Qeios

Peer Review

Review of: "SafeSynthDP: Leveraging Large Language Models for Privacy-Preserving Synthetic Data Generation Using Differential Privacy"

Weisan Wu[1]

1. Baicheng Normal University, China, King Fahad University of Petroleum and Minerals, Saudi Arabia

1. **Overall Coverage and Relevance**

   The references cover a broad range of topics relevant to your work, including data security under GDPR, differential privacy, synthetic data generation, large language models (LLMs), and their applications in various domains. This comprehensive coverage is commendable and helps situate your work within the current state of the art.

2. Formatting and Consistency Issues

- **Inconsistent Markers:** The reference list contains various markers such as "^" and prefixes like "a, b" or "a, b, c" before some citations. These markers are not standard in most citation styles and may confuse readers. It is recommended that you remove these markers or, if they are intended to indicate multiple authors or contributions, clarify their meaning in a footnote.
- **Multiple Versions of the Same Work:** For example, Brown et al.'s "Language Models are Few-Shot Learners" appears twice—once as a preprint on arXiv and once as part of the NeurIPS proceedings. If these refer to the same work, please consolidate the citations and choose the most appropriate version for citation.

3. Detail Consistency

- **Publication Information:** Please verify that all publication details (e.g., author names, publication year, page numbers, and DOIs) are correct and consistent across the references. Some entries may require further checking to ensure that they match the original sources.

doi.org/10.32388/4TIDJK

- **DOI and Links:** It is advisable to provide DOIs or stable URLs for all references where available. This will enhance the accessibility of the cited works for readers.

4. Suggestions for Improvement

- **Adopt a Standard Citation Style:** I recommend selecting a standard citation style (such as IEEE, ACM, or APA) and reformatting the entire reference list accordingly. This will ensure consistency and professionalism.
- **Eliminate Redundancy:** Remove duplicate citations and ensure that each work is cited only once unless there is a compelling reason to reference multiple versions.
- **Clarify Specific References:**
  - For "Chain-of-thought prompting elicits reasoning in large language models," ensure that the citation includes complete details such as the conference name, page numbers, and year in a consistent format.
  - For foundational works on differential privacy (e.g., Dwork & Roth, 2014), consider verifying if more recent reviews or updates are available and whether they should be included.
  - For Kingma and Welling's "An introduction to variational autoencoders," confirm the correct publication year and version, as multiple versions exist.

Overall, the algorithm design and motivation in this paper are very strong. However, I noticed a lack of theoretical analysis throughout the manuscript. Specifically, there is no presentation of theorems, lemmas, or properties that detail how the privacy guarantees are obtained, nor is there any discussion on the algorithm's complexity, the convergence of parameters, or error bounds and limits. Additionally, the nature of the aggregated dataset is not specified—for example, whether it is Boolean or of another type.

Furthermore, the algorithm only considers the Laplace and Gaussian noise mechanisms. Beyond the absence of any theoretical justification for choosing these two mechanisms, there is no discussion regarding the advantages or disadvantages of other noise mechanisms (such as the exponential mechanism) or alternative privacy frameworks (like Pufferfish privacy). Definitions of neighboring datasets and related concepts are also not provided in sufficient detail.

In summary, incorporating some theoretical results or guidelines would not only enhance the paper's rigor but also improve its readability by clearly outlining the assumptions, complexity, convergence, and error bounds associated with the proposed method.

Declarations

Potential competing interests: No potential competing interests to declare.