Foundation Model Transparency Index Indicator-Level Scores for Model, 2023

Source: 2023 Foundation Model Transparency Index

		∞ Meta	BigScience	- ⑤ OpenAl	stability.ai	Google	ANTHROP\C	c ohere	Al21 labs	Inflection	amazon
Subdomain	Indicator	Llama 2	BLOOMZ	GPT-4	Stable Diffusion 2		Claude 2	Command	Jurassic-2	Inflection-1	Titan Text
Model Basics	Input modality	1	1	1	1	1	1	1	1	1	1
	Output modality	1	1	1	1	1	1	1	1	1	1
	Model components	1	1	0	1	0	0	0	0	0	0
	Model size	1	1	0	0	0	0	0	0	0	0
	Model architecture	1	1	0	1	1	1	0	0	0	0
	Centralized model documentation	1	1	1	1	1	1	1	0	1	0
Model Access	External model access protocol	1	1	1	1	0	0	1	0	0	0
	Blackbox external model access	1	1	1	1	1	1	1	1	0	1
	Full external model access	1	1	0	1	0	0	0	0	0	0
Capabilities	Capabilities description	0	1	1	1	1	1	1	1	1	1
	Capabilities demonstration	1	1	1	1	1	1	1	1	0	0
	Evaluation of capabilities	1	1	1	0	1	1	0	0	1	0
	External reproducibility of capabilities evaluation	1	1	1	0	1	1	0	0	0	0
	Third-party capabilities evaluation	0	0	1	0	0	0	1	1	0	0
Limitations	Limitations description	1	1	1	1	1	1	1	0	1	0
	Limitations demonstration	0	0	0	0	0	0	1	0	0	0
	Third-party evaluation of limitations	1	1	1	1	1	1	1	1	0	1
Risks	Risks description	1	0	1	1	0	1	1	1	0	0
	Risks demonstration	1	0	1	0	0	0	0	0	0	0
	Unintentional harm evaluation	1	0	0	0	1	0	0	0	0	0
	External reproducibility of unintentional harm evaluation	1	0	0	0	1	0	0	0	0	0
	Intentional harm evaluation	0	0	0	0	0	0	0	0	0	0
	External reproducibility of intentional harm evaluation	0	0	1	0	0	0	0	0	0	0
	Third-party risks evaluation	0	0	1	0	0	1	1	1	0	0
Model Mitigations	Mitigations description	1	0	1	0	1	1	1	0	1	1
	Mitigations demonstration	1	0	1	0	0	0	0	0	0	0
Mitigations	Mitigations evaluation	1	0	1	0	1	1	0	0	0	0
	External reproducibility of mitigations evaluation	0	0	0	0	0	0	0	0	0	0
	Third-party mitigations evaluation	0	0	0	0	0	0	0	0	0	0
Trustworthiness	Trustworthiness evaluation	0	0	1	0	0	0	1	1	0	0
	External reproducibility of trustworthiness evaluation	0	0	0	0	0	0	1	1	0	0
Inference	Inference duration evaluation	1	1	0	0	0	0	0	0	0	0
	Inference compute evaluation	0	0	0	0	0	0	0	0	0	0
	Model Subtotal	67%	48%	61%	39%	45%	45%	48%	33%	21%	18%