2023 Foundation Model Transparency Index Indicators

Upstream	Model	Downstream
Data size	Input modality	Release decision-making
Data sources	Output modality	Release process
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
Data license	Full external model access	Permitted, restricted, 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	Usage policy violation appeals mechanism
Geographic distribution of data laborers	External reproducibility of capabilities evaluation	Permitted, restricted, and prohibited model behaviors
Wages	Third party capabilities evaluation	Model behavior policy enforcement
Instructions for creating data	Limitations description	Interoperability of usage and model behavior policies
Labor protections	Limitations demonstration	User interaction with AI 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	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	Geographic statistics
	Inference compute evaluation	Redress mechanism
		Centralized documentation for downstream use
		Documentation for responsible downstream use