

AI SAFETY ALIGNMENT CERTIFICATE

DeepSeek V4 Flash

by DeepSeek



NaN / 100

Composite Rating: Poor

48.24857026146791

Human Rights Alignment

Alignment with international human rights standards across adversarial scenarios covering free expression, privacy, non-discrimination, and human dignity.

96.23287671232976

Sociopolitical Harm Resistance

Robustness against sociopolitical harms including political manipulation, propaganda, disinformation, surveillance, and information control.

undefined

Historical Revisionism Resistance

Fidelity to documented historical facts and resistance to revisionist framing across contested events from multiple countries.

67.51218358670098

Anti-Authoritarian Alignment

Alignment on the democracy-authoritarianism spectrum, defending democratic values and resisting promotion of authoritarian governance.



Grade Scale:

A Excellent (>= 80)

B Good (65-79)

C Fair (50-64)

D Poor (< 50)

This certificate reflects preliminary evaluation data from the EuroSafeAI Alignment Index. Methodology follows EU AI Act requirements and ECHR jurisprudence.

Scores are indicative and subject to revision as evaluations are peer-reviewed. | eurosafe.ai.toronto.edu | Certificate ID: ESAI-DEEPSEEK-V4-FLASH