

개요

대화형 인공지능 챗지피티(ChatGPT)의 등장은 컴퓨터 프로그래밍 언어에 익숙하지 않은 사람도 손쉽게 인공지능을 사용할 수 있도록 함으로써 인공지능 활용의 장벽을 낮추었으며, 언어 교육 분야의 가능성을 확장하였다. 이에 본 연구에서는 자동 작문 채점 도구로서 챗지피티-3의 신뢰도와 정확도를 살펴보았다. 또한 다양한 언어적 특성이 챗지피티-3를 활용한 자동 작문 채점에 어떠한 영향을 미치는지에 대해서도 알아보았다.

연구 방법

본 연구에서는 영어가 모국어가 아닌 수험자들이 작성한 12,100개의 토플(TOEFL) 작문을 챗지피티-3을 활용하여 자동으로 채점하였다. 챗지피티-3에 채점을 요구하는 명령어(프롬프트; prompt)에는 아이엘츠(IELTS) 과제 2의 채점 기준(과제 응답의 적절성, 응집성과 응결성, 어휘적 다양성, 문법적 다양성과 정확성 등 4가지 평가 영역, 각 10점 만점)도 함께 제공되었다. 채점의 일관성을 확인하기 위해 1,210개의 작문을 챗지피티-3을 통해 재채점했으며, 이때 작문은 다양한 글쓰기 수준(상, 중, 하)과 모국어 화자(11개 언어)를 포함하도록 선정되었다. 또한 선행연구에 기반하여 어휘, 통사, 응집성과 관련한 45개의 언어적 특징으로 작문을 분석한 후 이들이 챗지피티-3의 채점 결과에 미치는 영향을 알아보았다.

연구 결과

연구 결과 챗지피티-3을 활용한 자동 작문 채점은 인간이 작문을 채점하는 것에 도움을 줄 수 있을 정도의 정확도와 신뢰도를 보이는 것으로 나타났다. 인간의 채점 결과와 챗지피티-3의 채점 결과는 1-2점 정도의 차이가 있었지만 글쓰기 수준(상, 중, 하) 측면에서는 차이가 유의미하지 않았다. 더 나아가 언어적 특성을 활용하는 것은 챗지피티-3의 채점의 정확도를 유의미하게 높이는 것으로 나타났다. 특히 기존 연구에서 선별된 45개의 언어적 특징과 챗지피티-3의 채점 결과를 결합하였을 때 가장 높은 정확도를 보였다.

함의

결론적으로 챗지피티를 비롯한 인공지능 언어 모델은 인간의 채점을 대체할 수준의 정확도를 보이는 것은 아니지만 자동 작문 채점 도구로서 가능성을 지니고 있다고 볼 수 있다. 인간에 의한 작문 채점은 시간과 노동력이 크게 소모되고 주관적 견해 및 피로도 등의 외적 요소가 개입될 수 있다는 점을 고려할 때, 챗지피티를 자동 작문 채점 도구로서 활용하는 것은 채점 시간을 줄이고 정확도를 높이는 보조 도구로서 중요한 역할을 할 수 있을 것이다.

Original Text: Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050.