

Journal Name:	<a href="#">Asian Journal of Research in Computer Science</a>
Manuscript Number:	Ms_AJRCOS_116113
Title of the Manuscript:	Veritas AI: The ChatGPT Polygraph
Type of the Article	Original Research Article

**PART 1: Review Comments**

	Reviewer's comment	Author's comment (if agreed with reviewer, correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)
<p><b>Compulsory REVISION comments</b></p> <ol style="list-style-type: none"> <li>1. Is the manuscript important for scientific community? (Please write few sentences on this manuscript)</li> <li>2. Is the title of the article suitable? (If not please suggest an alternative title)</li> <li>3. Is the abstract of the article comprehensive?</li> <li>4. Are subsections and structure of the manuscript appropriate?</li> <li>5. Do you think the manuscript is scientifically correct?</li> <li>6. Are the references sufficient and recent? If you have suggestion of additional references, please mention in the review form.</li> </ol> <p><u>(Apart from above mentioned 6 points, reviewers are free to provide additional suggestions/comments)</u></p>	<p>1. Yes 2. Yes 3. Yes 4. Yes 5. Can't tell. For example in figure 4 Veritas gets a 9 in User Transparency. Based on what? All I could see was: "A score from 1 to 10, with 10 being the most transparent, gauges the algorithm's explainability to users. Veritas AI scores the highest with 9, suggesting that its decisions are highly interpretable, which is crucial for trust and ethical considerations." But how did you determine that Veritas got a 9? How was it measured? Who measured it? If it is based only on the author's evaluations and claims that of course is meaningless. I assume there must be some objective criteria that were used but what they were was not clear to me. There are many other examples like that such as ease of integration. Again Veritas gets a 9 but I don't see any description of how that value was measured. Figure 6 is not at all convincing. ChatGPT could determine that the Beatles had 4 members and could correctly determine the truth or falsity of every statement in that table except the claim that someone climbed Mount Everest. And I don't see how one can automatically classify such a statement as a lie. If the statement is embedded in a respected magazine such as National Geographic it is probably true. If it is a random person one met in a bar it is probably false. On its own it is less likely to be true because only a very small percentage of people have done that but that again is something ChatGPT or other LLMs could determine. E.g., I asked MS Copilot: "how likely is it that any random person has climbed Mt. Everest" and it replied "very low" So nothing in Table 6 qualifies as "lie detection" it is more fact verification and each answer is well within the means of a standalone LLM. Same for Table 8. In fact the response from MS Pilot to "how likely is it that any random individual has never cheated on a test in their life" was far more nuanced than the blanket assumption that everyone has cheated at some point. The system claims that it is analyzing non-verbal cues based on facial recognition but there is no description of any studies. Who were the subjects? What kind of equipment was used to record their face as they spoke. What were their instructions? Were they told to lie at some point or were they simply asked questions and the system decided if they were likely lying? 6. References are adequate but if authors resubmit they <i>must</i> provide a URL for their Github site or provide a very compelling reason why they don't. The bullet points at the end of section 5 are very suspect. E.g., "Incorporation of specialized biometric analysis tools, such as Tobii Pro for ocular tracking and Noldus FaceReader" Why "such as" did you use those tools or not?</p>	
<p><b>Minor REVISION comments</b></p> <ol style="list-style-type: none"> <li>1. Is language/English quality of the article suitable for scholarly communications?</li> </ol>	<p>English is good. Tone is not appropriate. This reads more like a product brochure than a journal article. Things like "It embodies the zenith of current scientific understanding in these fields, pushing the boundaries of what machines can comprehend about human language." and "The</p>	

	<p>implications of Veritas AI are profound, promising a paradigm shift in security, forensics [4], and communication. Its deployment could redefine trust in digital interactions, offering a beacon of truth in an era increasingly mired by misinformation." That really is very self promotional for a journal article. Tone it down a bit and be more objective. Show don't tell. Give evidence for that conclusion and let readers make such value judgements. Or at a minimum save such self promotional text for the conclusion.</p> <p>Some more examples: "use_transformer_model applies a pre-trained transformer to generate embeddings." What transformer model and how was it trained? Did you build your own from scratch? If so why? Why not use one of the existing transformers? If you did use a transformer which one? Google? Skikit-Learn? ????</p> <p>" by asserting the falsehood of the statement based on perceived non-verbal cues" How were these non-verbal cues picked up? What software and hardware were used? What was the setup? Were subjects informed about the tests or were they just asked random questions. If the latter and if the answers subjects received were of the form: "That is a lie. You have cheated on a test at least once in your life" that is not appropriate scientific respect for human subjects. Also, the idea that no one has ever cheated is IMO not correct. I personally know some very religious and otherwise ethical people who I would hypothesize have never cheated on a test in their life.</p> <p>The psuedo code in section 4 is not very understandable and is suspicious. As far as I can tell there are never any iterations in the code. That seems strange A diagram of control or data flow between the various sub-systems would be much more appropriate. For a good example see Figure 1 Transformer - model architecture in the Google paper: Attention is All You Need by Vaswani et. al.</p>	
<b>Optional/General comments</b>	<p>There is absolutely no detail on the implementation. What programming language was used? Did the authors develop everything from scratch (highly doubt that) or did they build on some existing LLM such as BERT or existing Transformers from Google or Skikit-Learn? Where is the Github repository for the code? If there is no Github repository why not? Failure to make code open source and available for inspection makes the work automatically suspect.</p> <p>Actually, I'm beginning to wonder if any of this has actually been implemented or if this is all just a hypothetical architecture. If nothing has been implemented that should be made very clear from the very beginning because the way it is written it sounds as if they are talking about an actual working system. If it has been implemented they need to be much more detailed on things like programming language(s), reuse of transformers or other models, equipment and software used to analyze non-verbal cues, and actual experiments (if any) conducted with real people.</p>	

**PART 2:**

	<b>Reviewer's comment</b>	<b>Author's comment</b> (if agreed with reviewer, correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)
<b>Are there ethical issues in this manuscript?</b>	<i>(If yes, Kindly please write down the ethical issues here in details)</i>	

**Reviewer Details:**

Name:	<b>Michael De Bellis</b>
Department, University & Country	<b>USA</b>