

생성형 AI의 한국어 용언의 활용 능력 연구

송현주 (Hyunju Song), 경북대학교

남길임 (Kilim Nam), 연세대학교

본 연구는 국내외 생성형 AI가 한국어의 문법에 대해 모국어 화자와 유사한 수준의 문법적 능력을 갖고 있는지 살피기 위한 연구, 한국어문법 개론서 수준에서 용언의 활용과 관련하여 생성형 AI가 어느 정도 한국어를 이해하고 생성할 수 있는지를 살펴보고자 한다. 이를 위하여 남기삼·고영근 외(2019)의 『새로 쓴 표준국어문법론』의 '4장 용언과 어미' 부분, 국립국어원(2005)의 『외국인을 위한 한국어 문법 1』의 15장 단어의 갈래 가운데 '동사, 형용사, 이다' 부분, 고영근·구본관(2018)의 『개정판 우리말 문법론』의 6장 '용언과 그 쓰임' 부분의 용례를 대상으로 평가 데이터 세트를 구성하기로 한다. 이들 개론서를 대상으로 하여 평가 데이터 세트를 구성하는 것은 개별 학술 논문에 비해 주관성이 다소 덜하고, 해당 문장이 어떤 문법 항목의 문법성 판단에 관한 것인지 분명하기 때문이다. 생성형 AI의 과제 수행에 필요한 프롬프트는 인간과의 의사소통과 최대한 유사하게 작성하되, 비문을 교정할 때 평가 대상 문법 항목의 관점에서 접근할 수 있도록 평가 의도를 반영하였다.

한국어 용언의 활용에 대한 문법성 판단 테스트에서 '축약, 품사 선택 제약, 의미 제약, 시제 제약, 서법 제약'을 포함하고 있는 경우 오류를 보이는 경향이 있다. 국외의 생성형 AI와 비교했을 때, 국내산인 'CLOVA X'가 문법적 이해 능력에서는 상대적으로 높은 정확률을 보이는데 이는 '어미'가 형태와 주로 관련되며, 'CLOVA X'가 더 많은 한국어 데이터를 학습하고 있기 때문인 것으로 보인다. 그럼에도, '불규칙 용언의 활용'과 '서법 제약'에 대해서는 모든 생성형 AI가 높은 오류율을 보이는데, 용언의 품사와 어미의 서법 제약에 관한 정교한 평가 데이터 세트의 개발과 실험, 이를 통한 오류 유형의 확보가 필요한 것으로 보인다. 용언에 대한 문법적 생산 능력 평가는 용언의 이해 능력 평가에 비해 상당히 낮은 정확도를 보이며, 용언의 활용 시 시제, 서법, 의미 제약을 고려하여 올바른 문장을 생성해 내기 위해서는 현재 개발된 생성형 AI는 국내외산 모두 큰 개선이 필요할 것으로 보인다. 이러한 성능 개선을 위해서는 저자원 언어인 한국어 데이터에 대한 입력과 함께 오류 유형별로 학습 가능한 데이터 세트의 제공이 필요할 것이다.