2023 Foundation Model Transparency Index Indicators

Data size Input modality Release decision-making Data sources Output modality Release decision-making Data sources Output modality Release process Data creators Model components Distribution channels Data sources election Model size Products and services Data curation Model architecture Detection of machine-generated content Data augmentation Centralized model documentation Model License Harmful data filtration External model access protocol Terms of service Copyrighted data Blackbox atternal model access Permitted and prohibited users Data license Permitted, restricted, and prohibited user Personal information in data Capabilities demonstration Usage policy enforcement Use of human labor Capabilities demonstration Justification for enforcement exciton Usage policy enforcement External reproducibility of cepabilities evaluation Representation for the provided of the provided provided in the provided provided in the provided provided provided in the provided pro	Upstream	Model	Downstream
Data sources Data creators Model components Data source selection Model size Products and services Data cursion Model architecture Data augmentation Centralized model documentation Model License Harmful data filtration External model access protocol Terms of service Data license Blackbox external model access Permitted and prohibited users Personal information in data Use of human labor Capabilities description Use of human labor Capabilities of human labor			
Data creators Model components Distribution channels Data source selection Model size Products and services Data curation Model architecture Detection of machine-generated content Data augmentation Centralized model documentation Model License Harmful data filtration External model access protocol Terms of service Copyrighted data Blackbox external model access Permitted and prohibited users Para license Full external model access Permitted, and prohibited uses Personal information in data Capabilities description Usage policy enforcement Use of human labor Capabilities demonstration Justification for enforcement action Employment of data laborers Evaluation of capabilities External reproducibility of capabilities evaluation External reproducibility of capabilities evaluation Instructions for creating data Limitations description Usage policy enforcement Instructions for creating data Limitations description User interoperability of usage and model behavior policies Labor protections Limitations description User interoperability of usage and model behavior policies Labor protections Limitations description User interoperability of usage and model behavior policies Compute water data access Risks description User interaction with Al system Usage disclaimers Ourayable external data access Risks demonstration User data protection policy Permitted and prohibited use of user data Development duration External reproducibility of unintentional harm evaluation Compute hardware Intentional harm evaluation Permitted and prohibited use of user data Usage data access protocol Versioning protocol Third party risks evaluation Feedback mechanism Feedback mechanism Feedback mechanism Feedback mechanism Feedback summary Feedback summary Feedback summary Feedback summary Feedback summary Feedback summary Feedback mechanism Feedback mechanism Feedback summary F		<u> </u>	
Data source selection Model size Data curation Model architecture Data augmentation Centralized model documentation Model License Harmful data filtration External model access protocol Terms of service Copyrighted data Blackbox external model access Permitted and prohibited users Parsonal information in data Capabilities description Use of human labor Capabilities demonstration External reproducibility of capabilities evaluation Employment of data laborers Evaluation of capabilities evaluation Employment of data laborers Evaluation of capabilities evaluation External reproducibilities evaluation Instructions for creating data Limitations description Limitations description Limitations description Third party partners Third party evaluation of limitations Queryable external data access Risks description Direct external data access Risks demonstration External reproducibility of unintentional harm evaluation Compute usage Unintentional harm evaluation External reproducibility of unintentional harm evaluation Compute hardware Intentional harm evaluation External reproducibility of intentional harm evaluation Compute hardware Intentional harm evaluation External reproducibility of intentional harm evaluation Capabilities evaluation Change log Third party risks evaluation Change log Deprecation policy Permitted and prohibited use of user data Change log Unintentional harm evaluation Change log Deprecation policy Feedback summary Feedback summary Government inquiries Model stages Mitigations description Mitigations evaluation Affected market sectors Additional dependencies Third party mitigations evaluation Affected individuals Mitigations for copyright Inference compute evaluation Inference compute evaluation Interence understance in the machinal probability and interestic		· · · · · · · · · · · · · · · · · · ·	<u> </u>
Data curation Data sugmentation Centralized model documentation Model License Harmful data filtration External model access protocol Forms of service Copyrighted data Blackbox external model access Permitted and prohibited users Data license Full external model access Permitted, restricted, and prohibited users Data license Personal information in data Capabilities description Use of human labor Capabilities description Use of human labor External reproducibility of capabilities Geographic distribution of data laborers Full external reproducibility of capabilities evaluation Geographic distribution of data laborers Third party capabilities evaluation Business of creating data Instructions for creating data Limitations description Limitations description Limitations demonstration Limitations demonstration Limitations description User interaction with AI system Third party partners Third party evaluation of limitations Usage disclaimers Queryable external data access Risks description Direct external data access Risks demonstration Fermitted and prohibited use of user data Usage Direct external data access Direct external data access Risks demonstration Development duration External reproducibility of unintentional harm evaluation Permitted and prohibited use of user data Usage data access protocol Versioning protocol User data protocol Versioning protocol Third party risks evaluation Development duration External reproducibility of intentional harm evaluation Feedback summany Readback mechanism Mitigations description Mitigations description Feedback summany Feedback summa		·	
Data augmentation			
Harmful data filtration			-
Description of the party capabilities description User interaction policies			
Data license Full external model access Personal information in data Capabilities description Usage policy enforcement Use of human labor Capabilities description Usage policy enforcement Use of human labor Capabilities demonstration Usage policy enforcement action Employment of data laborers Evaluation of capabilities Usage policy violation appeals mechanism Geographic distribution of data laborers Capabilities evaluation Permitted, restricted, and prohibited model behaviors Permitted, restricted, and prohibited model behavior policies Permitted, restricted, and prohibited model behavior policies Permitted, restricted, and prohibited model behavior policies Permitted, restricted, and prohibited model pehavior policies Permitted, restricted, and prohibited model behaviors Permitted, restricted, and prohibited model behaviors Permitted, restricted, and prohibited model behaviors Permitted, restricted, and prohibited model behavior policies Permitted, restricted, and prohibited used model behavior policies Permitted, restricted, and prohibited model behavior policies Permitted, restricted, and prohibited used of user data Permitted, restricted, and prohibited used user data Permitted, restricted, and prohibited used user data Permitted, restricted, and prohibited used user data Permitted, rest		•	
Personal information in data Use of human labor Capabilities description Usage policy enforcement Use of human labor Capabilities Employment of data laborers Cagaphic distribution of data laborers External reproducibility of capabilities evaluation Usage policy violation appeals mechanism Permitted, restricted, and prohibited model behaviors Dehaviors Wages Third party capabilities evaluation Usage policy enforcement Instructions for creating data Limitations description Limitations description User interaction with Al system Usage disclaimers Usage disclaimers Queryable external data access Risks demonstration User data protection policy Direct external data access Risks demonstration Usage data access protocol External reproducibility of unintentional harm evaluation External reproducibility of unintentional harm evaluation Compute usage Unintentional harm evaluation External reproducibility of intentional harm evaluation Compute hardware Intentional harm evaluation Compute hardware Intentional harm evaluation External reproducibility of mitigations evaluation Feedback mechanism Model stages Mitigations description Model stages Mitigations evaluation External reproducibility of mitigations evaluation Downstream applications Affected individuals Mitigations for copyright Inference duration evaluation Inference compute evaluation Redress mechanism	., -		
Use of human labor Capabilities demonstration Justification for enforcement action Employment of data laborers Evaluation of capabilities Usage policy violation appeals mechanism Geographic distribution of data laborers Wages Third party capabilities evaluation Model behavior policy enforcement Instructions for creating data Limitations description Interoparability of usage and model behavior policies Labor protections Limitations demonstration User interaction with Al system Third party partners Third party evaluation of limitations Usage disclaimers Gueryable external data access Risks description User data protection policy Direct external data access Risks demonstration User data protection policy Direct external data access Unintentional harm evaluation Usage data access protocol Development duration External reproducibility of unintentional harm evaluation Compute hardware Intentional harm evaluation Change log Hardware owner External reproducibility of intentional harm evaluation Deprecation policy Energy usage Third party risks evaluation Feedback mechanism Garbon emissions Mitigations description Feedback mechanism Broader environmental impact Mitigations demonstration Monitoring mechanism Model stages Mitigations evaluation Downstream applications Model stages Mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Geographic statistics Mitigations for copyright Inference duration evaluation Geographic statistics Inference compute evaluation Redress mechanism			·
Employment of data laborers Geographic distribution of data laborers Geographic distribution of data laborers Wages Third party capabilities evaluation Instructions for creating data Limitations description Limitations demonstration User interaction with Al system Third party partners Third party partners Third party evaluation of limitations User data protection policy Direct external data access Risks description User data protection policy Direct external data access Risks demonstration Permitted and prohibited use of user data Compute usage Unintentional harm evaluation External reproducibility of unintentional harm evaluation Compute hardware Intentional harm evaluation External reproducibility of intentional harm evaluation Perection policy External reproducibility of intentional harm evaluation Energy usage Third party risks evaluation Feedback mechanism Carbon emissions Mitigations description Model objectives External reproducibility of mitigations evaluation Model objectives External reproducibility of mitigations evaluation Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Affected market sectors Affected market sectors Mitigations for copyright Inference duration evaluation Redress mechanism Redress mechanism Redress mechanism		<u> </u>	
Recographic distribution of data laborers		<u> </u>	
Baborers	Employment of data laborers	Evaluation of capabilities	Usage policy violation appeals mechanism
Instructions for creating data Limitations description Limitations demonstration Limitations demonstration User interaction with Al system Third party partners Third party evaluation of limitations Usage disclaimers Queryable external data access Risks description User data protection policy Permitted and prohibited use of user data Compute usage Unintentional harm evaluation External reproducibility of unintentional harm evaluation Compute hardware Intentional harm evaluation External reproducibility of intentional harm evaluation Energy usage Third party risks evaluation Feedback mechanism Carbon emissions Mitigations description Model stages Mitigations evaluation Monotring mechanism Model objectives External reproducibility of mitigations evaluation Downstream applications Affected market sectors Additional dependencies Trustworthiness evaluation Mitigations for privacy External reproducibility of trustworthiness evaluation Mitigations for copyright Inference compute evaluation Redress mechanism		External reproducibility of capabilities evaluation	·
Labor protections Limitations demonstration User interaction with Al system Third party partners Third party evaluation of limitations Usage disclaimers Queryable external data access Risks description User data protection policy Direct external data access Risks demonstration Permitted and prohibited use of user data Compute usage Unintentional harm evaluation Usage data access protocol Development duration External reproducibility of unintentional harm evaluation Compute hardware Intentional harm evaluation External reproducibility of intentional harm evaluation Deprecation policy Energy usage Third party risks evaluation Feedback mechanism Carbon emissions Mitigations description Feedback summary Broader environmental impact Mitigations evaluation Model stages Mitigations evaluation Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Mitigations for privacy External reproducibility of trustworthiness evaluation Mitigations for copyright Inference compute evaluation Redress mechanism Redress mechanism	Wages	Third party capabilities evaluation	Model behavior policy enforcement
Third party partners Queryable external data access Risks description Direct external data access Risks demonstration Compute usage Unintentional harm evaluation External reproducibility of unintentional harm evaluation Compute hardware Hardware owner External reproducibility of intentional harm evaluation External reproducibility of intentional harm evaluation Energy usage Third party risks evaluation Carbon emissions Mitigations description Feedback summary Broader environmental impact Mitigations evaluation Model stages Mitigations evaluation Model objectives External reproducibility of mitigations evaluation Government inquiries Model objectives External reproducibility of mitigations evaluation Model objectives External reproducibility of trustworthiness evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Mitigations for privacy External reproducibility of trustworthiness evaluation Mitigations for copyright Inference duration evaluation Redress mechanism	Instructions for creating data	Limitations description	
Queryable external data access Risks description User data protection policy Direct external data access Risks demonstration Permitted and prohibited use of user data Compute usage Unintentional harm evaluation Usage data access protocol Development duration External reproducibility of unintentional harm Versioning protocol Compute hardware Intentional harm evaluation Change log Hardware owner External reproducibility of intentional harm evaluation Deprecation policy Energy usage Third party risks evaluation Feedback mechanism Carbon emissions Mitigations description Feedback summary Broader environmental impact Mitigations demonstration Government inquiries Model stages Mitigations evaluation Monitoring mechanism Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Affected individuals Mitigations for privacy External reproducibility of trustworthiness evaluation Geographic statistics	Labor protections	Limitations demonstration	User interaction with AI system
Direct external data access Risks demonstration Compute usage Unintentional harm evaluation External reproducibility of unintentional harm evaluation Compute hardware Intentional harm evaluation External reproducibility of intentional harm evaluation Change log Hardware owner External reproducibility of intentional harm evaluation Energy usage Third party risks evaluation Carbon emissions Mitigations description Broader environmental impact Model stages Mitigations evaluation Model objectives External reproducibility of mitigations evaluation Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Mitigations for privacy External reproducibility of trustworthiness evaluation Mitigations for copyright Inference duration evaluation Redress mechanism	Third party partners	Third party evaluation of limitations	Usage disclaimers
Compute usage Unintentional harm evaluation Usage data access protocol Development duration External reproducibility of unintentional harm evaluation Versioning protocol Compute hardware Intentional harm evaluation Change log Hardware owner External reproducibility of intentional harm evaluation Deprecation policy Energy usage Third party risks evaluation Feedback mechanism Carbon emissions Mitigations description Feedback summary Broader environmental impact Mitigations demonstration Government inquiries Model stages Mitigations evaluation Monitoring mechanism Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Affected individuals Mitigations for privacy External reproducibility of trustworthiness evaluation Usage reports Mitigations for copyright Inference duration evaluation Redress mechanism	Queryable external data access	Risks description	User data protection policy
Development duration External reproducibility of unintentional harm evaluation Compute hardware Intentional harm evaluation External reproducibility of intentional harm evaluation External reproducibility of intentional harm evaluation Energy usage Third party risks evaluation Feedback mechanism Carbon emissions Mitigations description Freedback summary Broader environmental impact Mitigations demonstration Model stages Mitigations evaluation Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Mitigations for privacy External reproducibility of trustworthiness evaluation Usage reports Mitigations for copyright Inference duration evaluation Redress mechanism	Direct external data access	Risks demonstration	Permitted and prohibited use of user data
evaluation Compute hardware Intentional harm evaluation Change log Hardware owner External reproducibility of intentional harm evaluation Deprecation policy Energy usage Third party risks evaluation Feedback mechanism Carbon emissions Mitigations description Feedback summary Broader environmental impact Mitigations demonstration Government inquiries Model stages Mitigations evaluation Monitoring mechanism Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Affected individuals Mitigations for privacy External reproducibility of trustworthiness evaluation Usage reports Mitigations for copyright Inference duration evaluation Redress mechanism	Compute usage	Unintentional harm evaluation	Usage data access protocol
Hardware owner External reproducibility of intentional harm evaluation Deprecation policy Energy usage Third party risks evaluation Feedback mechanism Carbon emissions Mitigations description Feedback summary Broader environmental impact Mitigations demonstration Government inquiries Model stages Mitigations evaluation Monitoring mechanism Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Affected individuals Mitigations for privacy External reproducibility of trustworthiness evaluation Usage reports Mitigations for copyright Inference duration evaluation Geographic statistics Inference compute evaluation Redress mechanism	Development duration		Versioning protocol
Energy usage Third party risks evaluation Feedback mechanism Carbon emissions Mitigations description Feedback summary Broader environmental impact Mitigations demonstration Government inquiries Model stages Mitigations evaluation Monitoring mechanism Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Affected individuals Mitigations for privacy External reproducibility of trustworthiness evaluation Usage reports Mitigations for copyright Inference duration evaluation Geographic statistics Inference compute evaluation Redress mechanism	Compute hardware	Intentional harm evaluation	Change log
Carbon emissions Mitigations description Feedback summary Broader environmental impact Mitigations demonstration Government inquiries Model stages Mitigations evaluation Monitoring mechanism Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Affected individuals Mitigations for privacy External reproducibility of trustworthiness evaluation Usage reports Mitigations for copyright Inference duration evaluation Geographic statistics Inference compute evaluation Redress mechanism	Hardware owner	External reproducibility of intentional harm evaluation	Deprecation policy
Broader environmental impact Mitigations demonstration Government inquiries Model stages Mitigations evaluation Monitoring mechanism Model objectives External reproducibility of mitigations evaluation Downstream applications Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Affected individuals Mitigations for privacy External reproducibility of trustworthiness evaluation Usage reports Mitigations for copyright Inference duration evaluation Geographic statistics Inference compute evaluation Redress mechanism	Energy usage	Third party risks evaluation	Feedback mechanism
Model stagesMitigations evaluationMonitoring mechanismModel objectivesExternal reproducibility of mitigations evaluationDownstream applicationsCore frameworksThird party mitigations evaluationAffected market sectorsAdditional dependenciesTrustworthiness evaluationAffected individualsMitigations for privacyExternal reproducibility of trustworthiness evaluationUsage reportsMitigations for copyrightInference duration evaluationGeographic statisticsInference compute evaluationRedress mechanism	Carbon emissions	Mitigations description	Feedback summary
Model objectivesExternal reproducibility of mitigations evaluationDownstream applicationsCore frameworksThird party mitigations evaluationAffected market sectorsAdditional dependenciesTrustworthiness evaluationAffected individualsMitigations for privacyExternal reproducibility of trustworthiness evaluationUsage reportsMitigations for copyrightInference duration evaluationGeographic statisticsInference compute evaluationRedress mechanism	Broader environmental impact	Mitigations demonstration	Government inquiries
Core frameworks Third party mitigations evaluation Affected market sectors Additional dependencies Trustworthiness evaluation Mitigations for privacy External reproducibility of trustworthiness evaluation Mitigations for copyright Inference duration evaluation Inference compute evaluation Redress mechanism	Model stages	Mitigations evaluation	Monitoring mechanism
Additional dependencies Trustworthiness evaluation Affected individuals Mitigations for privacy External reproducibility of trustworthiness evaluation Usage reports Mitigations for copyright Inference duration evaluation Geographic statistics Inference compute evaluation Redress mechanism	Model objectives	External reproducibility of mitigations evaluation	Downstream applications
Mitigations for privacy External reproducibility of trustworthiness evaluation Usage reports Geographic statistics Inference compute evaluation Redress mechanism	Core frameworks	Third party mitigations evaluation	Affected market sectors
Mitigations for copyright Inference duration evaluation Geographic statistics Inference compute evaluation Redress mechanism	Additional dependencies	Trustworthiness evaluation	Affected individuals
Inference compute evaluation Redress mechanism	Mitigations for privacy	External reproducibility of trustworthiness evaluation	Usage reports
·			
Centralized documentation for downstream use		Inference compute evaluation	Redress mechanism
			Centralized documentation for downstream use
Documentation for responsible downstream use			Documentation for responsible downstream use