정하곤 한다 각각의 변화는 개별적으로는 미미해 보일 수 있지만, 논문 전체에 걸쳐 누적되면 저자의 인식적 입장(epistemic stance)에 상당한 변화를 가져올 수 있다. 따라서 담화 수준에서는 개별 문장의 적절성뿐만 아니라, 저자가 연구결과에 대해 어느 정도의 확신과 신중성을 나타내며 이러한 인식적 입장이 논문 전체에서 어떻게 형성되고 변화하는지를 함께 살펴보아야 한다. 본 섹션에서는 개별 인식적 표지를 넘어 논문 전체에서 형성되는 인식적 입장에 주목하고, 특히 섹션 간 입장 표류(stance drift)를 식별하고 조정하기 위해 편집인은 어떠한 판단이 필요한지 살펴본다.

결과와 해석의 경계

담화 수준에서 중요한 경계 중 하나는 연구결과의 보고와 그 의미에 대한 해석을 구분하는 것이다. 일반적으로 결과(Results) 섹션에서는 관찰되거나 분석된 결과를 중심으로 보고하고, 고찰(Discussion) 섹션에서는 그 결과를 해석하고 기존 연구와의 관계, 함의 및 의의를 논의한다. 이러한 구분의 정도는 학문 분야와 학술지의 관행에 따라 달라질 수 있다. 그러나 LLM을 활용한 수정은 이러한 수사적 인식적 경계를 흐릴 수 있다. 예를 들어, 결과 섹션에 다음과 같은 문장이 있다고 가정 했을 때, "The model achieved higher accuracy than the baseline, implying its potential clinical applicability," 이 문장은 문법적으로는 문제가 없지만, "implying its potential clinical applicability"는 관찰된 결과를 보고하는 데 그치지 않고 그 결과로부터 임상적 함의를 도출하는 해석을 덧붙인다. 즉

단순한 경험적 결과의 보고에서 그 결과의 의미에 대한 해석으로 담화 기능이 이동하게 된 것이다. LLM은 문장 간 연결을 매끄럽게 하거나 결과의 의미를 보다 명확하게 제시하는 과정에서 이러한 해석적 표현을 추가할 수 있다. 그러나 그 결과, 관찰된 사실과 그 사실로부터 도출된 해석 사이의 경계가 불분명해지고, 해석이 마치 연구결과 자체의 일부인 것처럼 제시될 수 있다. 따라서 영문 교정자는 해당 표현이 그 섹션의 수사적 기능에 적절한지를 판단하고, 필요한 경우 이를 고찰 섹션으로 옮기거나 결과 보고에 적합한 형태로 조정해야 한다.

섹션 간 인식적 확실성

논증이 전개됨에 따라 주장의 확실성이나 해석의 수준이 달라질 수 있지만, 이러한 변화는 자의적 이어서는 안 되며 연구 근거와 논증의 전개에 의해 정당화되어야 한다. 따라서 영문 교정자는 개별 문장의 적절성뿐 아니라 논문 전체에서 인식적 어조(epistemic tone)가 일관되게 유지되는지를 검토해야 한다. LLM으로 인해 발생할 수 있는 입장 표류(stance drift)의 한 형태는 논문의 후반부로 갈수록 별다른 근거 없이 주장이 강화되는 것이다.

결과 섹션의 "A positive association was observed between X and Y"와 고찰 섹션의 "These findings demonstrate that X plays a critical role in Y"를 비교해 볼 때, 각 문장을 개별적으로 보면 뚜렷한 언어적 문제는 없어 보인다. 그러나 두 문장을 논문 전체의 흐름에서 비교하면, 결과 섹션에서 보고된 연관성이 고찰 섹션에서는 X의 역할에 대한 보다 강한 해석으로 전환되고 있음을 알 수 있다. 이는 단순한 표현상의 변화가 아니라 주장의 성격과 그 주장에 부여되는 확실성이 함께 변화한 사례이다. "was observed"는 해당 연구에서 X와 Y 사이의 연관성이 관찰되었음을 보고하는 반면, "demonstrate"는 뒤따르는 해석이 근거에 의해 상당한 수준으로 확립되었음을 시사한다. 또한 "X plays a critical role in Y"는 단순한 연관성을 넘어 X의 인과적 또는 기능적 역할에 대한 해석을 제시한다.

고찰 섹션에서는 결과를 해석하고 그 의미를 논의하기 때문에 결과 섹션보다 더 강한

언어를 사용하는 것이 자연스럽다고 여겨질 수 있다. 그러나 새로운 근거나 방법론적 정당화가 없다면, 고찰 섹션이라는 이유만으로 관찰된 연관성을 더 강한 해석으로 전환하거나 이에 대한 확실성을 높여서는 안 된다. 편집인은 이러한 섹션 간 변화를 추적하여 주장의 성격과 이에 부여되는 확실성의 변화가 연구 근거와 논증의 전개에 의해 정당화되는지를 확인하고, 그렇지 않은 경우 적절한 수준으로 다시 조율해야 한다.

담화 수준의 과잉 일반화

과잉 일반화(overgeneralization)는 한 문장 안에서 발생할 수도 있지만, 논문이 전개되는 과정에서 범위를 제한하는 표현이 점차 약화되거나 삭제되면서 누적적으로 나타나기도 한다. 다음은 과잉 일반화의 예시이다.

(1) "In this cohort, the intervention improved outcomes."

(2) "The intervention improved outcomes."

(3) "The intervention was effective."

(1)에서 (2)로 이동하면서 연구대상 모집단을 한정하는 "in this cohort"가 삭제되고, (2)에서 (3)으로 이동하면서 구체적으로 관찰된 결과가 중재의 전반적인 효과성에 대한 보다 포괄적인 평가로 전환된다. 각 단계의 변화는 개별적으로는 미미해 보일 수 있지만, 이러한 변화가 논문 전반에 걸쳐 누적되면 주장의 적용범위가 근거가 직접 뒷받침하는 범위를 넘어 점차 확대될 수 있다.

LLM은 간결성이나 문체적 일관성을 높이는 과정에서 범위 표지를 삭제하거나 구체적인 결과를 보다 일반적인 표현으로 바꿀 수 있다. 따라서 영문 교정자는 개별 문장뿐 아니라 논문 전체에서 주장의 범위가 어떻게 변화하는지를 추적하고, 연구설계와 연구대상 모집단 등을 고려하면서, 그 범위가 연구 근거가 실제로 뒷받침하는 수준을 벗어나지 않는지 검토해야 한다.

논문 전반의 인식적·평가적 어조

문장 수준에서는 헤지나 평가적 수식어와 같은 언어적 장치가 개별 주장의 확실성이나 평가방식을 조절하지만, 담화 수준에서는 이러한 표현의 누적적 사용이 논문 전체의 인식적·평가적 어조를 형성한다. 예를 들어, LLM이 "may"나 "appears to"와 같은 헤지를 반복적으로 삭제하면, 각각의 수정은 사소해 보이더라도 논문 전체는 원래보다 지나치게 단정적인 어조를 띨 수 있다. 마찬가지로 한계나 문제점을 기술하는 과정에서 "serious limitation"이나 "critical issue"와 같은 표현이 반복되면 논문 전체가 필요 이상으로 부정적이거나 방어적인 어조를 띨 수 있다. 반대로 긍정적 평가 표현이 반복되면 연구의 의의가 과도하게 강조되어 과장된 어조가 형성될 수 있다. 따라서 영문 교정자는 개별 표현의 빈도 자체보다는 이러한 표현이 논문의 어느 부분에서 어떠한 기능으로 사용되는지 살펴보고, 그 결과 형성되는 인식적·평가적 어조가 연구 근거와 각 섹션의 수사적 기능에 비추어 적절한지를 판단해야 한다.

편집인을 위한 시사점

본 튜토리얼은 LLM이 생성하거나 수정한 텍스트의 언어적 유창성이 인식적 무결성(epistemic integrity)과 반드시 일치하지 않으며, 이로 인해 문장 및 담화 수준에서 인식적 오조율(epistemic miscalibration)이 발생할 수 있음을 보여주었다. 인식적 무결성을 유지하려면 연구 주장이 근거가 뒷받침하는 확실성, 범위 및 해석의 수준을 넘어서는 안 되며, 이러한 인식적 특성은 동사, 헤지, 시제, 평가적 수식어, 범위 표지와 같은 다양한 언어적 장치를 통해 부분적으로 표현된다. LLM을 활용한 수정과정에서는 이러한 언어적 장치가 근거와의 관계를 충분히 고려하지 않은 채 변경되면서 주장이 과도하게 강화되거나 약화되기도 한다. 또한 서로 다른 수준의 확실성이나 범위가 획일화될 수도 있다. 따라서 인공지능(artificial intelligence) 시대의 영문 교정자는 문법과 문체를 점검하는

데 그치지 않고, 이러한 국소적 인식적 표지가 연구 근거에 부합하도록 적절히 조율되어 있는지를 검토해야 한다.

담화 수준에서도 인간의 판단은 필수적이다. 논문의 인식적 입장은 모든 섹션에서 동일하게 유지되어야 하는 것이 아니라, 연구 근거와 각 섹션의 수사적 기능에 따라 적절하게 변화하고 전개되어야 한다. 그러나 LLM이 생성한 텍스트의 높은 유창성은 섹션 간 주장 범위나 확실성, 해석 수준의 미묘한 변화를 놓치게 만들기도 하기 때문에 입장 표류(stance drift)를 발견하기 어려운 경우도 있다. 따라서 편집인은 먼저 논문의 핵심 주장을 식별하고, 이들이 결과, 고찰, 결론을 거치면서 어떻게 변화하는지를 추적해야 한다. 특히 이러한 변화가 연구 근거에 의해 정당화되는지를 검토함으로써, 논문 전체에서 근거와 주장 사이의 적절한 조율이 유지되도록 해야 한다.

LLM이 학술 글쓰기와 편집과정에 점점 더 깊이 통합됨에 따라 인간의 판단이 불필요해지는 것이 아니라, 오히려 편집 및 검토과정에서 요구되는 판단의 범위가 확대될 수 있다. 문법, 문체, 명료성, 가독성에 대한 기존의 점검에 더해, 언어적 선택이 연구 근거와 주장의 관계를 왜곡하지 않는지를 검토하는 인식적 감독(epistemic oversight)이 요구되기 때문이다. 따라서 인간 편집인의 전문적 판단은 여전히 필수적이며, 특히 미세한 언어적 변화가 연구 결과에 부여되는 확실성, 적용범위 및 해석 수준을 실질적으로 변화시킬 수 있기 때문에, 학술 출판 맥락에서 편집인의 역할은 더욱 중요하다.

ORCID

Yunhee Whang: https://orcid.org/0000-0003-4519-9474

Andrew Dombrowski: https://orcid.org/0000-0002-8532-6653

Authors' contribution

Conceptualization: YW. Methodology/formal analysis/validation: YW, AD. Project administration: YW. Writing–original draft: YW. Writing–review & editing: YW, AD.

Conflict of interest

Andrew Dombrowski serves as the English editor for this journal, but ws not involved in the peer review or decision making process of this article. No other potential conflict of interest relevant to this article was reported.

Funding

None.

Data availability

Not applicable.

Acknowledgments

None.

Supplementary materials

None.

References

1. De Winter J, Kosolosky L. The epistemic integrity of scientific research. Sci Eng Ethics 2013;19:757-774. https://doi.org/10.1007/s11948-012-9394-3
2. Hyland K. The author in the text: hedging scientific writing. Hong Kong Pap Linguist Lang Teach 1995;18:33-42.
3. Hyland K. Boosting, hedging and the negotiation of academic knowledge. Text Talk 1998;18:349-382. https://doi.org/10.1515/text.1.1998.18.3.349
4. Hyland K. Stance and engagement: a model of interaction in academic discourse. Discourse Stud 2005;7:173-192. https://doi.org/10.1177/1461445605050365
5. Wijesinghe C. Epistemic integrity in human-AI interaction: philosophical and religious inquiry in a probabilistic world. SSRN [Preprint] 2025 Jul 29. https://doi.org/10.2139/ssrn.5370815
6. Yona G, Aharoni R, Geva M. Can large language models faithfully express their intrinsic uncertainty in words? arXiv [Preprint] 2024 Sep 26. https://doi.org/10.48550/arXiv.2405.16908
7. Ghafouri B, Mohammadzadeh S, Zhou J, Nair P, Tian JJ, Tsujimura H, Goel M, Krishna S, Rabbany R, Godbout JF, Pelrine K. Epistemic integrity in large language models. arXiv [Preprint]. 2025 Jun 8. https://doi.org/10.48550/arXiv.2411.06528
8. Zhou K, Hwang JD, Ren X, Sap M. Relying on the unreliable: the impact of language models' reluctance to express uncertainty. arXiv [Preprint] 2024 Jul 9. https://doi.org/10.48550/arXiv.2401.06730
9. Sun F, Li N, Wang K, Goette L. Large language models are overconfident and amplify human bias. arXiv [Preprint]. 2025 May 4. https://doi.org/10.48550/arXiv.2505.02151
10. Hosking T, Blunsom P, Bartolo M. Human feedback is not gold standard. arXiv [Preprint] 2024 Jan 16. https://doi.org/10.48550/arXiv.2309.16349
11. Finks C. From approval to truth: how RLHF creates systematic hallucination in large language models: a critical analysis of preference optimization and its epistemic consequences: technical white paper version 1.0 [Internet]. Constraint Layer AI Research; 2025 [cited 2026 Aug 10].

Available from: https://constraintlayer.ai/papers/rlhf-hallucination-analysis-2025.pdf

12. Malmqvist L. Sycophancy in large language models: causes and mitigations. arXiv [Preprint] 2024 Nov 22. https://doi.org/10.48550/arXiv.2411.15287

13. Baron R. AI editing: are we there yet? Sci Ed 2024;47:78-82. https://doi.org/10.36591/SE-4703-18