Model Indicators for the 2023 Foundation Model Transparency Index Model **Input modality:** Are the input modalities for the model disclosed? **Output modality:** Are the output modalities for the model disclosed? **Model components:** Are all components of the model disclosed? **Model size:** For all components of the model, is the associated model size disclosed? **Model architecture:** Is the model architecture disclosed? Centralized model documentation: Is key information about the model included in a centralized artifact such as a model card? External model access protocol: Is a protocol for granting external entities access to the model disclosed? Blackbox external model access: Is black box model access provided to external entities? Full external model access: Is full model access provided to external entities? Capabilities description: Are the model's capabilities described? Capabilities demonstration: Are the model's capabilities demonstrated? Evaluation of capabilities: Are the model's capabilities rigorously evaluated, with the results of these evaluations reported prior to or concurrent with the initial release of the model? **External reproducibility of capabilities evaluation:** Are the evaluations of the model's capabilities reproducible by external entities? **Third-party capabilities evaluation:** Are the model's capabilities evaluated by third parties? **Limitations description:** Are the model's limitations disclosed? Limitations demonstration: Are the model's limitations demonstrated? Third-party evaluation of limitations: Can the model's limitations be evaluated by third parties? **Risks description:** Are the model's risks disclosed? Risks demonstration: Are the model's risks demonstrated? Unintentional harm evaluation: Are the model's risks related to unintentional harm rigorously evaluated, with the results of these evaluations reported prior to or concurrent with the initial release of the model? External reproducibility of unintentional harm evaluation: Are the evaluations of the model's risks related to unintentional harm reproducible by external entities? Intentional harm evaluation: Are the model's risks related to intentional harm rigorously evaluated, with the results of these evaluations reported prior to or concurrent with the initial release of the model?. External reproducibility of intentional harm evaluation: Are the evaluations of the model's risks related to intentional harm reproducible by external entities? Third-party risks evaluation: Are the model's risks evaluated by third parties? Mitigations description: Are the model mitigations disclosed? **Mitigations demonstration:** Are the model mitigations demonstrated? Mitigations evaluation: Are the model mitigations rigorously evaluated, with the results of these evaluations reported? External reproducibility of mitigations evaluation: Are the model mitigation evaluations reproducible by external entities?

Third-party mitigations evaluation: Can the model mitigations be evaluated by third parties?

Trustworthiness evaluation: Is the trustworthiness of the model rigorously evaluated, with the results of these evaluations disclosed?

External reproducibility of trustworthiness evaluation: Are the trustworthiness evaluations reproducible by external entities?

Inference duration evaluation: Is the time required for model inference disclosed for a clearly-specified task on a clearly-specified set of hardware?

Inference compute evaluation: Is the compute usage for model inference disclosed for a clearly-specified task on a clearly-specified set of hardware?