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

1. Data sources
   * 2 points: Types of data are described (“web documents, books, code, mathematics, and conversational data” plus significant amounts of non-English language data), as well as size relative to previous version PaLM, though specific sources and size of dataset are not described.
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
     1. Note: CNBC [claimed](https://www.cnbc.com/2023/05/16/googles-palm-2-uses-nearly-five-times-more-text-data-than-predecessor.html) to receive information in a leak indicating PaLM 2 was trained on 5x more data than PaLM.
     2. Note 2: Could have given only 1 point. The technical report’s details on data processing do not count towards “Description of the data sources used in the development of the foundation model” in my view
2. Data governance
   * 3 points: Substantial discussion of “measures to examine the suitability of the data sources and possible biases and appropriate mitigation” in the technical report. These include “removal of sensitive-PII and filtering”, “special control tokens marking the toxicity of text”, an analysis of representation of people and identity-related topics in pretraining data, and measurement of various biases. But: datasets that are not subject to appropriate data governance are not excluded, as required by the AIA.
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
     1. Note: Could have given 2 points. While E.1 is impressive on this, there’s still far to go to get to “Process and incorporate only datasets that are subject to appropriate data governance measures”
3. Copyrighted data
   * 0 points: No description of copyrighted training data.
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
4. Compute
   * 0 points: In the model card, compute requirements are “not reported.” Google is the compute provider as PaLM 2 is trained on TPU v4, but no information on model size, training time, number of hardware units, or FLOPs are provided.
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
     1. Note: Section 2.2 estimates optimal number of parameters for a given number of FLOPs (or fixed compute cost) as compared to Chinchilla.
5. Energy
   * 0 points: no description of energy usage, emissions, measurement strategy, or mitigations.
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
6. Capabilities and limitations
   * 4 points: Detailed discussions and evaluations of capabilities and limitations. From the PaLM API - PaLM 2 “is intended to be used for a wide variety of natural language processing (NLP) applications such as chat bots, text summarization, and question and answer. The embedding service allows additional NLP use cases such as document search. Limitations - 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.” The technical report assesses performance on capabilities such as coding, translation, question answering, and natural language generation and highlights limitations such as toxic language generation, representational bias, and misgendering in translation.
   * Sources: <https://ai.google/static/documents/palm2techreport.pdf> ; <https://developers.generativeai.google/models/language> ; <https://developers.generativeai.google/guide/safety_guidance>
7. Risks and mitigations
   * 3 points: extensive list of risks and mitigations as well assessments of how effective mitigations are in reducing risks. However, there is little discussion of non-mitigated risks or justification for not mitigating such risks.
   * Sources: <https://ai.google/static/documents/palm2techreport.pdf> ; <https://policies.google.com/terms/generative-ai/use-policy> ; <https://developers.generativeai.google/guide/safety_guidance> ; <https://developers.generativeai.google/guide/safety_setting>
8. Evaluations
   * 2 points: significant evaluations of accuracy and unintentional harm in different tasks and datasets, but no evaluations of intentional harms such as malicious use, robustness, or other factors.
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
     1. Note: Could almost give 3 points for how extensive the evaluations are for unintentional harm, but the other dimensions are very important.
9. Testing
   * 2 points: Wide variety of internal testing beyond standard public benchmarks. No description of external evaluations, red teaming, or stress testing.
   * Source: <https://ai.google/static/documents/palm2techreport.pdf>
10. Machine-generated content
    * 3 points: clear policy against misrepresenting machine-generated content: “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” is prohibited by Google’s Generative AI Prohibited Use Policy. No discussion of watermarking or other mechanisms to ensure content is identifiable as machine-generated, though PaLM 2 has not yet been fully rolled out.
    * Sources: <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 currently available in the US.
    * Source: <https://developers.generativeai.google/available_regions>
12. Downstream documentation
    * 4 points: Substantial documentation is made available by Google to downstream providers in both the technical paper and guidance on using the API.
    * Source: <https://developers.generativeai.google/guide/safety_guidance>; <https://developers.generativeai.google/guide/safety_setting> ;

<https://ai.google/static/documents/palm2techreport.pdf>