3. Validation Plan

Systematic approach & framework
Start with small, controlled tests of each agent, then run the full chat flow end-to-end using thousands of scripted cases. Use a large library of de-identified real cases and clinician-written scenarios, including hard cases like pediatrics, pregnancy, medication interactions, and vague symptoms. Before any patient sees it, run the system in silent mode inside real clinics so clinicians can review its outputs without them reaching patients.

Safety & effectiveness measures
Safety is measured by how reliably it flags emergencies, how often it misses a red flag, and how often it produces unsafe or incorrect advice. Effectiveness is measured by clinician agreement with the assessment and plan, patient understanding of the advice, and whether the chat leads to timely follow-up care.

What convinces hospital admins
Clear, pre-set thresholds are met across multiple sites, with independent clinician review and complete audit logs. There is a staged rollout plan with human override, continuous monitoring, and a fast rollback process. Privacy and security controls are verified and documented.