AI shows you answers. It doesn't show you what to question.

Perplexity built its trust on citations. 48% of users never engage with them. · 37 surveys · 5 interviews · 1 observation · all on Slide 2.

MARKET CONTEXT

~30M	700M	$47B	51%	Trust layer
Perplexity MAU	ChatGPT MAU	AI tool TAM 2026	AI use is work-related	= the unclaimed moat

THE DATA · WHAT 37 USERS TOLD US

73%
misled by AI output they trusted for real work
Q11 · 27 of 37

81%
caught AI being CONFIDENTLY wrong
Q8 · 30 of 37

62%
say polished AI feels MORE trustworthy
Q4 · 23 of 37

48%
heavy users skip the sources entirely
Q7 · 14 of 29

THE CONTRADICTION USERS CAN'T ARTICULATE

"Polish feels reliable" — but they pick CALIBRATED output when shown both.

49% chose calibrated Output B

30% picked A (decisive)

Of 23 polish-trusters → 13 still chose B.
People can't ASK for calibration.
They RECOGNIZE it when shown.

EVIDENCE BASE

n=37 valid · 78% heavy users · 95% ChatGPT · 24% Perplexity · Anonymized · Triangulated

THE STORY · ONE USER, ₹40K ON THE LINE

HERO OBSERVATION

Operations Analyst, 26, Mumbai

Task: "Should I invest ₹40,000 in a 6-month Go Lang course?"
Think-aloud protocol · Perplexity vs ChatGPT

ON PERPLEXITY:

0 sources clicked

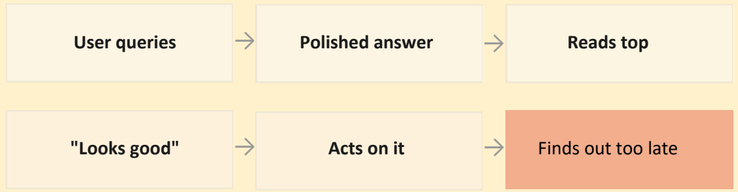
0 cross-checks · 0 re-prompts

7/10 post-task confidence

"I think it looks good."

— would have trusted a wrong answer equally.

THE BROKEN JOURNEY · TODAY



Zero friction at the decision moment. No signal to pause. No per-claim visibility.

THE URGENCY · WHY PERPLEXITY, WHY NOW

PERPLEXITY'S PROMISE

"The answer engine for research."

Citations were the original trust scaffold.
But 48% of heavy users never click them.

Citations ≠ Calibration. Showing sources isn't the same as helping users evaluate.

WHY NOW · THE WINDOW IS CLOSING

- 2025-26 inflection: Daily AI use crossed the frequency threshold. Polish caught up — accuracy has not.
- Every model launch widens the gap. GPT-5, Claude 4.5, Gemini 2.5 — same polish, no calibration.
- Citations were Perplexity's moat. Open-sourced now. Next moat: calibration.
- First mover wins. Second mover plays catch-up.

Defensive: when ChatGPT adds search, citations alone won't hold the moat.

WHAT'S AT STAKE

Trust failure isn't an edge case. It's the dominant pattern in daily AI use.

The trust problem isn't that AI is wrong sometimes.
It's that AI doesn't help users know WHEN to question it. SyncIQ closes that gap.

Users don't lack trust in AI. They lack a way to be calibrated.

37 surveys · 5 in-depth interviews · 1 task-based observation · 5 first-person Perplexity sessions

METHOD

PRIMARY SURVEY

n=37

78% heavy users (daily+)
95% ChatGPT · 68% Gemini
54% Claude · 24% Perplexity

→ Form → Raw data

5 IN-DEPTH INTERVIEWS · 30-45 min

ROLE · AGE	TOOLS	KEY PAIN
SAP Dev,	ChatGPT	Wrong
Backend	ChatGPT	"2 of 10
SDE, 25	ChatGPT-	Wrong
Ops	Perplexity+	0 clicks
Network	ChatGPT-	Resume

TASK OBSERVATION

"Decide whether to invest ₹40,000 in a 6-month Go Lang course"
Think-aloud · Perplexity vs ChatGPT
→ Protocol doc

WHAT USERS ACTUALLY DO — 8 BEHAVIORAL PATTERNS

1. SPOON-FEEDING TAX

4 of 5 interviewees re-prompted 3-5 times for adequate answer.

"It gives a one-shot answer of two lines. I have to keep telling it do this, do that." — SDE, 25, Mumbai

2. CITATION THEATER

OBSERVED: 0 source clicks during ₹40K purchase decision task.

"Out of 10 sources only 2 were useful." — Backend Dev, 25, Mumbai

3. INVISIBLE FRAMING → VALIDATES F2

AI assumes context about user's intent. Never reveals what it assumed.

"AI should first understand my context before answering." — Network Eng (CCNA), 27, Mumbai

4. STAKES-ADJUSTED TRUST

Users self-segment trust by risk. Code = trust. Terminal cmds = verify.

"For coding I trust AI. For terminal commands I cross-verify." — SDE, 25, Mumbai

5. THE REPETITION LOOP

AI gives the same wrong answer when corrected. Doesn't update.

"It was rolling around the same answer, but it was wrong." — Survey Q14

6. HALLUCINATION OF SATISFACTION

AI invents details to please the user. No flag that it's fabricating.

"It started hallucinating names — added numbers that weren't mine." — Survey Q14

7. DECISION PARALYSIS RISK

Too many options from AI causes choice overload.

"Too many options becomes difficult — which one we should pick?" — SAP Dev, 26, Mumbai

8. SKIM & CONFIDENCE

Read top of answer, feel 7/10 sure, never verify. Default pattern.

"I only read top few lines or one paragraph" — Backend Dev, 25, Mumbai

WHAT USERS WANT (Q13)

WHEN ASKED "WHAT WOULD HELP?"

68% want per-claim granular flags
→ *validates F1 design*

46% want assumptions surfaced
→ *validates F2 design*

46% want "3 expert Qs to ask"
→ *validates F3 design*

Only 8% said "current state is fine."

THE SELF-REPORT GAP

**Users believe they evaluate carefully.
The data says otherwise.**

62% claim "I read carefully"
22% admit: "looks right, so I trust it"
Self-report ≠ behavior. SyncIQ must work without trusting self-report.

RESEARCH VERDICT

The problem is experiential, not technical. The fix is a per-claim calibration layer — not documentation, not another AI, not trust scores.

H1 POLISH TRUST BIAS

VERDICT: STRONGLY CONFIRMED

62% find polish more trustworthy (Q4). 4 of 5 interviews confirm.

H2 CITATIONS = THEATER

VERDICT: CONFIRMED

"2 of 10 sources useful" + observation: 0 clicks during ₹40K decision.

H3 SOURCE CLICK GAP

VERDICT: STRONGLY CONFIRMED

48% of heavy users skip sources (Q7).
₹ 40K observation: 0 clicks.

H4 NO EVALUATION VOCAB

VERDICT: CONFIRMED

59% pick "well-written, clear" (Q5) - style criteria, not substance.

H5 COUNTER-INTUITIVE

VERDICT: REVERSED

Experts trust polish MORE, fooled by technical-sounding language.

H6 USERS WANT GRANULAR

VERDICT: STRONGLY CONFIRMED

68% want per-claim flags (Q13). Only 8% chose "current state fine."

8 behavioral patterns. 6 hypotheses tested. 3 feature-validating Q13 signals. 1 counter-intuitive reversal.

The problem isn't AI accuracy — it's calibration. Users want decomposed help, not aggregate trust scores.

Two daily users. One trust gap. Five distinct segments behind them.

75% of Perplexity Pro daily heavy users fall into 2 personas — but the underlying 37-user sample breaks into 5 distinct behavioral segments.

BEHAVIORAL LANDSCAPE

← *Acts on instinct* — *Cross-checks* →

QUICK TRUSTER
~40%
Trusts polish, acts fast, often burned.
Ops Analyst 26 · Network Eng 27 Mumbai

CAREFUL VERIFIER
~35%
Knows to check, cross-checks across tools, exhausted.
SDE 25 · Backend Dev 25 Mumbai

CASUAL DRIFTER
~15%
Low-stakes queries, calibration not yet needed.

AI EXPERT
~10%
Already self-calibrating, doesn't need SyncIQ.

WHY THESE TWO ARE OUR TARGET

40% + 35% = 75% of Perplexity Pro daily heavy users. SyncIQ ships to both.

QUICK TRUSTER · DAILY CONTEXT
Opens Perplexity 4-6x/day. Reads top of answer. Acts on first read. Rarely clicks sources. Burns ~$200/month on bad AI decisions.

CAREFUL VERIFIER · DAILY CONTEXT
Opens Perplexity 8-12x/day. Cross-checks with ChatGPT/Gemini. Spends 20+ min verifying single answers. Exhausted by manual eval.

PRIMARY PERSONAS

PRIMARY · QUICK TRUSTER



"It looks good. I'll go with this."

— Operations Analyst, 26, Mumbai (task observation participant)

JOBS-TO-BE-DONE

"When I get an AI answer for a high-stakes question, I want to know what to question — so I don't act on something hallucinated."

CORE PROBLEM

Doesn't know they need help. SyncIQ: Wake them up — surface the moment of decision they currently skip.

PRIMARY · CAREFUL VERIFIER



"For terminal commands I cross-verify on the official site : wrong command affects my system."

— Software Engineer, 25, Mumbai

JOBS-TO-BE-DONE

"When I get an AI answer, I want to instantly see which parts to verify — so I stop wasting 20 minutes cross-checking the whole thing."

CORE PROBLEM

Knows they need help, doesn't have a tool. SyncIQ: Make checking effortless — per-claim flags surface what to verify.

SIDE-BY-SIDE · WHAT MAKES THEM DIFFERENT, WHAT MAKES THEM THE SAME

DIMENSION	QUICK TRUSTER (40%)	CAREFUL VERIFIER (35%)
Core problem	Doesn't know they need help	Knows they need help, no tool
Trust trigger	Polish + speed	Multi-source agreement
Decision style	Skim-and-act	Cross-check-and-act
Success looks like	Started verifying ? marks before acting	Reduced cross-check time from 20min → 3min

5-SEGMENT BREAKDOWN · ANOTHER ANGLE ON THE SAME 37 USERS

Polish Trusters · 11/37 (30%)
Trust output based on polish. Picked Output A. Most receptive to F1 marks.
Ops Analyst 26 · Network Eng 27 Mumbai

Forced Cross-Checkers · 13/37 (35%)
Got burned before. Verify across multiple AIs. Most receptive to F3.
SDE 25 · Backend Dev 25 Mumbai

Selective Evaluators · 7/37 (19%)
Verify only on high-stakes. Most receptive to F2 assumption surfacing.
SAP Dev 26 Mumbai

Burned but Unchanged · 4/37 (11%)
Aware of problem, no behavior shift. Hidden growth segment.
Survey Q12 open-ended

AI-Skeptical Avoiders · 2/37 (5%)
Light AI users. Not a SyncIQ target — self-protecting.

Math check: 11+13+7+4+2 = 37 ✓ matches valid response count.

Two personas cover 75% of users. Five segments reveal the underlying behavioral distribution.

The trust problem isn't AI being wrong.
It's AI being opaque about what it doesn't know.

Four structural opacities mapped to a real market. Trust layer is unclaimed — for now.

PROBLEM CANVAS · 6 BLOCKS

TRUE PROBLEM

Users can't tell when to trust AI's polished output.

Result: over-trust on high-stakes decisions. Over-skepticism on simple ones. No calibration layer inside Perplexity.

81% have caught AI confidently wrong. 73% still got misled by AI they trusted.

Not a discovery problem. Not an accuracy problem. A calibration problem.

WHO FACES IT

75% of daily heavy AI users (see Slide 3).

Quick Trusters (40%) — don't know they need help · skim → "looks good" → act

Careful Verifiers (35%) — know they need help but no efficient tool exists

Profiles: Ops Analyst, SDE, Backend Dev, Network Eng, SAP Dev — all 25-27, Mumbai.

Daily-stakes use + high trust velocity.

HOW WE KNOW IT'S REAL

73 · 81 · 62 · 48%

misled · caught wrong · trust polish · skip

+ 5 in-depth interviews triangulating the same patterns across roles

+ Task observation: 0 source clicks on ₹40K decision · 7/10 post-task confidence

+ 5 personal Perplexity research sessions

5 hypotheses confirmed · 1 counter (H5). See Slide 2.

MARKET SOLUTIONS & GAPS

Three categories exist · none calibrate

All address output ACCURACY. None address user CALIBRATION.

Citations (Perplexity, You.com) — Decorative · 48% of heavy users skip them

Fact-checkers (Originality.ai) — Post-hoc · outside the answer flow

Verifier AIs (Devin, Cursor review) — Shifts trust target · doesn't help user judge

The "decision moment" has no product layer.

VALUE IF SOLVED

For users

Avoid wrong-decision losses (₹40K type)
Build evaluation skill over time
Cut cross-check time 40-60% for Verifiers
For Perplexity-
Trust = retention · Pro tier conversion lever
Differentiation moat as competition floods in
Defensive: when ChatGPT adds search, citations alone won't hold the moat

3-year revenue impact: $52M → $312M ARR

WHY NOW

Trust layer is unclaimed — but not for long

2025-26 inflection: Daily AI use crossed the frequency threshold. Polish caught up · accuracy has not.

Every model launch widens the gap. GPT-5, Claude 4.5, Gemini 2.5 — same polish.

Citations were Perplexity's moat. Open-sourced now. Next moat: calibration.

First mover wins. Second mover plays catch-up.

MARKET SIZING

TAM · AI ASSISTANT MARKET 2026

$47B

Projected $200B+ by 2030
Source: Grandview Research 2026

SAM · DAILY HIGH-STAKES AI USERS

~120M

Perplexity + ChatGPT + Claude + Gemini
Filter: daily use, high-stakes (work, research)

SOM · YEAR-1 ADDRESSABLE (Perplexity)

~22M

30M MAU · 75% daily heavy users
Per Slide 3 segmentation

3-YEAR ARR PROJECTION · PRO $20/MO

Y1: 1% SOM	$52M
Y2: 3% SOM + cross-AI	$158M
Y3: 6% SOM + partners	$312M

FOUR STRUCTURAL OPACITIES · WHAT THE CALIBRATION LAYER MUST CLOSE

1 · CORRECTNESS OPACITY

Which claims are right?

AI doesn't distinguish high-confidence claims from low-confidence ones.

Today: every claim looks equally certain.

2 · COMPLETENESS OPACITY

What's missing?

AI doesn't surface what it doesn't know about the user. Incomplete picture.

Today: the user is invisible to the answer.

3 · REASONING OPACITY

What did AI assume?

AI makes context assumptions silently. Framing may or may not match intent.

Today: reasoning is a black box.

4 · CALIBRATION OPACITY

What should I question?

AI doesn't help users build evaluation skill. Every output starts from zero.

Today: every output is its own island.

Four opacities. One root cause: AI is built to ANSWER, not to CALIBRATE.

Three ways to solve trust. Only one survives the brief.

And every flexible tool got structure the same way.

SyncIQ doesn't invent a new pattern. It applies 40 years of HCI calibration design to AI outputs — Perplexity's brand, not ChatGPT's.

THE PATTERN · REPEATED ACROSS DECADES

Every flexible tool that became too uncertain got a CALIBRATION LAYER on top.

1980s · SPELL-CHECK IN WORD PROCESSORS

Typing was flexible — anything was a word.

Red squiggle = "you might want to check this."

→ Per-word confidence marking. Same pattern as SyncIQ F1.

2004 · GOOGLE SEARCH AUTOCOMPLETE

Search queries became infinite.

Google predicted intent BEFORE you finished typing.

→ Surfacing assumptions about user intent. Same as F2.

2009 · GRAMMARLY'S "3 ALTERNATIVE PHRASINGS"

Writing flexibility produced bad sentences.

Grammarly didn't fix it — it asked: "Did you mean...?"

→ Suggesting questions to ask of your own work. Same as F3.

2026 · SYNCIQ · THE PATTERN APPLIED TO AI

LLM outputs are the most flexible tool yet built.

SyncIQ is what spell-check, autocomplete, and Grammarly already proved you need.

Not a new pattern. The right pattern, applied to a new flexibility.

3 SOLUTIONS · 2 REJECTED · 1 WINS

SOLUTION 1 · TRUSTSCORE · REJECTED

Single confidence % per answer

"This answer is 87% confident."

✗ Brief explicitly bans "generic trust scores"
✗ One number can't capture per-claim variance
✗ 68% of users want per-claim flags, not aggregate scores (Q13)

SOLUTION 2 · SECOND-OPINION AI · REJECTED

Another AI verifies the first one

"Claude says ChatGPT's answer is..."

✗ Recreates the trust problem one level deeper
✗ Replaces user judgment instead of supporting it (brief pillar)
✗ "Who verifies the verifier?" — recursive trust target

SOLUTION 3 · SYNCIQ · ★ WINNER



Inline calibration layer

per-claim, in-flow, judgment-supporting

Three MVP features close 3 of 4 structural opacities. Fourth deferred to Phase 2.

✓ Per-claim granular (not aggregate) — passes the trust-score ban
✓ In-flow during the answer (not post-hoc) — solves the "decision moment"
✓ Supports judgment, doesn't replace it — passes the brief's pillar
✓ Builds calibration skill over repeated use — addresses what users WANT (68% Q13)

✓ **Perplexity-native by design** NEW — closes "why not ChatGPT?"

Citations are Perplexity's brand identity. SyncIQ extends that brand.
For ChatGPT, calibration is off-brand — they're selling decisiveness.

"Users don't fail to USE AI well. They fail at the decision moment — that 2-second window: 'Should I trust this, or check it?'"

Three solutions. Two banned by the brief. One survives — and rhymes with 40 years of HCI calibration design.

Three MVP features. One per opacity. Designed to ship in 8 weeks.

F1 leads adoption. F2 + F3 ship in the same release bundle. F4 + F5 deferred with named purposes.

THREE MVP FEATURES · EACH WITH BUILT-IN PREVIEW

F1 · MUST · MVP CENTERPIECE — RICE 62

Inline Confidence Marking

Per-claim ✓ ~ ? marks rendered inline with the answer.

Go offers strong ROI in tech. ✓
Instructor has 8 years experience. ?
The course is widely used. ~

CLOSES
Correctness Opacity

SERVES
Quick Truster — catches eye mid-skim
Careful Verifier — focuses verification

EVIDENCE
68% Q13 want granular flags · top user ask
4 of 5 interviews validate per-claim need

BUILD ~4 weeks · parallel to answer generation

F2 · MUST · MVP BUNDLE — RICE 21

Assumptions Card

AI surfaces what it inferred about you before answering.

SYNCIQ INFERRED ABOUT YOU
Goal: backend engineer (from past chats)
Status: working pro, not student
→ Wrong? Tap to correct.

CLOSES
Reasoning Opacity

SERVES
Quick Truster — reveals hidden framing
Careful Verifier — cuts re-prompt cycle

EVIDENCE
46% Q13 want assumptions surfaced
"AI should understand my context first."

BUILD ~3 weeks · uses M1 Intent Inferrer

F3 · MUST · MVP BUNDLE — RICE 18

Ask-Back Prompts

3 questions to ask of this output, generated by AI.

QUESTIONS YOU SHOULD ASK
1. What's the actual completion rate?
2. Are free Go resources adequate?
3. Does Go fit YOUR career path?

CLOSES
Calibration Opacity

SERVES
Quick Truster — seeds evaluation habit
Careful Verifier — prioritizes checks

EVIDENCE
46% Q13 want "3 expert Qs to ask"
Highest desirability flag in survey

BUILD ~2 weeks · grounded in low-conf claims

F1 IN ACTION · SAME QUERY, BEFORE vs AFTER SYNCIQ MARKS

WITHOUT SYNCIQ · Perplexity default

Q: Should I take the ₹40,000 Go course?

Go offers strong ROI in tech.
The course is widely used.
Instructor has 8 years experience.
Completion rate is high.

All claims rendered with identical polish. User can't tell which to question.

WITH SYNCIQ · F1 marks inline

Q: Should I take the ₹40,000 Go course?

Go offers strong ROI in tech. ✓
The course is widely used. ~
Instructor has 8 years experience. ?
Completion rate is high. ?

✓ high · ~ assumed · ? limited data. User sees which 2 claims to verify.

RICE · ALL 5 FEATURES SCORED

FEATURE	R·I·C·E	SCORE	DECISION
F1 · Inline	5·5·5·2	62	**Lead MVP**
F2.Assumption	4·4·4·3	21	Bundle
F3 · Ask-Back	3·4·3·2	18	Bundle
F4 · Don't Know	3·3·2·5	~4	**Phase 2**
F5 · Evaluation	—	~3	Phase 3

FEATURE-TO-OPACITY MAP · 3 OF 4 CLOSED AT MVP · 1 DEFERRED WITH NAMED PURPOSE

OPACITY 1 · CORRECTNESS
F1 · Inline Confidence Marking — Per-claim ✓ ~ ? marks make uncertainty visible — CLOSED at MVP

OPACITY 2 · REASONING
F2 · Assumptions Card — Surfaces inferred context before the answer — CLOSED at MVP

OPACITY 3 · CALIBRATION
F3 · Ask-Back Prompts — Builds evaluation skill through 3 grounded Qs — CLOSED at MVP

OPACITY 4 · COMPLETENESS
F4 · "What I Don't Know" — Needs user-profile system from MVP data first — DEFERRED Phase 2

Three features. Three opacities closed at MVP. One deferred — by design, not by accident.

F4 earns its place when MVP data proves users want it. Then F5 builds the long-term moat.

SyncIQ runs alongside Perplexity. Adds visibility, not latency.

6 system layers · 1 user flow · runs in parallel · <300ms P50 latency budget · zero blocking on Perplexity

6-LAYER STACK · CLIENT TO DATA

LAYER 1 · CLIENT · PERPLEXITY.AI BROWSER/MOBILE

Composer · SyncIQ overlay · Marker UI · Card UI · Ask-back chips

Renders ✓ ~ ? marks inline · Shows Assumptions Card above answer

Stack: React Native + Perplexity SDK · zero new auth, uses existing session

LAYER 2 · CLASSIFICATION · CLAIM PARSING + SCORING

Claim Parser · Confidence Scorer · Intent Inferrer · Ask-back Generator

Splits answer into claims · scores each per-claim · thresholds ≥0.85 ✓ / 0.5–0.85 ~ / <0.5 ?

Built on Claude Haiku 4.5 · sub-200ms per claim batch

LAYER 3 · ORCHESTRATION · COORDINATOR + SAFETY

Threshold Manager · Display Rules · Override Flags · Constitutional Filter

Caps marks at 3 per output · falls back to plain render if budget exceeded

Routes through safety filters · refuses to mark in sensitive topics

LAYER 4 · EXECUTION · PERPLEXITY'S EXISTING ENGINE

Perplexity LLM · Source Retrieval · Citation Generator

Unchanged · SyncIQ runs PARALLEL to answer generation, not sequential

If SyncIQ fails, Perplexity output renders as today — zero degradation

LAYER 5 · FEEDBACK · CORRECTION EVENTS + LEARNING

Correction Listener · Calibration Tracker · Assumption Updates

"Wrong? Tap to correct" feeds into Intent Inferrer · stores anon by default

Powers F4 (M3+) and F5 (M5+) — earned by MVP data

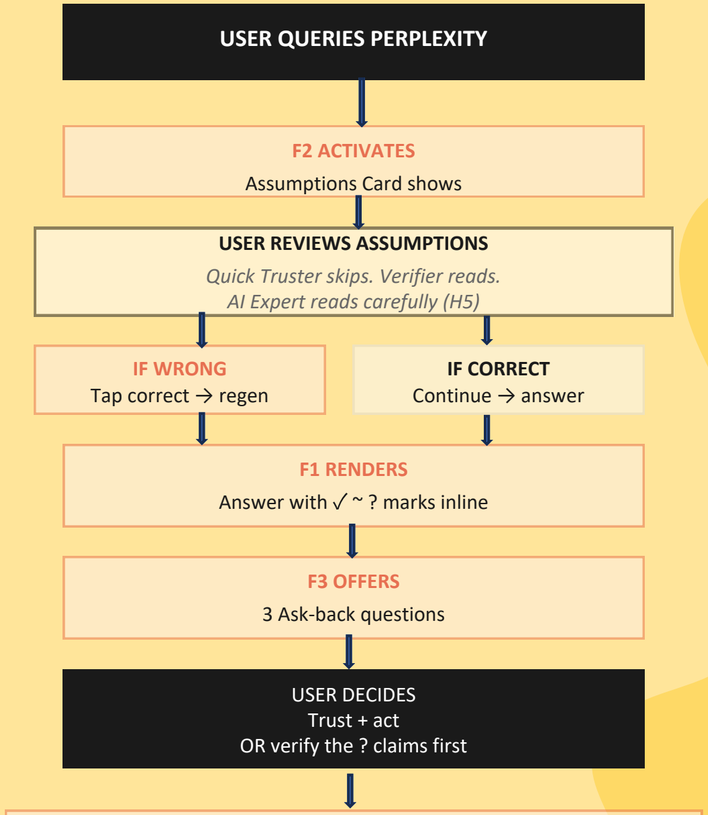
LAYER 6 · DATA · TELEMETRY, NO PII STORAGE

Anonymous event log · per-claim outcome telemetry

Stores: claim hash + mark + user action · NO query text, NO user identity

GDPR-aligned · monthly PII audit

USER FLOW · BRANCHING TREE



DECISION POINT · CALIBRATED ACTION

User retains all judgment. SyncIQ surfaces — never decides for the user.
Brief pillar: "Supports judgment, doesn't replace it."

PARALLEL SYSTEM PROCESSES

PARALLEL · WHAT HAPPENS BEHIND THE SCENES

- Layer 2 parses output into 5–8 claims · ~150ms · Claude Haiku batch scoring
- Layer 3 applies thresholds + display rules · ~30ms · max 3 marks/output, ≥0.65 confidence to flag
- Layer 1 renders inline marks + assumption card · ~80ms · React composition, no full re-render

PERFORMANCE BUDGET

Total: P50 < 300ms · P90 < 500ms
SyncIQ overlay appears AFTER answer renders, never blocks it

- Layer 4 (Perplexity) renders unmodified · If SyncIQ fails or exceeds budget, user sees normal Perplexity
- Layer 5 captures user actions for learning · "Marked ✓ but user verified anyway" → threshold tuning
- Layer 6 stores anonymized telemetry · No query text, no user identity — claim hash + mark + outcome

FALLBACK BEHAVIOR · 3 GUARANTEES

1. Layer 2 fails → Perplexity renders normally, no marks
2. Latency budget exceeded → marks render late, never blocking
3. User opts out → SyncIQ disabled per session, no telemetry

SCALABILITY

Per-claim parsing scales linearly. Tested: 8 claims/answer × 30M MAU = ~5B claim scores/day at peak.

SECURITY POSTURE · KEY GUARANTEES

- No PII in telemetry · monthly audit
- Constitutional filter on Layer 3
- 500-claim eval suite before threshold changes
- User opt-out preserves Perplexity baseline

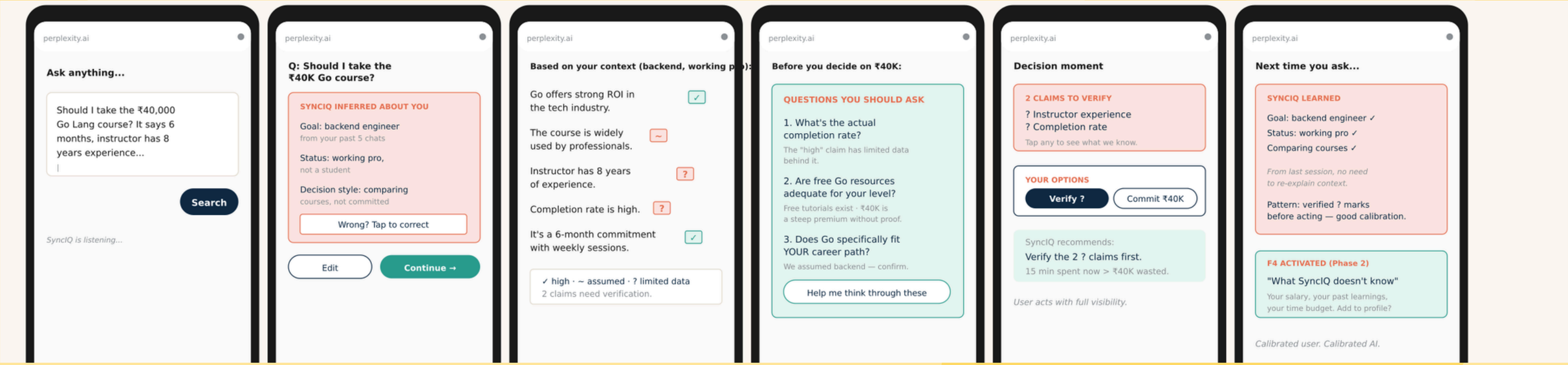
User flow drives 7 emotional shifts in <8 seconds — from Uncertainty to Calibrated. See interactive prototype for full progression.

6 layers running in parallel. 1 user flow. Sub-300ms P50. Zero blocking on Perplexity's answer.

Six screens. One full calibration journey, end to end.

Same query (₹40K Go course decision) — see how SyncIQ moves the user from default trust to calibrated action.

MOBILE PRODUCT FLOW · 6 SCREENS · PROTOTYPE → WIREFRAME →



SCREEN 1 · QUERY

User types question. Standard Perplexity composer. SyncIQ is invisible at this stage — listening to intent (Layer 2). No friction added. No new auth.

SCREEN 2 · F2 ASSUMPTIONS CARD

AI surfaces 3 inferences first. User sees what SyncIQ thinks they want. Tap to correct → cuts re-prompt cycle. Closes Reasoning Opacity. ~3 sec from query to this card.

SCREEN 3 · F1 INLINE MARKS

5 claims marked per-claim. 2 ✓ (high) · 1 ~ (assumed) · 2 ? (limited). User sees which to verify before acting. Closes Correctness Opacity. MVP centerpiece — RICE 62.

SCREEN 4 · F3 ASK-BACK PROMPTS

3 grounded questions appear. Each tied to a flagged claim (~ or ?). User can tap to deepen on any one. Closes Calibration Opacity. Builds eval skill over repeated use.

SCREEN 5 · DECISION MOMENT

User chooses with confidence. 2 ? claims surfaced for verification. Recommendation is suggestive, not forced. User retains all judgment. No automatic action.

SCREEN 6 · NEXT TIME (F4 PREVIEW)

Calibration compounds over time. Previous corrections feed Intent Inferrer. F4 surfaces what AI still doesn't know. Powered by Layer 5 + 6 data. Phase 2 capability.

Six screens. One full calibration journey. Every screen earns its place by closing a different opacity.

From default trust to calibrated action — in under 8 seconds, end to end.

We measure the behavior shift. We name the ways SyncIQ can fail.

NSM: % users who verify a flagged claim before acting. 5 leading + 5 lagging + 5 guardrails. 4 risks ranked P0–P3.

METRICS · NORTH STAR + LEADING + LAGGING + GUARDRAILS

NORTH STAR METRIC · TIME-PHASED PROGRESSION

% of users who verify a flagged claim before acting

Sessions with ≥1 ? or ~ claim tap and verification engagement ÷ total sessions with flags shown · measured weekly

TODAY	3 MONTHS	12 MONTHS	BENCHMARK	FAILURE
0%	35%	65%	Skill Nudge: 40%/6mo	< 25% by M3
no signal exists	*MVP traction*	*habit formed*	*Blueprint: 0→3.5/12mo*	*= re-evaluate*

LEADING INDICATORS · 5 EARLY SIGNALS THAT PREDICT NSM

1. Flag engagement rate	≥40% / M3	% users who hover/tap any ✓ ~ ? mark
2. Assumptions Card correction	≥15% / M2	% tap "wrong? correct" on F2 — proves F2 is read
3. Ask-back click-through	≥25% / M3	% tap any F3 question — proves want to interrogate
4. Time to first verification	<15 sec median	From answer render to first ? tap
5. Per-claim hover dwell	≥2 sec avg	Sub-2s = scanning, not evaluating

LAGGING INDICATORS · 5 BUSINESS OUTCOMES DOWNSTREAM OF NSM

1. Repeat use D30	target ≥55% / M3	% returning within 30 days
2. Perplexity Pro conversion lift	target +1.5pp / M6	SyncIQ users vs A/B control
3. NPS shift (calibrated users)	target +12 / M6	NSM-hitting vs control
4. Pro tier churn reduction	target -2pp / M6	SyncIQ users churn less
5. High-stakes query share	target +30% / M6	% queries SyncIQ classifies high-stakes

REVENUE LINE · PROJECTION

Project ~$8M ARR uplift by Y2 — modeled not benchmarked (SyncIQ is novel).

GUARDRAILS · 5 HARD LIMITS · ANY BREACH TRIGGERS FALLBACK

Latency P50/P90	False-flag rate	Dismissal rate	PII in telemetry	Baseline sat.
<300/<500ms	< 8%	< 15%	0%	no drop

RISKS · RANKED P0-P3 · PROBABILITY + TRIGGER + MITIGATION + OWNER

P0 SAFETY · HIGHEST PRIORITY — PROBABILITY: MEDIUM

Wrong ✓ marks erode trust in Perplexity itself

User blames Perplexity (not SyncIQ) when ✓ claim proves false. Each false ✓ damages parent brand.

TRIGGER: False-flag rate ≥8% in any 7-day window · or ≥2 viral X/Reddit complaints in same week

MITIGATION: ≥0.85 confidence threshold for ✓ · 500-claim eval suite · rollback flag · Owner: Safety Lead

P1 PRODUCT · HIGH PRIORITY — PROBABILITY: HIGH

Cognitive overload — too many flags kill the experience

3+ marks per output → users paralyzed, dismiss SyncIQ, revert to "trust polish" baseline.

TRIGGER: Dismissal rate ≥15% in any cohort · or NPS drop ≥5pts in SyncIQ-on group vs control

MITIGATION: Hard cap at 3 marks/output · collapsed-by-default UI · A/B test density · Owner: PM

P2 TECHNICAL · MEDIUM PRIORITY — PROBABILITY: MEDIUM

Latency spikes degrade per-claim parsing

Per-claim Haiku scoring exceeds 500ms budget. Marks render late or not at all.

TRIGGER + MITIGATION:

P90 >500ms 1hr → parallel batch scoring · per-region flag · auto-fallback · Eng Lead

P3 STRATEGIC · LONGER-TERM — PROBABILITY: LOW-MED

Behavior doesn't shift — knowing isn't acting

Users see ? marks, ignore them, act anyway. NSM stalls below 25% by M3.
Risk: marks suffer same fate as citations (48% currently skip).

WHY MARKS DIFFER FROM CITATIONS: Citations = work (click→read→return). Marks = zero-work (inline glance). Citations transfer trust externally; marks signal AI's own honesty.

THE MVP BAR · WHAT SUCCESS MEANS FOR SYNCIQ AT MONTH 3

BEHAVIOR ≥35% verify a flagged claim before acting (incl. H5 expert cohort) | **TRUST** False-flag <8% across eval suite | **BUSINESS** Pro retention +1.5pp target vs A/B control

5 leading · 5 lagging · 5 guardrails · 4 ranked risks. Each metric has a target. Each risk has a trigger and owner.

If 1 of 3 MVP Bar conditions misses by Month 3, we pause and re-evaluate. Honesty before scale.

SyncIQ doesn't add a new SKU. It protects the revenue Perplexity already has.

Bundled in Pro · 3-phase rollout · 3 distribution channels · $312M Y3 ARR projected via retention lift.

MARKET SIZING + 3Y ARR

TAM · AI ANSWER ENGINES

$47B

Global market 2026

38% CAGR · projected $200B+ by 2030.

SAM · PERPLEXITY USERS

30M

Perplexity MAU end-2025

10M Pro subscribers · 20M free tier

SOM · BEHAVIOR-SHIFT TARGET

6.5M

Pro users hitting NSM by M12

65% of 10M Pro = calibrated user base

3-YEAR ARR PROJECTION

Y1 2026 · MVP	$52M
Y2 2027 · full rollout	$158M
Y3 2028 · Enterprise	$312M

All revenue is RETAINED Pro revenue.
Zero new SKU. Bundled in Pro.

REVENUE MODEL · 4 PRICING TIERS

TIER	PRICE	SYNCIQ ACCESS	USER COUNT
Free **Acquisition tier**	$0	F1 free · F2/F3 capped 5/day Soft paywall on F4	20M users
Perplexity Pro **PRIMARY REVENUE TIER**	$20 / mo	F1 + F2 + F3 unlimited F4 + F5 included (Phase 2+)	10M users
Teams (Phase 3) **Multi-seat option**	$35 / user / mo	+ shared calibration audits Cross-team learning signal	500K (M12)
Enterprise (Phase 3) **Custom pricing**	Custom	SLA + audit logs + SSO Finance + Legal verticals	50 contracts

Projected SyncIQ retention lift: ~$8M ARR by Y2 *(modeled estimate · validated via A/B in Phase 1)*

LAUNCH PLAN · 3 PHASES OVER 6 MONTHS

PHASE 1 · M1-M2 · PRO USERS ONLY · 10M COHORT

Validate behavior shift with high-intent cohort

10M Pro users get F1 + F2 + F3 at MVP launch · A/B 50/50 control vs SyncIQ-on

Decision gate: NSM ≥ 25% in cohort by M2 before Phase 2 expansion

PHASE 2 · M3-M4 · FREE TIER + SOFT PAYWALL · 30M COHORT

Expand to all 20M free users · introduce F4 as Pro-only differentiator

F1 free for all · F2 + F3 free capped 5/day · F4 "What I Don't Know" Pro-only

Pro conversion lift target: +1.5pp by M6 · Reddit AMA at Week 12

PHASE 3 · M5-M6 · TEAMS + ENTERPRISE · 50 PILOTS

F5 Evaluation Memory · Teams tier · Enterprise outbound

F5 ships · per-user profiles · Teams $35/user/mo · shared calibration audits

Enterprise pilot with 5 finance + legal firms · M6 target: 50 Enterprise contracts

GTM CHANNELS · 3 DISTRIBUTION PATHS

CHANNEL 1 · IN-PRODUCT (PRIMARY · 70% reach)

SyncIQ activates inline at first eligible query

Zero acquisition cost · the product IS the GTM · expected reach 30M MAU in 90 days

Compass precedent: 5 steps, zero clicks to discover · same model for SyncIQ.

CHANNEL 2 · REDDIT + SOCIAL (20% reach)

Organic A/B results post at Week 8 · community-led validation

r/Perplexity 240K members · r/AI 4M · Twitter/X SyncIQ launch thread

Skill Nudge precedent: 747K r/ClaudeAI = strong validation channel

CHANNEL 3 · ENTERPRISE OUTBOUND (10% reach · Phase 3 only)

Direct outreach to 100 finance/legal firms

Audit log + SLA differentiator · partnership with Big 4 consulting firms

Phase 3 only · M5-M6 · 50 contracts target by M6

$47B TAM · 30M SAM · 6.5M SOM · 3 phases over 6 months · 3 channels · 4 pricing tiers.

$312M Y3 ARR projection — all retained Pro revenue. Bundled in Pro, not a new SKU.