Scores for Cohere on the 2023 Foundation Model Transparency Index

Background

1. Please see the paper describing the Foundation Model Transparency Index in order to understand what this document includes. The paper provides necessary background on (i) what these indicators are and why they were chosen, (ii) our standardized process for scoring the transparency of foundation model developers, and (iii) what these scores mean in context.

2. This document contains only information that was publicly available before September 15, 2023. It has not been updated and should be interpreted as a snapshot of transparency as of September 15, 2023.

3. In order to assess the transparency of foundation model developers, we used a rigorous, standardized search protocol to find publicly available information related to these indicators. You can find more information about this search protocol in the paper describing the Foundation Model Transparency Index.

4. We evaluate every company in this same way–you can find scoring documents for the other companies here.

5. We evaluate each company on 100 indicators of transparency. You can find the definition of each indicator and additional information about how each indicator was scored here.

6. Scores for each indicator are either 0 or 1. If the score is a 0, we do not provide a source for the score because our standardized search protocol (which includes many relevant sources) did not yield enough information to award a point. If the score is a 1, we provide a source that includes the information we cite in the justification for the score.

7. We evaluate each company on the basis of its flagship foundation model; in the case of Cohere, we evaluate Command.

8. In advance of releasing the Foundation Model Transparency Index, we reached out to Cohere for comment (along with the 9 other companies we evaluated) and offered an opportunity to provide feedback on the index and the organization's scores.

Scores for Each Indicator

1. Upstream → Data → <u>Data Size</u>
   - Score: 1
   - Justification: Cohere states that coheretext-filtered is ~200GB in its Data Statement and that it is used to train their Generation models.
   - Source: https://web.archive.org/web/20230831064017/https://docs.cohere.com/docs/data-statement

2. Upstream → Data → <u>Data Sources</u>
   - Score: 0
   - Justification: Underlying sources are not provided in the data statement or elsewhere, though they are described as "similar in composition to many other large, Internet-sourced language modeling datasets".
   - Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

3. Upstream → Data → <u>Data Creators</u>
   - Score: 0
   - Justification: Information is not provided on creators in the data statement or elsewhere, though data is described as "similar in composition to many other large, Internet-sourced language modeling datasets".
   - Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

4. Upstream → Data → <u>Data Source Selection</u>
   - Score: 0
   - Justification: No information found related to selection protocols for including and excluding data sources disclosed.
   - Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

5. Upstream → Data → <u>Data Curation</u>
   - Score: 1
   - Justification: Cohere provides documentation on coheretext curation practices.
   - Source: https://web.archive.org/web/20230831064017/https://docs.cohere.com/docs/data-statement#document-curation

6. Upstream → Data → Data Augmentation
   ○ Score: 0
   ○ Justification: No information found related to steps Cohere takes to augment its data sources.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

7. Upstream → Data → Harmful Data Filtration
   ○ Score: 0
   ○ Justification: Cohere discloses prior work on harm filtration that is documented in a paper (Ngo et al., 2021), which would be sufficient to award this point, but does not confirm the filters used to produce the coheretext-filtered dataset are the same as (or how they relate to) the filters documented in Ngo et al. (2021).
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

8. Upstream → Data → Copyrighted data
   ○ Score: 0
   ○ Justification: No information found regarding the copyright status of the data used to build Command.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

9. Upstream → Data → Data License
   ○ Score: 0
   ○ Justification: No information found regarding the license status of the data used to build Command.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

10. Upstream → Data → Personal Information in Data
   ○ Score: 0
   ○ Justification: Cohere discloses its policy for making use of user data and how commercial customers can access opt-out flags; however, no discussion of personal information in other data sources (e.g. Internet-based data) is provided, as is required by this indicator.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

11. Upstream → Data Labor → <u>Use of Human Labor</u>
    ○ Score: 0
    ○ Justification: No information found about the phases of the data pipeline where human labor is involved.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

12. Upstream → Data Labor → <u>Employers of Data Laborers</u>
    ○ Score: 0
    ○ Justification: No information found about the organization that directly employs the people involved in data labor.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

13. Upstream → Data Labor → <u>Geographic Distribution of Data Laborers</u>
    ○ Score: 0
    ○ Justification: No information found about geographic distribution of data laborers.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

14. Upstream → Data Labor → <u>Wages</u>
    ○ Score: 0
    ○ Justification: No information found about the wages for people who perform data labor.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

15. Upstream → Data Labor → <u>Instructions For Creating Data</u>
    ○ Score: 0
    ○ Justification: No information found about the instructions given to people who create the data used to build Command.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

16. Upstream → Data Labor → <u>Labor Protections</u>
    ○ Score: 0
    ○ Justification: No information found about labor protections for data laborers.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

17. Upstream → Data Labor → <u>Third Party Partners</u>
   ○ Score: 0
   ○ Justification: No information found about any third parties who were or are involved in the development of Command.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

18. Upstream → Data Access → <u>Queryable External Data Access</u>
   ○ Score: 0
   ○ Justification: No information found about whether external entities are provided with queryable access to the data used to build Command.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

19. Upstream → Data Access → <u>Direct External Data Access</u>
   ○ Score: 0
   ○ Justification: No information found about whether external entities are provided with direct access to the data used to build Command.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

20. Upstream → Compute → <u>Compute Usage</u>
   ○ Score: 0
   ○ Justification: No information found related to compute usage. Some compute information is provided for models other than Command (e.g. the base-light model and some embedding models) tied to environmental impact documentation.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

21. Upstream → Compute → <u>Development Duration</u>
   ○ Score: 0
   ○ Justification: No information found about the time required to build Command. Some compute information is provided for models other than Command (e.g. the base-light model and some embedding models) tied to environmental impact documentation.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

22. Upstream → Compute → <u>Compute Hardware</u>
   ○ Score: 0
   ○ Justification: No information found related to compute hardware. Some compute information is provided for models other than Command (e.g. the base-light model and some embedding models) tied to environmental impact documentation.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

23. Upstream → Compute → <u>Hardware Owner</u>
   ○ Score: 1
   ○ Justification: Cohere discloses all models are trained on Google Cloud TPUs. We do note that the page, as written, does not make clear if the page is updated in a way that would show this statement holds for Command in particular.
   ○ Source: https://docs.cohere.com/docs/environmental-impact

24. Upstream → Compute → <u>Energy Usage</u>
   ○ Score: 0
   ○ Justification: No information found related to energy usage. Some compute information is provided for models other than Command (e.g. the base-light model and some embedding models) tied to environmental impact documentation.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

25. Upstream → Compute → <u>Carbon Emissions</u>
   ○ Score: 0
   ○ Justification: No information found about the amount of carbon emitted in building Command, though the emissions for other models is disclosed.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

26. Upstream → Compute → <u>Broader Environmental Impact</u>
   ○ Score: 0
   ○ Justification: No information found any broader environmental impacts from building the model.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

27. Upstream → Methods → Model Stages
   ○ Score: 0
   ○ Justification: No information found about the stages involved in the model development process.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

28. Upstream → Methods → Model Objectives
   ○ Score: 0
   ○ Justification: No information found about the learning objectives for each stage in the model development process.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

29. Upstream → Methods → Core Frameworks
   ○ Score: 0
   ○ Justification: No information found about the core frameworks used for model development and data.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

30. Upstream → Methods → Additional Dependencies
   ○ Score: 0
   ○ Justification: No information found about additional dependencies required to build the model.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

31. Upstream → Data Mitigations → Mitigations for Personally Identifiable Information
   ○ Score: 0
   ○ Justification: Cohere discloses how it uses and permits opt-out for user data, but this does not clarify how personal information and privacy concerns are mitigated in the data used to train Command (given no user data is disclosed as being part of the data used to train Command).
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

32. Upstream → Data Mitigations → <u>Mitigations for Copyright</u>
   ○ Score: 0
   ○ Justification: No information found about steps Cohere takes to mitigate the presence of copyrighted information in the data.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

33. Model → Model Basics → <u>Input Modality</u>
   ○ Score: 1
   ○ Justification: The input modality for Command is text (i.e. it "responds well with instruction-like prompts")
   ○ Source: https://web.archive.org/web/20230831063026/https://docs.cohere.com/docs/command-beta and https://docs.cohere.com/docs/models

34. Model → Model Basics → <u>Output Modality</u>
   ○ Score: 1
   ○ Justification: The output modality Command is text. Cohere states "Command is Cohere's flagship text generation model. It is trained to follow user commands and to be instantly useful in practical business applications."
   ○ Source: https://cohere.com/models/command

35. Model → Model Basics → <u>Model Components</u>
   ○ Score: 0
   ○ Justification: No information found about the components of the model.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

36. Model → Model Basics → <u>Model Size</u>
   ○ Score: 0
   ○ Justification: Architectures are disclosed in existing model cards, but existing model cards (namely the Generation model card) do not indicate they apply to the Command model and other documentation distinguishes between the two types of models.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

37. Model → Model Basics → Model Architecture
   ○ Score: 0
   ○ Justification: No information found about the overall structure and organization of Command.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

38. Model → Model Basics → Centralized Model Documentation
   ○ Score: 1
   ○ Justification: Documentation, when available, is predominantly in the Cohere API documentation website, and mostly centralized under the "Responsible Use" category, including Generation model card.
   ○ Source: https://web.archive.org/web/20230831064552/https://docs.cohere.com/docs/responsible-use

39. Model → Model Access → External Model Access Protocol
   ○ Score: 1
   ○ Justification: The Cohere for AI Research Grant Application provides a form and criteria for granting external academic researchers model access.
   ○ Source: https://txt.cohere.com/c4ai-research-grants/

40. Model → Model Access → Black Box External Model Access
   ○ Score: 1
   ○ Justification: Command is available via an API, so black box external access is possible.
   ○ Source: https://web.archive.org/web/20230606004425/https://docs.cohere.com/reference/generate

41. Model → Model Access → Full External Model Access
   ○ Score: 0
   ○ Justification: Full access to the model is not provided to external entities via model weights.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

42. Model → Capabilities → <u>Capabilities Description</u>
   ○ Score: 1
   ○ Justification: Tutorials provide some description of capabilities.
   ○ Source:
     http://web.archive.org/web/20230720183957/https://docs.cohere.com/docs/quick-start-guides

43. Model → Capabilities → <u>Capabilities Demonstration</u>
   ○ Score: 1
   ○ Justification: Tutorials provide some demonstration of capabilities.
   ○ Source:
     http://web.archive.org/web/20230720183957/https://docs.cohere.com/docs/quick-start-guides

44. Model → Capabilities → <u>Evaluation of Capabilities</u>
   ○ Score: 0
   ○ Justification: No information found related to rigorous evaluations of the Command's capabilities.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

45. Model → Capabilities → <u>External Reproducibility of Capabilities Evaluation</u>
   ○ Score: 0
   ○ Justification: No information found indicating third parties have evaluated Command's capabilities.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

46. Model → Capabilities → <u>Third Party Capabilities Evaluation</u>
   ○ Score: 1
   ○ Justification: Cohere discloses that Stanford's Center for Research on Foundation Models conducted a third-party evaluation of capabilities on the HELM benchmark.
   ○ Source:
     http://web.archive.org/web/20230914092039/https://docs.cohere.com/docs/the-command-model

47. Model → Limitations → <u>Limitations Description</u>
   ○ Score: 1
   ○ Justification: Cohere lists limitations related to language, information about the Global South, historical information, biases, and statistical generation of model outputs.
   ○ Source: https://docs.cohere.com/docs/model-limitations

48. Model → Limitations → <u>Limitations Demonstration</u>
   ○ Score: 1
   ○ Justification: Examples are laid out for multiple limitations.
   ○ Source: https://docs.cohere.com/docs/model-limitations

49. Model → Limitations → <u>Third-Party Evaluation of Limitations</u>
   ○ Score: 1
   ○ Justification: Third parties can evaluate limitations if given access to Cohere's API.
   ○ Source: https://web.archive.org/web/20230606004425/https://docs.cohere.com/reference/generate

50. Model → Risks → <u>Risks Description</u>
   ○ Score: 1
   ○ Justification: Risks are described in Generation model card, namely for unintentional harms (toxic content generation, social biases) and intentional harms (astroturfing, misinformation, posing as a human).
   ○ Source: https://web.archive.org/web/20230824222334/https://docs.cohere.com/docs/generation-card

51. Model → Risks → <u>Risks Demonstration</u>
   ○ Score: 0
   ○ Justification: No information found related to explicit demonstrations of described risks.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

52. Model → Risks → <u>Unintentional Harm Evaluation</u>
   ○ Score: 0
   ○ Justification: No information found related to rigorous evaluations of the potential for the model to facilitate unintentional harm.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

53. Model → Risks → <u>External Reproducibility of Unintentional Harm Evaluation</u>
    ○ Score: 0
    ○ Justification: No information found related to the reproducibility of unintentional harm evaluations as there is no information found about the evaluations themselves.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

54. Model → Risks → <u>Intentional Harm Evaluation</u>
    ○ Score: 0
    ○ Justification: No information found related to rigorous evaluations of the potential for the model to facilitate intentional harm.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

55. Model → Risks → <u>External Reproducibility of Intentional Harm Evaluation</u>
    ○ Score: 0
    ○ Justification: No information found related to the reproducibility of intentional harm evaluations as there is no information found about the evaluations themselves.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

56. Model → Risks → <u>Third-Party Risk Evaluation</u>
    ○ Score: 1
    ○ Justification: Cohere points to external evaluation on the HELM benchmark from Stanford's Center for Research on Foundation Models, which evaluates risks such as toxicity and bias.
    ○ Source: https://docs.cohere.com/docs/the-command-model#cutting-edge-performance

57. Model → Mitigations → <u>Mitigations Description</u>
- ○ Score: 1
- ○ Justification: Cohere's harm prevention section of its responsible use documentation states "We aim to mitigate adverse use of our models with the following: Responsible AI Research: We've established a dedicated safety team which conducts research and development to build safer language models, and we're investing in technical (e.g., usage monitoring) and non-technical (e.g., a dedicated team reviewing use cases) measures to mitigate potential harms. // Cohere Responsibility Council: We've established an external advisory council made up of experts who work with us to ensure that the technology we're building is deployed safely for everyone. // No online learning: To safeguard model integrity and prevent underlying models from being poisoned with harmful content by adversarial actors, user input goes through curation and enrichment prior to integration with training."
- ○ Source: https://web.archive.org/web/20230914091917/http://web.archive.org/screenshot/https://docs.cohere.com/docs/responsible-use#harm-prevention

58. Model → Model Mitigations → <u>Mitigations Demonstration</u>
- ○ Score: 0
- ○ Justification: No information found related to illustrative examples of mitigations.
- ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

59. Model → Model Mitigations → <u>Mitigations Evaluation</u>
- ○ Score: 0
- ○ Justification: No information found related to rigorous evaluations of model mitigations.
- ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

60. Model → Model Mitigations → <u>External Reproducibility of Mitigations Evaluation</u>
- ○ Score: 0
- ○ Justification: No information found to indicate mitigations evaluations are externally reproducible.
- ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

61. Model → Model Mitigations → <u>Third Party Mitigations Evaluation</u>
   ○ Score: 0
   ○ Justification: No information found to indicate mitigations can be evaluated by third parties.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

62. Model → Trustworthiness → <u>Trustworthiness Evaluation</u>
   ○ Score: 1
   ○ Justification: Cohere points to evaluations of Command on the HELM benchmark, an external evaluation conducted by Stanford's Center for Research on Foundation Models that includes evaluations of calibration and robustness.
   ○ Source: Cohere's documentation points to HELM evaluation here https://cohere.com/models/command

63. Model → Trustworthiness → <u>External Reproducibility of Trustworthiness Evaluation</u>
   ○ Score: 1
   ○ Justification: Cohere points to evaluations of Command on the HELM benchmark, an external evaluation conducted by Stanford's Center for Research on Foundation Models that includes standard benchmarks of calibration and robustness and gives detailed protocols for how to reproduce evaluations.
   ○ Source: Cohere's documentation points to HELM evaluation here https://cohere.com/models/command

64. Model → Inference → <u>Inference Duration Evaluation</u>
   ○ Score: 0
   ○ Justification: No information found related to the time required for model inference.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

65. Model → Inference → <u>Inference Compute Evaluation</u>
   ○ Score: 0
   ○ Justification: No information found related to compute usage for model inference.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

66. Downstream → Distribution → <u>Release Decision-Making</u>
   ○ Score: 0
   ○ Justification: No information found about Cohere's protocol for deciding to release Command.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

67. Downstream → Distribution → <u>Release Process</u>
   ○ Score: 0
   ○ Justification: No information found about the process of how Command was released.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

68. Downstream → Distribution → <u>Distribution Channels</u>
   ○ Score: 1
   ○ Justification: Cohere specifies that its platform (and products on Amazon sagemaker) constitute its intended distribution channels.
   ○ Source: https://web.archive.org/web/20230720183938/https://docs.cohere.com/docs/the-cohere-platform and https://web.archive.org/web/20230720183938/https://docs.cohere.com/docs/amazon-sagemaker-setup-guide

69. Downstream → Distribution → <u>Products and Services</u>
   ○ Score: 1
   ○ Justification: Cohere discloses that Coral is powered by Command.
   ○ Source: https://web.archive.org/web/20230914103701/https://txt.cohere.com/introducing-coral/

70. Downstream → Distribution → <u>Detection of Machine-Generated Content</u>
   ○ Score: 0
   ○ Justification: No information found related to a mechanism for the detection of content generated by Command.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

71. Downstream → Distribution → <u>Model License</u>
- ○ Score: 1
- ○ Justification: Commercial SaaS Agreement constitutes a type of model license.
- ○ Source: https://web.archive.org/web/20230914095803/https://cohere.com/saas-agreement

72. Downstream → Distribution → <u>Terms of Service</u>
- ○ Score: 1
- ○ Justification: Cohere's Terms of Use satisfy this indicator.
- ○ Source: https://cohere.com/terms-of-use

73. Downstream → Usage Policy → <u>Permitted, Restricted, and Prohibited Users</u>
- ○ Score: 1
- ○ Justification: Cohere's Terms of Use ban direct competitors from using Cohere's models without prior consent, constituting a disclosure of prohibited users.
- ○ Source: https://cohere.com/terms-of-use

74. Downstream → Usage Policy → <u>Permitted, Restricted, and Prohibited Uses</u>
- ○ Score: 1
- ○ Justification: Usage guidelines outline many prohibited uses, Generation model card outlines permitted uses.
- ○ Source: https://web.archive.org/web/20230914091906/https://docs.cohere.com/docs/usage-guidelines and https://web.archive.org/web/20230914092045/https://docs.cohere.com

75. Downstream → Usage Policy → <u>Usage Policy Enforcement</u>
- ○ Score: 0
- ○ Justification: Cohere discloses that API access may be restricted or terminated, but this is insufficient to satisfy this indicator.
- ○ Source: https://cohere.com/terms-of-use and https://docs.cohere.com/docs/usage-guidelines

76. Downstream → Usage Policy → <u>Justification for Enforcement Action</u>
- ○ Score: 0
- ○ Justification: No information found related to whether users receive a justification when they are subject to an enforcement action for violating the usage policy.
- ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

77. Downstream → Usage Policy → <u>Usage Policy Violation Appeals Mechanism</u>
    ○ Score: 0
    ○ Justification: No information found related to a mechanism for appealing the developer's decision that the usage policy has been violated.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

78. Downstream → Model Behavior Policy → <u>Permitted, Restricted, and Prohibited Model Behaviors</u>
    ○ Score: 0
    ○ Justification: No information found related to model behaviors that are permitted, restricted, and prohibited.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

79. Downstream → Model Behavior Policy → <u>Model Behavior Policy Enforcement</u>
    ○ Score: 0
    ○ Justification: No information found related to the enforcement protocol for the model behavior policy.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

80. Downstream → Model Behavior Policy → <u>Interoperability of Usage and Model Behavior Policies</u>
    ○ Score: 0
    ○ Justification: No information found about how the usage policy and the model behavior policy interoperate.
    ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

81. Downstream → User Interface → <u>User Interaction with AI System</u>
    ○ Score: 1
    ○ Justification: On the playground, there are many indications that content is machine generated (e.g., clear input-output delineation) and that a user is interacting with an AI system.
    ○ Source: https://web.archive.org/web/20230914094557/https://docs.cohere.com/docs/playground-overview and https://dashboard.cohere.com/playground/generate

82. Downstream → User Interface → <u>Usage Disclaimers</u>
   ○ Score: 1
   ○ Justification: Upon sign-up for playground, Cohere's platform says "By signing up, you agree to the Terms of Use".
   ○ Source: https://dashboard.cohere.com/welcome/login?redirect_uri=%2Fplayground%2Fgenerate

83. Downstream → User Data Protection → <u>User Data Protection Policy</u>
   ○ Score: 1
   ○ Justification: Cohere's privacy policy includes protocols for collecting, using, and transferring user data (namely personal information). Cohere's data usage policy also addresses these issues.
   ○ Source: https://web.archive.org/web/20230914095709/https://cohere.com/data-usage-policy and https://web.archive.org/web/20230914095757/https://cohere.com/privacy

84. Downstream → User Data Protection → <u>Permitted and Prohibited Use of User Data</u>
   ○ Score: 1
   ○ Justification: The data usage policy has sections on "What do we do with personal information customers share with us?" and states that service providers with which Cohere shares user data "are not authorized to use or disclose personal information for their own marketing or other purposes."
   ○ Source: https://web.archive.org/web/20230914095709/https://cohere.com/data-usage-policy

85. Downstream → User Data Protection → <u>Usage Data Access Protocol</u>
   ○ Score: 0
   ○ Justification: No information found related to a protocol for granting external entities access to usage data.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

86. Downstream → Model Updates → <u>Versioning Protocol</u>
   ○ Score: 0
   ○ Justification: No information found related to a protocol for uniquely identifying each instance of Command.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

87. Downstream → Model Updates → <u>Change Log</u>
   ○ Score: 1
   ○ Justification: Cohere discloses a high-level change log with release notes.
   ○ Source: https://docs.cohere.com/changelog

88. Downstream → Model Updates → <u>Deprecation Policy</u>
   ○ Score: 0
   ○ Justification: No information found related to a description of what it means for Command or other foundation models from Cohere to be deprecated and how users should respond to the deprecation.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

89. Downstream → Feedback → <u>Feedback Mechanism</u>
   ○ Score: 1
   ○ Justification: Cohere provides a feedback endpoint, a thumbs up option on the playground, and a contact email.
   ○ Source: https://web.archive.org/web/20230914094557/https://docs.cohere.com/docs/playground-overview and https://web.archive.org/web/20230914101743/https://docs.cohere.com

90. Downstream → Feedback → <u>Feedback Summary</u>
   ○ Score: 0
   ○ Justification: No information found related to any summary of user feedback.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

91. Downstream → Feedback→ <u>Government Inquiries</u>
   ○ Score: 0
   ○ Justification: No information found in connection with a summary of government inquiries related to the model.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

92. Downstream → Impact → Monitoring Mechanism
   ○ Score: 1
   ○ Justification: Data usage policy says "We do not use raw input prompts or model generations to profile how our models are being used. We may use aggregated metadata, such as a count of input prompt tokens, to understand usage and improve our system performance."
   ○ Source: https://web.archive.org/web/20230914095709/https://cohere.com/data-usage-policy

93. Downstream → Impact → Downstream Applications
   ○ Score: 0
   ○ Justification: No information found related to the number of applications dependent on the foundation model.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

94. Downstream → Impact → Affected Market Sectors
   ○ Score: 0
   ○ Justification: No information found related to the fraction of applications corresponding to each market sector.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

95. Downstream → Impact → Affected Individuals
   ○ Score: 0
   ○ Justification: No information found related to the number of individuals affected by the model.
   ○ Source: Search protocol produced no source containing sufficient information to award this point

96. Downstream → Impact → Usage Reports
   ○ Score: 0
   ○ Justification: No information found related to usage statistics describing the impact of the model on users.
   ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

97. Downstream → Impact → <u>Geographic Statistics</u>
- ○ Score: 0
- ○ Justification: No information found regarding statistics of model usage across geographies.
- ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

98. Downstream → Impact → <u>Redress Mechanism</u>
- ○ Score: 0
- ○ Justification: No information found regarding any mechanism to provide redress to users for harm caused by the model.
- ○ Source: No source is provided because the search protocol did not produce a source containing sufficient information to award this point.

99. Downstream → Documentation for Deployers → <u>Centralized Documentation for Downstream Use</u>
- ○ Score: 1
- ○ Justification: Centralized hub with usage guidelines, responsible use information, and API documentation is sufficiently centralized
- ○ Source: https://web.archive.org/web/20230914094632/https://docs.cohere.com/reference/about and https://web.archive.org/web/20230914092011/https://docs.cohere.com/docs/data-statement and https://web.archive.org/web/20230914091906/https://docs.cohere.com/docs/usage-guidelines

100. Downstream → Documentation for Deployers → <u>Documentation for Responsible Downstream Use</u>
- ○ Score: 1
- ○ Justification: Responsible use guidelines constitute such documentation
- ○ Source: https://web.archive.org/web/20230914091906/https://docs.cohere.com/docs/usage-guidelines