**Google (PaLM 2) - total: 27/48**

1. Data sources
   * 2 points: Dataset pipeline is disclosed: training on a large corpus of web documents, books, code, mathematics, and conversational data (including a significant fraction of non-English and parallel data across hundreds of language pairs with English as one half of the pair) with various forms of data processing (“deduplication, removal of sensitive PII and filtering”). The size of the dataset is not disclosed (though is acknowledged to be larger than PaLM), nor is finer-grained sourcing. as a singular dataset derived from The Pile with all relevant preprocessing, with The Pile itself being documented extensively (including dataset size and fine-grained sourcing) in the associated paper and datasheet. Note: The size of the dataset was potentially leaked following the release.
   * Sources: <https://ai.google/static/documents/palm2techreport.pdf> Sections 3 and E.1
2. Data governance
   * 3 points: The subject of data governance and curation is raised, with specific practices to mitigate harms (e.g. filtration practices) and measurement of biases/representation. With that said, no constraint is imposed to formally preclude the use of inappropriately-governed data, though many dimensions of governance are considered.
   * Sources: <https://ai.google/static/documents/palm2techreport.pdf> Sections 3 and E.1
3. Copyrighted data
   * 0 points: Copyright is not discussed.
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
4. Compute (additive)
   * 0 points: The model size and compute requirements are not disclosed (this is acknowledged in the model card). The hardware used is disclosed to be TPU v4s, but the amount or duration are undisclosed. Note: The FLOPs and associated information are reported in comparison to the Chinchilla paper, but no confirmation is provided that these numbers (as opposed to the broader scaling law) are instantiated in the PaLM v2 models.
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>, especially E.10.
5. Energy (additive)
   * 0 points: The energy usage, emissions, energy sources, and measurement methodology are not disclosed
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
6. Capabilities and limitations
   * 4 points: Capabilities are described (“ PaLM 2, which excels at a variety of capabilities. PaLM 2 has been optimized for ease of use on key developer use cases and the ability to follow instructions with precision and nuance.”), with several more specific use cases (“natural language processing (NLP) applications such as chat bots, text summarization, and question and answer”). Several limitations (e.g. “Large language models are powerful tools, but they are not without their limitations. Their versatility and applicability can sometimes lead to unexpected outputs, such as outputs that are inaccurate, biased, or offensive. Post-processing, and rigorous manual evaluation are essential to limit the risk of harm from such outputs. See the safety guidance for additional safe use suggestions.”), largely aligning with the literature on language model limitations. Some grounding to evaluation/use cases.
   * Sources: <https://ai.google/static/documents/palm2techreport.pdf>, <https://developers.generativeai.google/models/language> and <https://developers.generativeai.google/guide/safety_guidance>
7. Risks and mitigations (additive)
   * 3 points: Risks are listed (e.g. bias, inaccuracy, offensive language), mitigations are provided (e.g. PII filtration, input controls, safety filters, prohibited use case policy misuse of many forms). Risk mitigations, especially around safety filters, are contextualized to particular risk categories fairly clear, though the exact effects and what remains unmitigated are still not fully clear.
   * Source: <https://developers.generativeai.google/guide/safety_guidance>, <https://developers.generativeai.google/guide/safety_setting>, <https://policies.google.com/terms/generative-ai/use-policy>
8. Evaluations (additive)
   * 2 points: Extensive evaluation for accuracy across many tasks/datasets. Evaluation of unintentional harm (toxicity in several forms, memorization, bias via BBQ). No evaluation of malicious use or other desiderata (e.g. robustness, calibration).
   * Source: <https://ai.google/static/documents/palm2techreport.pdf> Section 4 and Appendix E
9. Testing (additive)
   * 2 points: Extensive discussion of various forms of internal testing beyond standard public benchmarks. No discussion of external evaluation (e.g. HELM) or redteaming (e.g. ARC).
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
10. Machine-generated content
    * 3 points: Usage policy places several forms of prohibitions: “Generate and distribute content intended to misinform, misrepresent or mislead, including misrepresentation of the provenance of generated content by claiming content was created by a human, or represent generated content as original works, in order to deceive. Generation of content that impersonates an individual (living or dead) without explicit disclosure, in order to deceive”. Within context of PaLM API, it is clear that the content is machine-generated. No disclosure of how to ensure content remains disclosed as machine-generated, though given “external or production use is strictly prohibited”.
    * Source: <https://policies.google.com/terms/generative-ai/use-policy>, <https://developers.generativeai.google/terms>, <https://developers.generativeai.google/products/palm>
11. Member states
    * 4 points: Only available in the United States, so this point is largely not applicable. Note: Given the model is not put on market in the EU currently, we recognize the AI Act does not have bearing at present, but nonetheless assess compliance as if it were put on the market.
    * Sources: <https://developers.generativeai.google/available_regions>
12. Downstream documentation
    * 4 points: Extensive documentation of the model in the paper, including a model card, and further deployment-centric documentation is available with discussion of requirements/obligations of downstream developers.
    * Source: <https://ai.google/static/documents/palm2techreport.pdf> and <https://developers.generativeai.google/guide/safety_guidance>