

Score ID	Criteria	Weight	Anthropic	OpenAI	Google DeepMind	Meta	x.AI
	2. Risk Analysis and Evaluation	25%	26%	34%	31%	36%	31%
	2.1 Setting a Risk Tolerance	35%	7%	16%	3%	24%	57%
	2.1.1 Risk tolerance is defined	80%	8%	20%	3%	30%	71%
C13	2.1.1.1 Risk tolerance is at least qualitatively defined for most risk	33%	25%	50%	10%	90%	90%
C14	2.1.1.2 Risk tolerance is expressed fully quantitatively (cf. criterion above) or at least partly quantitatively as a combination of scenarios (qualitative) and probabilities (quantitative) for most risk	33%	0%	10%	0%	0%	75%
C15	2.1.1.3 Risk tolerance is expressed fully quantitatively as a product of severity (quantitative) and probability (quantitative) for most risks	33%	0%	0%	0%	0%	50%
	2.1.2 Process to define the tolerance	20%	0%	0%	0%	0%	0%
C16	2.1.2.1 AI developers engage in public consultations or seek guidance from regulators where available.	50%	0%	0%	0%	0%	0%
C17	2.1.2.2 Any significant deviations from risk tolerance norms established in other industries are justified and documented (e.g., cost-benefit analyses)	50%	0%	0%	0%	0%	0%
	2.2 Operationalizing Risk Tolerance	65%	36%	44%	47%	43%	18%
	2.2.1 Key Risk Indicators (KRI)	30%	51%	51%	51%	33%	33%
C18	2.2.1.1 KRI thresholds are at least qualitatively defined for most risks	45%	90%	90%	90%	50%	50%
C19	2.2.1.2 KRIs thresholds are quantitatively defined for most risks	45%	50%	25%	10%	0%	75%
C20	2.2.1.3 KRI also identifies and monitors changes in the level of risk in the external environment	10%	0%	0%	0%	10%	0%
	2.2.2 Key Control Indicators (KCI)	30%	31%	45%	38%	18%	14%
	2.2.2.1 Containment KCIs	35%	43%	5%	63%	38%	5%
C21	2.2.2.1.1 Most KRI thresholds have corresponding qualitative containment KCI thresholds	50%	75%	10%	75%	75%	10%
C22	2.2.2.1.2 Most KRI thresholds have corresponding quantitative containment KCI thresholds	50%	10%	0%	50%	0%	0%
	2.2.2.2 Deployment KCIs	35%	38%	45%	25%	13%	25%
C23	2.2.2.2.1 Most KRI thresholds have corresponding qualitative deployment KCI thresholds	50%	75%	90%	50%	25%	50%
C24	2.2.2.2.2 Most KRI thresholds have corresponding quantitative deployment KCI thresholds	50%	0%	0%	0%	0%	0%
C25	2.2.2.3 For advanced KRIs, Assurance processes KCI are defined	30%	10%	90%	25%	0%	10%
C26	2.2.3 Pairs of thresholds are grounded in risk modeling to show that risks remain below the tolerance	20%	10%	50%	90%	50%	10%
C27	2.2.4 Policy to put development on hold if the required KCI threshold cannot be achieved, until sufficient controls are implemented to meet the threshold	20%	50%	25%	10%	90%	10%