Robots Cannot Read NAPLAN Essays

Leading US academic, Les Perelman's recent analysis of a 2015 paper by Australian Curriculum, Assessment and Reporting Authority (ACARA) has revealed there are potential flaws in using automated essay scoring (AES) to mark NAPLAN Online writing tasks, indicating that AES is virtually worthless as an educational tool.

According to Perelman's analysis, the most glaring flaw is that AES has not been shown to measure the content of written work, but rather qualities like sentence structure and vocabulary. Consequently, AES platforms may award high marks to random gibberish, as long as it uses sophisticated vocabulary and sentence structure.

A dramatic demonstration of this is the BABEL Generator: developed at the Massachusetts Institute of Technology (MIT), this program can generate nonsense essays which have repeatedly been shown to score high marks from AES platforms.

Not surprisingly, a recent independent survey commissioned by P&C Federation has revealed that only 7% respondents believe that a computer system can mirror human marking.

P&C Federation President, Susie Boyd commented "If a student just takes time to memorise some well-written gibberish, could that improve their NAPLAN scores under AES? If so, that would completely undermine the value of NAPLAN results, and it's disturbing that ACARA hasn't even looked into that possibility. Students may not be able to game the system like this during the initial trial period, as ACARA has said automated marking will be double checked by humans. However, how will ACARA ensure this doesn't happen if human markers are removed completely?"

Even research from Pearson Education, a leading developer of AES platforms, explicitly states that AES systems "*are not good at assessing rhetorical voice, the logic of an argument, the extent to which particular concepts are accurately described, or whether specific ideas presented in the essay are well founded.*" The developers also conceded that "*Assessment of creativity, poetry, irony, or other more artistic uses of writing is beyond such systems.*"

Ms Boyd commented "If you look at the NAPLAN marking criteria for writing, it includes qualities like persuasive devices and the ability to engage the reader. Given Pearson Education's own conclusions, we shouldn't trust the ability of AES systems to implement NAPLAN's marking criteria."

Ms. Boyd went on to say, "There are numerous holes that have been identified in ACARA's research, including citing studies as showing AES is better at marking than humans when the studies really indicate the opposite. We call on ACARA to focus on what's in the best interests of our children and abandon its plans to introduce AES to assess our children's creative work."

Sydney, NSW, 20 October 2017

Authorised By:
President   Susie Boyd

Media Contact:
Communications Officer   Hetaal Badiyani   1300 885 982   media@pandc.org.au