

Evaluation of the difficulty of a geometric statement: comparing ChatGPT and GeoGebra Discovery

Piedad Tolmos
Juan Carlos University, Spain

30th Applications of Computer Algebra - ACA 2025

Our communication will present some initial results from an experience that we are developing, comparing the “complexity” measure assigned by GeoGebra Discovery’s ShowProof command to a variety of well-known, elementary geometric statements, and the performance (i.e. correctness, clarity, and level of detail in the answer) of ChatGPT when asked about the same statements. Let us recall that the ShowProof command algorithmically outputs a proof by contradiction of a given geometric statement, expressing 1 as a combination of the hypotheses and the negation of the thesis. And ranks the “interest” or “difficulty” of the statement by computing the highest degree of the polynomials required to describe such contradiction. Measuring the interest of the output of automated reasoning tools is a classical challenge, but we think that the rank computed by the ShowProof command could be the first algorithmic approach towards establishing such measure in the context of geometric statements, although yet requiring a careful experimental work, such as the one we are initiating now, regarding its practical performance..