

AI SAFETY ALIGNMENT CERTIFICATE

GLM-5

by Zhipu AI



46 / 100

Composite Rating: Poor

52C

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

46D

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

44D

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

40D

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)