Pre-Screen Assignment

American Express · Digital Banking · July 2026

Enterprise KB Chatbot
Product Strategy & Roadmap

Udit Sehgal · Pre-Screen Assignment · Technical Product Owner

01 Problem & Users

1a *What problem is the chatbot solving?* Two people need a trustworthy answer from American Express and can't get one fast. A **prospect** comparing cards can't get straight answers about benefits and fit, so they bounce or call — and the application stalls. A **cardmember** with a servicing question finds the answer already exists somewhere in the help center, but can't reach it fast, so they call or chat. Both show up as avoidable calls — and, for prospects, lost applications.

1b *Who are the primary users?* **Cardmembers *and* prospects.**

- **Prospects (pre-login).** Honest, grounded answers about benefits and fit, help comparing, then a hand-off to apply. It educates — it doesn't give regulated financial advice.
- **Cardmembers (post-login).** One correct, current answer in seconds — grounded servicing answers first, account-aware help once validated.

Also served along the way: care agents (answers beside them mid-call), operations teams, new hires, knowledge owners (what's missing or stale), and compliance partners (confidence it won't answer outside its lane).

1c *What are the key challenges today?*

Grounded in a first-hand review of the live Amex experience, July 2026 — the specifics are in "How it improves," below.

1. **Trust.** One confidently wrong answer and people stop using the bot. The failure mode isn't "no answer" — it's a made-up one.
2. **Coverage vs. quality.** Ingesting everything imports noise and contradictions along with the knowledge.
3. **Freshness.** A bot serving last quarter's policy is worse than no bot.
4. **Closing the loop.** Collecting a thumbs-up is easy; turning it into ranked retrieval fixes and content tickets is the hard part. A signal that only lands in a dashboard never improves the next answer.

02 Product Strategy

2a *What should this product aim to achieve?* **Trusted answers at the speed of conversation.** The product's job isn't to answer every question — it's to be *right when it answers, honest when it isn't sure, and measurably better every week.*

2b *How does it improve over existing solutions?* Help-center search returns pages and makes the user do the reading — this answers directly, with citations back to those same pages. The logged-in chat handles fixed servicing intents (transactions, disputes, balance) and hands open questions to humans — this adds the "how does X work" lane it doesn't cover. And where today's feedback is one star for the whole session, every answer here carries its own thumbs — so quality is attributable, and provably improves week over week.

Grounded, cited, or silent. Every answer comes strictly from retrieved source passages and shows its citations; when confidence is low, it returns the top sources instead of guessing. The single biggest driver of adoption.

The feedback loop is the product. Every answer carries its own thumbs-up/down, tied to the sources it cited — that's what turns feedback into ranked retrieval fixes and content tickets.

Guardrails as a feature. A hard, code-level firewall keeps the bot off topics it must never free-lance on (legal and compliance determinations) and routes them to a human with context. In a regulated enterprise, this is what makes deployment possible at all.

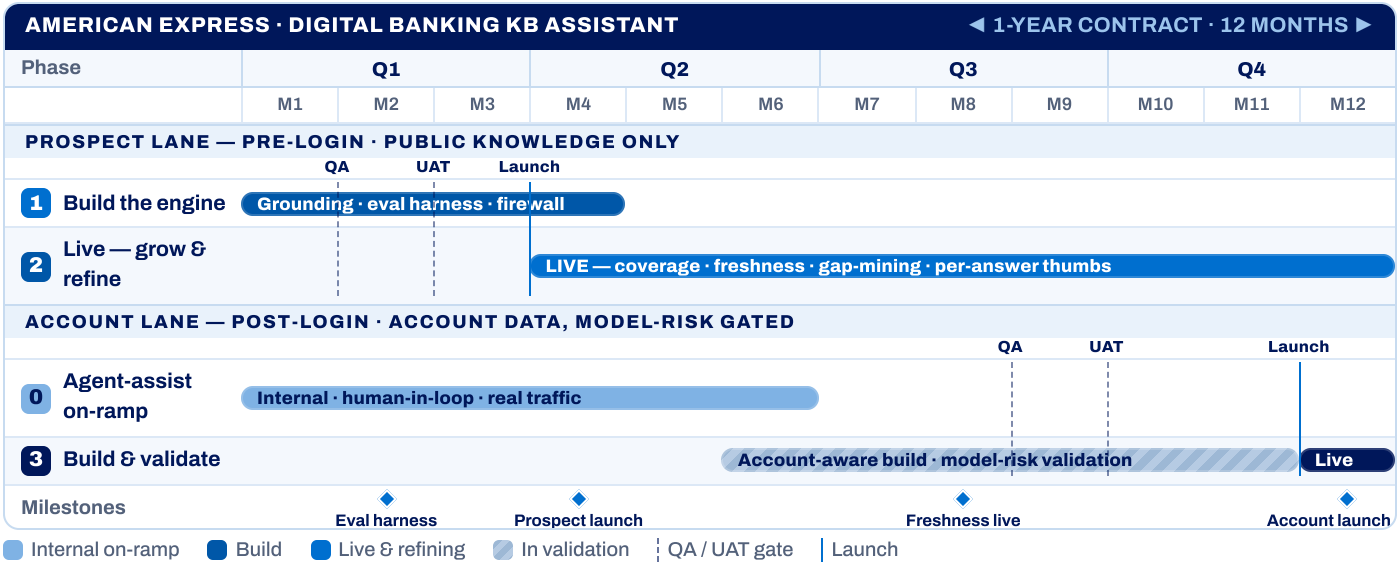
Where this fits — EP2, two surfaces, one engine.

This assistant is the **AI Assistant & Knowledge Experiences** capability of the EP2 self-service platform — a reusable capability alongside Capability Maps, Self-Service Services, and the Workflow Composer, not a one-off bot. **Pre-login** it educates prospects and hands off to apply (application starts, never financial advice); **post-login** it adds grounded knowledge to the servicing lane Amex runs today. Two front doors, one engine — **and the live demo is that engine.**

03 Roadmap (Phased)

3a *key features/capabilities — each **What** line* 3b *why they are prioritized — each **Why** line*

Two customer-facing lanes, each with its own build, hardening, and launch — riding one shared engine. The phase badges map to the phases below: the prospect lane builds and launches first (no account data); the account lane seasons on agent-assist, then builds and validates before it launches. Bars show relative timing, not committed dates.



0 Agent-assist (optional on-ramp)

What: the same grounded assistant, deployed beside care agents — the answer and its citations appear next to the agent, who stays the human in the loop.

Why: it proves quality on real traffic with zero customer-facing risk — the standard way regulated enterprises graduate GenAI, and a safe way to season the account lane before it ever faces a cardmember.

1 Trust Foundation

What: strict grounding with inline citations; source-search fallback when confidence is low; per-answer thumbs; an evaluation harness — a golden question set scored on every change; the sensitive-topic firewall.

Why first: trust is the adoption bottleneck — you can't improve what you can't measure. Once the foundation holds, the prospect lane ships (*Prospect launch*): no account data, so grounded-and-cited is enough.

2 Coverage & Freshness

What: new knowledge sources, each gated by an evaluation pass; a freshness pipeline (stale-content detection, change-triggered re-indexing); a dashboard for knowledge owners; gap-mining that turns unanswered questions into content tickets.

Why second: with the trust mechanics in place, coverage growth compounds value instead of compounding noise.

3 Scale & Action

What: role-aware and account-aware answers under full model-risk validation; action-taking for the top validated intents — open the ticket, not just describe it — turning answers into self-service journeys other Enterprise Platforms teams can compose against; more channels; latency and cost tuning.

Why third: personalization and action-taking multiply value but also multiply blast radius, so they're built last, on a measured, trusted foundation. This is where the account lane goes customer-facing (*Account launch*) — deliberately last, not sooner.

04 Prioritization — Top 5

4a *What are your top 5 priorities?*

#	Priority	Rationale
1	**Grounded answers with citations**	Attacks the #1 failure mode (hallucination); prerequisite for everything
2	**Evaluation harness + golden question set**	Makes quality measurable — every change becomes evidence, not opinion
3	**Confidence fallback ("cited or silent")**	Converts worst-case from *wrong answer* to *useful search result*
4	**Per-answer feedback loop (👍/👎)**	Ties each miss to a specific answer and its retrieval; feeds prioritization forever
5	**Freshness pipeline**	Stale answers erode trust as fast as wrong ones

4b *How did you decide (criteria)?* **Ranked by** impact on trust, reach, risk reduction, effort, and dependency order — measurement before optimization, trust before scale.

05 Trade-offs

5 *How would you approach each pair below? With a mechanism, not a vibe.*

Accuracy vs. speed. Cache validated answers for the common questions; spend extra retrieval passes on the long tail — a correct answer in 6 seconds beats a wrong one in 2. When retrieval still isn't confident, return sources instead of guessing.

Coverage vs. noise. A new source ships only if it raises answer quality on the golden set, not just document count. Better to be narrow and right, then widen deliberately — users forgive "I don't cover that yet"; they don't forgive confidently wrong.

Feature delivery vs. system reliability. Give answer quality a hard floor, measured on a continuously audited sample: when it's healthy, ship features; when it slips, fix quality first. New capabilities roll out to one segment first, with automatic rollback if ratings drop.

06 Success Metrics

6a *How would you measure success?* **One north star per audience.** Cardmembers: **self-serve resolution** — % of questions resolved without a human or ticket. Prospects: **assisted-apply conversion** — % of conversations that lead to a qualified application start.

Quality (leading): groundedness and citation accuracy on an audited sample — the trust number that gates every release; retrieval hit rate on the golden set — separates "can't find it" from "can't say it."

Engagement (leading): per-answer helpfulness (👍/👎) — the fastest user-truth signal, tied to the specific answer.

Outcomes (lagging): self-serve resolution and escalation deflection (the servicing case); assisted-apply conversion (the acquisition case).

Guardrails: zero sensitive-topic violations; content freshness; worst-case response time.

6b *What metrics matter most?* **Groundedness first** — it gates everything — then the two north stars. Every roadmap item above exists to move one of these numbers.

Postscript — we built it

Rather than only writing this strategy, we built it: a live Phase-1 banking assistant answering real questions from 80 indexed public amex.com help articles — every answer cites its live source, declines to guess when retrieval is weak, and takes a thumbs-up/down. The improvement loop ran for real: a 20-question test scored 17/20, drove same-day tuning, re-ran at 20/20 — this role's first line of quality. Build and test results are on the site.

Prefer to chat with the strategy instead of reading it?

demo.millionautomations.com/amex

Prepared by Udit Sehgal · July 2026 · Unofficial — not affiliated with or endorsed by American Express.