




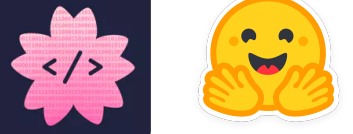










Foundation Model Transparency Index Indicator-Level Scores for Model, May 2024

Source: May 2024 Foundation Model Transparency Index

		 ADEPT	 AI21 Labs	 ALEPH ALPHA	 amazon	 ANTHROPIC	 servicenow	 Google	 IBM	 Meta	 Microsoft	 MISTRAL AI	 OpenAI	 stability.ai	 WRITER
Subdomain	Indicator	Fuyu-8B	Jurassic-2	Luminous	Titan Text Express	Claude 3	StarCoder	Gemini 1.0 Ultra	Granite	Llama 2	Phi-2	Mistral 7B	GPT-4	Stable Video Diffusion	Palmyra-X
Model Basics	Input modality	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Output modality	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Model components	0	1	1	1	0	1	1	1	1	1	1	0	1	1
	Model size	1	1	1	0	0	1	0	1	1	1	1	0	1	1
	Model architecture	1	1	1	1	0	1	1	1	1	1	1	0	1	1
	Centralized model documentation	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Model Access	External model access protocol	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Blackbox external model access	1	1	1	1	1	1	1	1	1	1	1	1	1	0
	Full external model access	1	0	1	0	0	1	0	0	1	1	1	0	1	0
Capabilities	Capabilities description	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Capabilities demonstration	1	1	1	1	1	1	1	0	1	1	1	1	0	1
	Evaluation of capabilities	1	0	1	1	1	1	1	1	1	1	1	1	1	1
	External reproducibility of capabilities evaluation	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Third-party capabilities evaluation	0	1	1	0	1	1	0	0	1	1	1	1	0	1
Limitations	Limitations description	1	1	1	1	1	1	0	1	1	1	0	1	1	1
	Limitations demonstration	0	1	1	1	0	0	0	0	0	0	0	0	0	0
	Third-party evaluation of limitations	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Risks	Risks description	0	1	1	1	1	1	1	1	1	0	1	1	1	1
	Risks demonstration	0	1	1	1	1	1	1	0	1	0	0	1	0	0
	Unintentional harm evaluation	0	0	0	0	1	1	0	1	1	1	0	0	0	0
	External reproducibility of unintentional harm evaluation	0	1	1	1	1	1	0	1	1	1	0	0	0	0
	Intentional harm evaluation	0	0	0	0	1	1	1	1	0	0	0	0	0	0
	External reproducibility of intentional harm evaluation	0	0	0	0	0	1	0	1	0	0	0	1	0	0
	Third-party risks evaluation	0	1	1	0	1	1	0	0	1	0	0	1	0	0
	Mitigations description	0	1	1	1	1	0	1	1	1	0	1	1	0	1
Mitigations	Mitigations demonstration	0	1	0	0	1	0	0	0	1	0	1	1	0	0
	Mitigations evaluation	0	0	0	0	0	0	0	1	1	0	1	1	0	0
	External reproducibility of mitigations evaluation	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	Third-party mitigations evaluation	0	0	0	0	0	0	1	1	0	0	0	0	0	0
Trustworthiness	Trustworthiness evaluation	0	1	1	1	1	0	1	0	0	0	0	1	0	0
	External reproducibility of trustworthiness evaluation	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Inference	Inference duration evaluation	0	1	1	0	0	1	0	0	1	0	0	0	1	1
	Inference compute evaluation	0	1	0	0	0	1	0	0	0	0	0	0	0	1
Model Subtotal		42%	76%	76%	58%	64%	76%	55%	67%	76%	55%	58%	61%	48%	52%