Authorship Verification for Academic Dishonesty in the Era of AI

Jang, Jun (School: Oxford High School)

With the rise of Generative AI (Gen AI), academic dishonesty in the classroom has skyrocketed, with over 90% of students now admitting to using it for schoolwork. While Gen AI offers various learning benefits, empirical studies show that an overreliance on these tools erodes creativity and critical thinking, which are crucial for the future of innovation. Unfortunately, the existing solutions often fall short. AI detectors merely analyze one text at a time, failing to account for students' previous writings. Meanwhile, many authorship verification (AV) models fail to analyze the nuances in writing styles that truly distinguish authorship. To fill this existing gap, this project proposes a novel authorship verification system designed as a feature vector model. The model incorporates a comprehensive feature set, which is a combination of standard transformer-based token-level features (e.g., POS tag patterns) and handcrafted stylistic features (e.g., vocabulary richness & sentence structure variation). Extracted from a variety of datasets and real-world high school student essays, these combined vectors were used to train and evaluate multiple machine learning (ML) binary classification models for detecting academic dishonesty. The proposed comprehensive AV feature vector model filled the critical gaps in the literature, outperforming the standard token-based approaches by over 25%, while also enhancing transparency through interpretable decisions. This research highlights the importance of writing style features within a holistic, feature-based AV system. It assists educators across the globe in detecting academic dishonesty more reliably and responsibly, preserving critical thinking and innovation in the age of AI.

Awards Won: