

The FDE Customer Feedback Question Bank

43 questions to uncover the real problems in AI products: a field guide for Forward Deployed Engineers, AI product managers, and customer success teams.

How to Use This Question Bank

The purpose of this question bank is to move every conversation from opinion to evidence. A user saying "it's not accurate" gives an engineer nothing to work with. A user showing the exact query, the AI's response, and the answer they expected gives an engineer a reproducible test case.

When to use it

Use these questions at any point where you have direct access to a customer or end user:

- **Onboarding calls:** establish a baseline of how users work before the AI becomes part of their routine.
- **Quarterly business reviews:** check whether adoption is growing and which issues are blocking it.
- **Escalations:** when a customer raises a complaint, use the diagnostic sections to find the root cause quickly.
- **Post-release check-ins:** after shipping a fix, use Section 6 to confirm it worked and caught no regressions.

You do not need to ask all 43 questions in one session. A typical 30-minute call uses 8 to 12 questions picked from two or three sections.

Four interview rules

1. **Ask about specific recent tasks, not general opinions.** "How do you feel about the assistant?" invites a vague answer. "What did you use the assistant for yesterday?" invites a concrete story.
2. **Ask "show me" before "tell me."** Ask users to share their screen and repeat the task. You will often see friction they have stopped noticing, such as copying text between three windows.
3. **Keep asking follow-ups until a root cause emerges.** The first answer usually describes a symptom. The third or fourth follow-up usually reveals the cause.
4. **Record the evidence triplet.** For every reported problem, capture the exact user query, the AI's actual output, and the output the user expected. Without these three items, the issue cannot be reproduced.

Example: A user says, "The HR bot is useless." Instead of logging that comment, the FDE asks for a recent example. The user shares that they asked "How many days of parental leave do I get?" and the bot answered "12 weeks." The current policy says 16 weeks. That single example points straight to a stale document in the knowledge base.

Question type legend

Each question in this guide is tagged with one of four types:

Tag	Purpose	Typical moment
Discovery	Understand the user's context and workflow	Onboarding, first meetings
Diagnostic	Isolate the technical cause of a problem	Escalations, bug triage
Prioritization	Rank issues by business impact	Roadmap planning, QBRs
Validation	Confirm a fix worked without side effects	Post-release check-ins

Section 1: Workflow & Context Discovery

This section answers one question: where does the AI fit in the user's day? Many complaints that sound like AI problems are actually workflow problems, so start here before diagnosing anything.

Q1. Which task is the AI product helping you complete? (*Discovery*) This anchors the conversation in a real job rather than the product's feature list.

- *Example:* A sales rep says "writing follow-up emails after demos." You now know to evaluate the AI on email tone, CRM context, and speed, not on general knowledge.

Q2. Walk me through the last time you used it, step by step. (*Discovery*) Recent, specific memories are far more accurate than general impressions.

- Ask the user to share their screen and repeat the task live if possible.
- Note every pause, copy-paste, and window switch.

Q3. Which tools or applications do you switch between to finish that task?

(Discovery) Context switching is one of the most common hidden sources of frustration.

- *Example:* A support agent uses the AI assistant, then copies the answer into Zendesk, then checks the policy in SharePoint. The real fix may be a Zendesk integration, not a better model.

Q4. Who else is involved before or after the AI's output is used? *(Discovery)* AI output often passes through reviewers or approvers. Their requirements shape what "good" means.

- *Example:* A legal team member must approve every AI-drafted contract clause. If the AI never cites its source clause, the reviewer rejects it, and the drafter blames the AI.

Q5. How did you complete this task before the AI tool existed? *(Discovery)* The old process sets the user's baseline expectations. If the AI is slower or less reliable than the old method, adoption will stall.

Q6. How often do you use it: daily, weekly, or occasionally? *(Discovery)* Usage frequency tells you how much weight to give the feedback.

- Daily users spot patterns and edge cases.
- Occasional users reveal onboarding and discoverability problems.

What to listen for: repeated copy-paste, manual reformatting, and phrases like "then I just do it myself." These signal integration gaps rather than model quality issues.

Section 2: Output Quality & Accuracy

This section isolates wrong, vague, or outdated responses. The goal is to collect concrete examples that engineers can reproduce and test against.

Q7. How often does the AI give an incorrect or irrelevant response? *(Diagnostic)*

Ask for a rough ratio, such as "1 in 10 answers." This creates a baseline you can measure against after a fix.

Q8. Can you share an example where the answer was wrong? What did you expect? *(Diagnostic)* This is the most valuable question in the guide because it produces the evidence triplet: query, actual output, and expected output.

- *Example:* Query: "What's the refund window for annual plans?" AI answer: "14 days." Expected: "30 days, per the 2026 pricing policy."

Q9. Are the answers wrong, incomplete, or just too generic? *(Diagnostic)* Each failure type points to a different cause:

- **Wrong:** usually bad retrieval or stale data.
- **Incomplete:** often too little context passed to the model, or chunks cut mid-section.
- **Too generic:** the model is answering from general knowledge instead of company documents.

Q10. Does the AI ever reference outdated policies, data, or procedures?

(Diagnostic) Outdated answers are among the most common complaints in RAG systems. They usually mean old document versions are still indexed alongside new ones.

- *Example:* Both "Travel Policy 2024" and "Travel Policy 2026" exist in the index, and the retriever ranks the older one higher because it has more matching keywords.

Q11. Do you trust the sources the AI cites? Why or why not? *(Diagnostic)* Trust drives adoption. If users don't trust citations, they will verify every answer manually, which removes most of the time savings.

Q12. What types of questions does it handle well vs. poorly? *(Diagnostic)* This maps the boundary of the system's competence.

- *Example:* "It's great at product specs but terrible at pricing questions." This suggests the pricing documents are missing, poorly formatted (for example, tables in scanned PDFs), or not indexed.

Q13. Does the tone or format of the output match what you need? *(Diagnostic)* A factually correct answer can still fail if it's too long, too formal, or in the wrong structure.

- *Example:* A support team needs two-sentence replies they can paste into chat. The AI writes five paragraphs. The fix is a prompt template, not a model change.

What to listen for: dates, version numbers, and words like "old," "last year," or "not anymore." These point to data freshness issues.

Section 3: Friction & Workarounds

This section surfaces pain points users never report. People adapt quietly to broken tools, so their workarounds are often the clearest evidence of a problem.

Q14. What do you do when the AI can't solve your problem? (Diagnostic) The answer reveals the fallback path and how costly each failure is.

- *Example:* "I message Priya on Slack because she knows the policies." Every failed AI answer now costs two people's time.

Q15. How much do you edit AI-generated content before using it? (Diagnostic) Heavy editing is a strong signal that the output is close but not right.

- Ask what they change: facts, tone, length, or structure.
- *Example:* A marketing user rewrites every opening line because the AI always starts with "In today's fast-paced world." A simple prompt rule fixes it.

Q16. Have you rephrased the same question multiple times? What changed? (Diagnostic) Repeated rephrasing shows a gap between how users speak and how the system understands them.

- *Example:* A user asks "PTO policy" with no result, then "vacation days" with no result, then "annual leave" and gets the answer. The retriever needs synonym handling or a terminology glossary.

Q17. At what point do you give up and ask a human? (Diagnostic) This identifies the exact moment trust breaks. Map it to the type of query to find the weakest areas.

Q18. Is there any part of the workflow you skip or avoid? (Discovery) Avoided features are invisible in complaint logs but obvious in usage data.

- *Example:* Nobody uses the "summarize ticket history" button because it once produced a wrong summary in front of a customer. One bad experience can permanently kill a feature.

Q19. What's the most frustrating thing about using it today? (Prioritization) Ask this near the end of the section. After the specific questions, users give sharper and more honest answers.

What to listen for: "I just," "I usually," and "I don't bother." These phrases describe workarounds that have become habits.

Section 4: Integration & Data

This section uncovers system-level and data-source issues. These questions are where an FDE's technical background matters most, because users describe symptoms and you must connect them to pipelines, permissions, and infrastructure.

Q20. Which internal systems or documents should the AI know about but doesn't? *(Diagnostic)* Users know where the knowledge lives. Engineers often don't.

- *Example:* "The real troubleshooting steps are in the Confluence runbooks, not the public help center." The runbooks were never connected as a data source.

Q21. How frequently do your source documents or policies change? *(Diagnostic)* This determines how often the index must refresh.

- If policies change weekly and the index syncs monthly, users will see outdated answers for up to three weeks of every month.
- Ask who owns updates and whether old versions are archived or deleted.

Q22. Are there terms, acronyms, or jargon the AI misunderstands? *(Diagnostic)* Internal language is rarely in a model's training data.

- *Example:* At one company, "QBR" means quarterly business review. At another, it means a quality-based release. The AI needs a company glossary in its context to resolve these correctly.

Q23. Does the AI respect access levels? Does it show you things you shouldn't see, or hide things you should? *(Diagnostic)* Permission errors are high-severity issues.

- **Over-exposure:** a junior employee sees salary bands or board documents. This is a security incident.
- **Under-exposure:** a manager can't get answers from documents they are entitled to. This is a usability problem.
- Escalate any over-exposure report immediately rather than logging it as routine feedback.

Q24. How fast are responses? Does speed affect your work? *(Diagnostic)* Latency tolerance depends on the task.

- *Example:* A live-chat agent with a customer waiting can tolerate about 2 to 3 seconds. An analyst generating a weekly report can tolerate 30 seconds.

Q25. Have you seen errors, timeouts, or failed actions? *(Diagnostic)* Ask for screenshots and approximate timestamps. These let you match the user's report to specific entries in the system logs.

What to listen for: names of tools (SharePoint, Confluence, Salesforce, Google Drive), version numbers, and any mention of seeing data meant for someone else.

Section 5: Prioritization & Impact

This section ranks issues by business value. Not every complaint deserves engineering time, and these questions give product managers the evidence they need to decide what to fix first.

Q26. Which single improvement would make the biggest difference to your work? *(Prioritization)* Forcing a single choice cuts through long wish lists and reveals what actually matters.

Q27. How much time would that save per week? *(Prioritization)* Time estimates turn feedback into a business case.

- *Example:* "About 3 hours a week" across a 40-person support team equals 120 hours per week, or roughly three full-time employees.

Q28. What happens to your business if the AI gives a wrong answer here? *(Prioritization)* This measures severity, which matters more than frequency for some issues.

- *Example:* A wrong answer about office parking is an annoyance. A wrong answer about drug dosage, loan eligibility, or a compliance deadline creates legal or financial risk and must be fixed first.

Q29. How many people on your team face this same issue? *(Prioritization)* A problem reported by one vocal user may affect only that user. Confirm the scope before committing engineering effort.

Q30. Is this blocking adoption, or is it an annoyance? *(Prioritization)* Blockers stop people from using the product at all. Annoyances slow them down. Blockers almost always come first.

Q31. If we could fix only one thing this quarter, what should it be? *(Prioritization)* Ask this to the team lead or budget owner, not only to end users. Their answer often reflects strategic priorities that individual users don't see.

Turning answers into a priority score

Score each issue from 1 to 5 on four factors, then divide by effort:

$$\text{Priority} = \frac{\text{Impact} + \text{Frequency} + \text{Severity} + \text{Reach}}{\text{Effort}}$$

Issue	Impact	Frequency	Severity	Reach	Effort	Priority
Outdated refund policy answers	4	4	5	5	2	9.0
No Zendesk integration	5	5	2	4	4	4.0
Responses too long for chat	3	5	1	4	1	13.0

In this example, shortening responses ranks first because it is cheap to fix and affects everyone. The refund policy issue ranks second but should be fast-tracked anyway because of its high severity.

Section 6: Post-Change Validation

This section confirms that a fix worked and caught no regressions. Shipping a change is not the finish line. The feedback loop closes only when users confirm the problem is gone.

Q32. After the update, has the original problem gone away? (Validation) Return to the exact example from the original complaint and test it again with the user.

- *Example:* Re-ask "What's the refund window for annual plans?" and confirm the AI now answers "30 days" with the correct source.

Q33. Are there questions that worked before but don't now? (Validation) Fixes often have side effects. Changing retrieval filters to favor recent documents might hide older documents that are still valid.

- *Example:* After prioritizing 2026 documents, the AI stops finding the 2023 warranty terms that still apply to older products.

Q34. Do you edit the outputs less than before? (Validation) This is a direct measure of output quality from the user's point of view. Compare it against the answer to Q15.

Q35. Are you escalating to humans less often? *(Validation)* Compare against the answer to Q17. Fewer escalations mean the AI is handling more of the workload.

Q36. How confident are you in the answers now, on a scale of 1 to 10? *(Validation)* Track this score across releases. A rising trust score is one of the strongest predictors of continued adoption.

Q37. What should we look at next? *(Prioritization)* End every validation call by opening the next cycle. This keeps the feedback loop continuous rather than a one-time event.

Before-and-after comparison

Pair each validation question with a measurable metric so the improvement can be proven, not just reported:

Question	Metric to track	Example before	Example after
Q32	Answer accuracy on the original test set	62%	91%
Q34	Responses accepted without edits	35%	68%
Q35	Escalations to human agents per week	140	85
Q36	Average user confidence score	5.2	7.8

Bonus: Six Follow-Up Probes

These short probes work after any question in the guide. They are how a symptom becomes a root cause.

Q38. "Can you show me?" Converts a description into observable evidence.

Q39. "Why does that matter to you?" Reveals the business impact behind a complaint.

Q40. "What happened next?" Uncovers the downstream cost of a failure.

Q41. "How often does that happen?" Separates one-off incidents from patterns.

Q42. "What did you expect instead?" Defines the correct output for testing.

Q43. "Is there anything I didn't ask that I should have?" Catches issues outside your question list. Always end a call with this.

Example: probes in action

The dialogue below shows how three probes move a vague complaint to a precise engineering task.

- **User:** "The assistant is slow."
- **FDE:** "Can you show me?" *(the user runs a query, and it takes 4 seconds)*
- **FDE:** "Is it always this slow?"
- **User:** "No, only when I ask about contracts."
- **FDE:** "What happens next when it's slow?"
- **User:** "The customer on the call gets impatient, so I stop using it during calls."

Outcome: The issue is not general latency. Contract queries retrieve large PDF chunks, which inflates the context size. The engineering task is to re-chunk contract documents, with a target of under 2 seconds for live-call use.

Translating Answers into Engineering Tasks

A customer complaint describes the problem from the user's side. An FDE's job is to translate it into the system component that needs to change. Use this table as a starting hypothesis, then confirm it with logs and test queries.

What users say	Likely root cause	Investigate	Typical fix
"Answers are outdated"	Stale index or duplicate document versions	Indexing pipeline, sync frequency, archive rules	Increase sync frequency; de-index superseded versions

"Answers are vague"	Weak retrieval relevance	Chunking strategy, ranking, metadata filters	Re-chunk by section; add a reranker
"It doesn't understand our terms"	Missing domain context	Glossary, system prompt, query expansion	Add a company glossary to the prompt; add synonym mapping
"I keep switching apps"	Workflow integration gap	APIs, connectors, where output is used	Build a direct integration into the user's main tool
"I edit everything"	Tone or format mismatch	Prompt templates, output examples	Add style rules and few-shot examples
"It's too slow"	High latency	Model choice, caching, context size	Cache frequent queries; reduce context; use a smaller model
"I saw something I shouldn't"	Broken access control	Permission sync, document-level ACLs	Enforce permission filtering at retrieval time; escalate as a security issue
"It makes things up"	Hallucination with no grounding	Retrieved context, citation checks	Require citations; return "I don't know" when no source is found

Worked example: from complaint to fix

This example follows a single complaint through the full process.

1. **Complaint:** "The HR assistant keeps giving wrong leave policy answers."
2. **Evidence (Q8):** The user asked about parental leave. The AI answered 12 weeks. The current policy says 16 weeks.
3. **Investigation:** The FDE checks the retrieval logs. Both the 2024 and 2026 policy PDFs are indexed. The 2024 version ranks higher because it contains the phrase "parental leave" more often.

4. **Root cause:** Superseded documents are never removed from the index.
5. **Fix:** Add an "effective date" metadata field, filter out superseded versions, and set a weekly sync with the HR document library.
6. **Validation (Q32, Q33):** The parental leave answer is now correct. A 50-question HR test set improves from 62% to 91% accuracy, with no regressions on other topics.

CERTIFIED FORWARD DEPLOYED ENGINEER

LEARN DIRECTLY FROM INDUSTRY EXPERTS, AI PRACTITIONERS, AND ENGINEERING LEADERS WHO ARE BUILDING AND DEPLOYING SCALABLE AI DRIVEN SOLUTIONS FOR ENTERPRISES WORLDWIDE.



ABOUT GSDC CERTIFICATION



EBOOK

Extensive and exclusive Ebook created by world's experts to help you with understanding core concepts.



EBOOK

Extensive and exclusive Ebook created by world's experts to help you with understanding core concepts.



CREATED BY EXPERTS

GSDC certifications are created and authored by world's leading experts in the field.



LEARNING MATERIALS

Get access to learning materials such as videos, ebooks, templates, and practice exams, which will help you clear the certification exam.

LEARNING OBJECTIVE

- Validate your Forward Deployed Engineering expertise.
- Increase your earning potential with an in-demand credential.
- Get complimentary career assistance after certification.

Enroll now with the code **LEARN20** To avail **20%** discount

Enroll Now



www.gsdccouncil.org