**Cohere (Command) - total: 24/48**

1. Data sources
   * 3 points: Dataset description denotes the size of the dataset, the filtering (e.g. English-only), and the general sourcing/demographic factors associated with the data. The dataset lacks finer-grained sourcing and more concrete discussion of the stages/forms of data (e.g. we expect that given Cohere Command is optimized to follow instructions, the model is trained on data more specific to instruction-following but this is not discussed in any form and only Internet-type data is mentioned).
   * Source: Data source is disclosed in <https://docs.cohere.com/docs/generation-card>, pointing to the data statement itself at <https://docs.cohere.com/docs/data-statement>
2. Data governance
   * 3 points: The subject of data governance and curation is raised, with specific practices to mitigate harms (e.g. filtration practices) amidst the broader theme of “compliance with best practices, industry guidelines and applicable regulations.” With that said, no constraint is imposed to formally preclude the use of inappropriately-governed data.
   * Source: <https://docs.cohere.com/docs/data-statement>
3. Copyrighted data
   * 0 points: No acknowledgement of copyright in any form.
   * Source: <https://docs.cohere.com/docs/data-statement>
4. Compute (additive)
   * 0 points: The model size, training time, hardware, and FLOPs are not disclosed. The models are named suggestive of a size (e.g. medium, extra large and model sizes were disclosed previously (e.g. 52B as cited in <https://arxiv.org/pdf/2210.14986.pdf>), but we are not aware of such size disclosure for Cohere Command)
   * Source: <https://docs.cohere.com/docs/environmental-impact>, <https://docs.cohere.com/docs/generation-card>, and <https://docs.cohere.com/docs/command-beta>.
5. Energy (additive)
   * 1 point: The carbon emissions for other models (i.e. embedding models) are disclosed, but the emissions or energy usage for Command are not disclosed. We do see the measure methodology used in the past is based on <https://mlco2.github.io/impact/>.
   * Source: <https://docs.cohere.com/docs/environmental-impact>
6. Capabilities and limitations
   * 3 points: Capabilities are described as extensive (“Command is Cohere's generative model that responds well with instruction-like prompts; Command is trained to follow user commands and to be instantly useful in practical business applications.”), with several more specific use cases (e.g. “Command can write product descriptions, help draft emails, suggest example press release copy”). Limitations are taxonomized as belonging to five categories: “languages, socio-economic, historical, ungrounded, biases”, largely aligning with the literature on language model limitations.
   * Source: <https://docs.cohere.com/docs/model-limitations>, <https://docs.cohere.com/docs/the-command-model>, and <https://docs.cohere.com/docs/command-beta>.
7. Risks and mitigations (additive)
   * 2 points: Risks are listed spanning usage guidelines and limitations, with some discussion of mitigation in terms of data filtration and documentation. No discussion of efficacy of mitigations. No discussion of what risks persist unmitigated and why that is justifiable.
   * Sources: <https://docs.cohere.com/docs/usage-guidelines> and <https://docs.cohere.com/docs/model-limitations>
8. Evaluations (additive)
   * 2 points: Accuracy on multiple datasets (HellaSwag, COPA) and some bias/toxicity results (RealToxicityPrompts, StereoSet, BOLD). No results for malicious use (e.g. disinformation) or other desiderata (e.g. robustness, calibration). Note: Results appear outdated given name change for Command models, but we give credit at present.
   * Source: <https://docs.cohere.com/docs/generation-card>
9. Testing (additive)
   * 2 points: Results of external evaluations disclosed (HELM) as well as ongoing third-party audits and penetration testing. No substantive results of additional internal testing are disclosed beyond above evaluations on public benchmarks (though the Cohere safety team has published work 2 years ago in <https://arxiv.org/abs/2108.07790>). Note: The results are not released for the aforementioned audits/penetration testing, but we give credit at present.
   * Source: <https://docs.cohere.com/docs/the-command-model> and <https://cohere.com/security>
10. Machine-generated content
    * 2 points: In the context of the web interface, it is clear that content is machine-generated. No discussion of machine-generated content being disclosed in documentation or further downstream obligations.
    * Source: <https://dashboard.cohere.ai/playground/generate>, <https://cohere.com/saas-agreement> and <https://cohere.com/terms-of-use>
11. Member states
    * 2 points: Organization-level prohibited use in US-embargoed countries in relation to Apple app store, but no further restrictions in relation to EU member states. No discussion of how the models are deployed across different EU member states.
    * Sources: <https://cohere.com/terms-of-use>
12. Downstream documentation
    * 4 points: Documentation of model, data, and related agreements. Downstream use is gated by “ every developer must clearly outline their use case and have it approved by Cohere through our application process. The application requires thoroughly understanding of our models and their limitations, which will be updated as the models improve. Beyond these Usage Guidelines, you should refer to the Generation and Representation model cards for detailed information about each model.”
    * Source: <https://docs.cohere.com/docs/usage-guidelines>, <https://docs.cohere.com/docs/generation-card>, <https://cohere.com/saas-agreement> and <https://cohere.com/terms-of-use>