**Cohere (Command) - total: 22 / 48**

1. Data sources
   * 3 points: Data statement notes size of overall training dataset (200GB), and the datasheet includes some sources and relative sizes (i.e., the dataset includes the Google Books dataset, CommonCrawl, and text from the internet scraped by the Cohere infrastructure team. The top ten domains scraped were: wordpress.com, medium.com, stackexchange.com, tumblr.com, elsevier.com, genius.com, bbc.co.uk, libsyn.com, yahoo.com, nytimes.com). No description of data used for fine tuning; specific relative size of sources and sources other than those explicitly named not included.
   * Sources: Data source is disclosed in the model card <https://docs.cohere.com/docs/generation-card> as coheretext-filtered; Data statement <https://docs.cohere.com/docs/data-statement> provides more detail.
     1. Note: the ecosystem graphs [node](https://crfm.stanford.edu/ecosystem-graphs/index.html?asset=coheretext) for the coheretext-filtered dataset says the datasheet provides more information, but the link to the datasheet just goes to the data statement (which does not provide the information ecosystem graphs includes). I searched for a while, but could not find the datasheet. I assume the datasheet exists and is accurately represented in EG, but we should find it and include it in the EU AI Act info gathering tab of the TI [v.04] sheet. (The EG entry inaccurately says coheretext-filtered is not used to train Generation models.)
2. Data governance
   * 3 point: describes several measures to mitigate biases in datasets, including investing resources in dataset curation, filtering harmful/biased/racist/undesirable documents. Risks/biases raised include overrepresentation of white male perspectives, imprecise filtering resulting in silencing of marginalized voices or removal of counterspeech.
   * Sources: Data statement <https://docs.cohere.com/docs/data-statement>
     1. Note: easily could have given 2 points as it does not adequately touch on other parts of data governance beyond mitigating biases, though this provision in the AI act is especially focused on mitigating biases
3. Copyrighted data
   * 0 points: No summary of which data used to train the model copyrighted or how the company addresses the copyright issues.
   * Sources: Data statement <https://docs.cohere.com/docs/data-statement> ; also checked terms of use <https://cohere.com/terms-of-use>
     1. Note: might change framing to “could not find any summary”
4. Compute (additive)
   * 0 points: No details on model size, number of GPUs, compute provider, or FLOPS.
   * Sources: N/A. Papers on other cohere foundation models discuss model size, compute provider, and FLOPS. Cohere says that all its models are trained on Google’s TPUs, but it also appears to use Nvidia GPUs via Amazon so despite its long term agreement with Google Cloud the specific compute provider for Cohere Command is unclear. <https://arxiv.org/pdf/2210.14986.pdf> <https://arxiv.org/abs/2204.06514> <https://nvidianews.nvidia.com/news/aws-and-nvidia-collaborate-on-next-generation-infrastructure-for-training-large-machine-learning-models-and-building-generative-ai-applications> <https://docs.cohere.com/docs/environmental-impact>
5. Energy (additive)
   * 1 points: Provides an emissions upper bound for Generation models, of which Cohere Command is one. Gives methodology for calculating upper bound (ML CO2 Impact Tool).
   * Sources: <https://docs.cohere.com/docs/environmental-impact>
     1. Note: could have given 2 points for providing an emissions upper bound, but I think there’s enough room for improvement here for this to be a 1.
6. Capabilities and limitations
   * 3 points: Provides a solid list of limitations with specific examples, as well as descriptions of capabilities via several use cases. Lacks technical paper outlining limitations and capabilities.
   * Sources: <https://docs.cohere.com/docs/model-limitations> ; <https://docs.cohere.com/docs/command-beta> ; <https://docs.cohere.com/docs/the-command-model> ; <https://docs.cohere.com/docs/generation-card>
7. Risks and mitigations (additive)
   * 2 points: good list of risks and mitigations. Minimal description of to what extent mitigations successfully reduce risk (none for Cohere Command specifically) and no justification of non-mitigated risks.
   * Sources: <https://docs.cohere.com/docs/usage-guidelines> ; https://docs.cohere.com/docs/model-limitations; <https://docs.cohere.com/docs/data-statement> ; <https://arxiv.org/abs/2108.07790>
8. Evaluations (additive)
   * 2 points: accuracy is measured on multiple benchmarks as is bias. Intentional harms, robustness, calibration, UX are not evaluated.
   * Sources: <https://docs.cohere.com/docs/generation-card>
9. Testing (additive)
   * 2 points: Cohere Command has test results via HELM and Cohere states that it undergoes third party audits (though does not reveal the results). No other significant disclosure of process re: internal testing or details on red-teaming or stress testing. So: relatively few results reported
   * Sources: <https://docs.cohere.com/docs/the-command-model> ; <https://cohere.com/security>
10. Machine-generated content
    * 3 points: Cohere’s platform makes clear it is machine generated content. Cohere’s usage policy says that in order to use Cohere’s APIs, users must not engage in “unsafe unsupervised uses…No transparency: Applications that do not disclose that the content is generated through automated means” or “Misrepresentation: Representing raw generations as coming from humans, using supervised generations with false identities, or a single person using generations with many identities that appear to be independent.” But no disclosed mechanism for watermarking.
    * Source: <https://docs.cohere.com/docs/usage-guidelines> ; <https://dashboard.cohere.ai/playground/generate>
      1. Note: The output itself from the dashboard could do a much better job of indicating it is AI generated (e.g., by leading with “Cohere Command is an AI system that cannot give an object answer on X” or similar)
11. Member states
    * 0 points: No description of deployment practices in relation to the EU.
    * Source: <https://cohere.com/terms-of-use>
12. Downstream documentation
    * 3 points. Model card (for Generation model, not for Cohere Command), usage policy, and data statement are all highly useful documentation. But does not provide close to sufficient information for users to comply with each provision in the AI act.
    * Sources: <https://docs.cohere.com/docs/usage-guidelines> ; <https://docs.cohere.com/docs/generation-card> ; <https://docs.cohere.com/docs/data-statement>
      1. Note: this should maybe be a 2. Assessing the quality of the downstream technical documentation for use is difficult for me as I am not using that documentation.