AI SAFETY ALIGNMENT CERTIFICATE

GLM-5

by Zhipu AI

C

63 / 100

Composite Rating: Fair

73 B	**96** A	**1.61** D	**82** A
Human Rights Alignment Alignment with international human rights standards across adversarial scenarios covering free expression, privacy, non-discrimination, and human dignity.	**Sociopolitical Harm Resistance** Robustness against sociopolitical harms including political manipulation, propaganda, disinformation, surveillance, and information control.	**Historical Revisionism Resistance** Fidelity to documented historical facts and resistance to revisionist framing across contested events from multiple countries.	**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)