Model Indicators for the 2025 Foundation Model Transparency Index

Indicator	Definition
Basic model properties	Are all basic model properties disclosed?
Deeper model properties	Is a detailed description of the model architecture disclosed?
Model dependencies	Is the model(s) the model is derived from disclosed?
Benchmarked inference	Is the compute and time required for model inference disclosed for a clearly-specified task on clearly-specified hardware?
Researcher credits	Is a protocol for granting external entities API credits for the model disclosed?
Specialized access	Does the developer disclose if it provides specialized access to the model?
Open weights	Are the model's weights openly released?
Agent Protocols	Are the agent protocols supported for the model disclosed?
Capabilities taxonomy	Are the specific capabilities or tasks that were optimized for during post-training disclosed?
Capabilities evaluation	Does the developer evaluate the model's capabilities prior to its release and disclose them concurrent with release?
External reproducibility of capabilities evaluation	Are code and prompts that allow for an external reproduction of the evaluation of model capabilities disclosed?
Train-test overlap	Does the developer measure and disclose the overlap between the training set and the dataset used to evaluate model capabilities?
Risks taxonomy	Are the risks considered when developing the model disclosed?
Risks evaluation	Does the developer evaluate the model's risks prior to its release and disclose them concurrent with release?
External reproducibility of risks evaluation	Are code and prompts to allow for an external reproduction of the evaluation of model risks disclosed?
Pre-deployment risk evaluation	Are the external entities have evaluated the model pre-deployment disclosed?
External risk evaluation	Are the parties contracted to evaluated model risks disclosed?
Mitigations taxonomy	Are the post-training mitigations implemented when developing the model disclosed?
Mitigations taxonomy mapped to risk taxonomy	Does the developer disclose how the post-training mitigations map onto the taxonomy of risks?
Mitigations efficacy	Does the developer evaluate and disclose the impact of post-training mitigations?
External reproducibility of mitigations evaluation	Are code and prompts to allow for an external reproduction of the evaluation of post-training mitigations disclosed?
Model theft prevention measures	Does the developer disclose the security measures used to prevent unauthorized copying (“theft”) or unauthorized public release of the model weights?
Release stages	Are the stages of the model's release disclosed?
Risk thresholds	Are risk thresholds disclosed?
Versioning protocol	Is there a disclosed protocol for versioning and deprecation of the model?
Change log	Is there a disclosed change log for the model?
Foundation model roadmap	Is a forward-looking roadmap for upcoming models, features, or products disclosed?
Top distribution channels	Are the top-5 distribution channels for the model disclosed?
Quantization	Is the quantization of the model served to customers in the top-5 distribution channels disclosed?
Terms of use	Are the terms of use of the model disclosed?