

Policy & Safety Guardrails (cross-cutting): PII minimization • harmful-content filtering • anti-cheating • human escalation

