

AIR-Bench: refusal rate

Source: Zeng et al., 2024 | Chart: 2025 AI Index report

