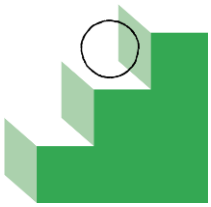
User Needs + Defining Success

Chapter worksheet



Instructions

Block out time to get as many cross-functional leads as possible together in a room to work through these exercises & checklists.

Exercises

1. Evidence of user need [multiple sessions]

Gather existing research and make a case for using AI to solve your user need.

2. Augmentation versus automation [multiple sessions]

Conduct user research to understand attitudes around automation versus augmentation.

3. Design your reward function [~1 hour]

Weigh the trade offs between precision and recall for the user experience.

4. Define success criteria [~1 hour]

Agree on how to measure if your feature is working or not, and consider the second order effects.

1. Evidence of user need

Before diving into whether or not to use AI, your team should gather user research detailing the problem you're trying to solve. The person in charge of user research should aggregate existing evidence for the team to reference in the subsequent exercises.

User research summary

List out the existing evidence you have supporting your user need. Add more rows as needed.

Date	Source	Summary of findings
10/14/2025	https://pathway.com/blog/ai-for-sec-filings	Anyone who has spent time poring over 10-Ks, 10-Qs, and 8-Ks knows the challenge. These documents often span hundreds of pages, filled with complex financial data, risk disclosures, and management discussions. Hidden within this sea of information are crucial insights that could impact investment decisions – subtle changes in risk factors, shifts in business strategy, or emerging competitive threats.
10/14/2025	https://www.captide.ai/insights/how-genai-transforms-sec-document-analysis	By analyzing filings across multiple companies simultaneously, GenAI can surface competitive dynamics that might not be obvious when looking at companies in isolation. This includes identifying similar strategic initiatives, overlapping market expansions, or contrasting approaches to industry challenges..
10/16/2025	https://search10k.com/how-to-use-ai-for-10-k-and-10-q-analysis/	According to a recent study by Microsoft, AI-powered analysis deliver between 45% to 82% accuracy in sentiment analysis. , improved accuracy in data extraction, and the ability to process multiple companies simultaneously for comparative analysis. These advantages are transforming how financial professionals approach due diligence and investment research.
10/17/2025	https://www.investopedia.com/articles/fundamental-analysis/09/form-10k.asp	Form 10-K is the most comprehensive compilation of information on a company. The Securities and Exchange Commission

		(SEC) requires it for all public companies. It is the best source of information on a company, providing—among other information—a description of the business and industry, risks, a summary of legal proceedings, and financial statements. It is both a quantitative and qualitative review.

Make a case for and against your AI feature

Meet as a team, look at the existing user research and evidence you have, and detail the user need you're trying to solve.

Next, write down a clear, focused statement of the user need and read through each of the statements below to identify if your user need is a potential good fit for an AI solution.

At the end of this exercise your team should be aligned on whether AI is a solution worth pursuing and why.

How might we solve **Financial Analysis of SEC 10-K Filings**?

Can AI solve this problem in a unique way?

AI probably better	AI probably **not** better
• AI can perform information retrieval faster • AI can detect cross company and multi year trends faster and generate some gaphs quickly • With proper training and deployment can process thousands of filings in minutes • Good with repetitive manual work such as KPI tagging and cross references • AI will augment the abilities of financial analysts	• Contextual understanding beyond text – Financial insight often depends on reading between the lines — something current AI models can't do. • Absence of Domain-Specific Judgment - Financial analysis isn't just pattern recognition; it's judgment. • AI trained solely on 10-K data doesn't access: Off-balance-sheet activities, earnings calls, investor presentations, or macro trends. • AI explanations for "why" a metric matters are not legally defensible in compliance contexts.

We think AI can help solve SEC 10-K financial analysis because these filings are long, complex, and unstructured, making it time-consuming for analysts to extract key insights manually. AI can rapidly retrieve, summarize, and compare information across multiple filings, allowing analysts to focus on interpretation rather than data gathering. Ultimately, AI amplifies analyst expertise , transforming 10-K review from a manual, document-heavy process into an efficient, insight-driven workflow that supports faster and more informed business decisions.

2. Augmentation versus automation

Conduct research to understand user attitudes

If your team has a hypothesis for why AI is a good fit for your user's need, conduct user research to further validate if AI is a good solution through the lens of automation or augmentation.

If your team is light on field research for the problem space you're working in, contextual inquiries can be a great method to understand opportunities for automation or augmentation.

Research protocol questions

- If you were helping to train a new coworker for a similar role, what would be the most important tasks you would teach them first?

  I Would start by teaching them What are the important metrics to look out for , and where they are normally located. Next , I would show them how to standardize and structure said data so they can be analyzed consistently across company and years.

- Tell me more about that action you just took, is that an action you repeat:
  This action would be repeated Yearly for every SEC 10-k Document , but a real financial analyst would to the same for each 10-Q filing too

- If you had a human assistant to work with on this task, what, if any, duties would you give them to carry out?

  I would delegate data Prep and validation to the assistant to get the filings, clean tables , and verify the values we have.

Below are some example questions you can ask to learn about how your users think about automation and augmentation.

If going to meet your users in context isn't feasible, you can also look into prototyping a selection of automation and augmentation solutions to understand initial user reactions.

The Triptech method is an early concept evaluation method that can be used to outline user requirements based on likes, dislikes, expectations, and concerns.

Research protocol questions

- Describe your first impression of this feature.

  It feels like a powerful assistant that could drastically reduce the time spent reviewing long SEC 10-K filings. The ability to ask natural-language questions and instantly retrieve relevant sections makes financial analysis faster, more consistent, and less tedious. It seems especially valuable for identifying insights and cross-company comparisons that would otherwise take hours to compile manually.

- How often do you encounter the following problem: "Spending excessive time manually searching through 10-K filings to find specific disclosures or financial metrics."

  Often , Analysts frequently need to cross-check figures or identify narrative trends, which involves scrolling through hundreds of pages of text for each company. This task repeats across multiple clients or coverage lists, making it a recurring challenge.

- How important is it to address this need or problem?

  Analysts frequently need to cross-check figures or identify narrative trends, which involves scrolling through hundreds of pages of text for each company. This task repeats across multiple clients or coverage lists, making it a recurring challenge.

3. Design your reward function

Once your team has had a chance to digest your recent research on user attitudes towards automation and augmentation, meet as a team to design your AI's **reward function**. You'll revisit this exercise as you continue to iterate on your feature and uncover new insights about how your AI performs.

Use the template below to list out instances of each reward function dimension.

Reward function template

Prediction

Reference		**Positive**	**Negative**
	Positive	True Positive {Example 1} {Example 2} {Example 3}	**False Negative** {Example 1} {Example 2} {Example 3}
	Negative	**False Positive** {Example 1} {Example 2} {Example 3}	True Negative {Example 1} {Example 2} {Example 3}

Take a look at the false positives and false negatives your team has identified.

- If your feature offers the most user benefit for **fewer false positives**, consider optimizing for **precision**.

- If your feature offers the most user benefit for **fewer false negatives**, consider optimizing for **recall**.

Our AI model will be optimized for: precision
because: Financial analysts depend on factual accuracy and trust in the extracted insights. Misinterpretations or hallucinated data from filings could lead to incorrect conclusions and financial risk. Prioritizing precision ensures every number, statement, and citation reflects the true content of the original SEC 10-K document.

We understand that the tradeoff for choosing this method means our model will:
sometimes retrieve fewer but higher-quality results. Analysts might miss some tangential or loosely relevant context (lower recall), but they can trust that what is returned is correct, verifiable, and grounded in the source documents, which aligns with the professional and regulatory standards of financial analysis.

4. Define success criteria

Now that you've done the work to understand whether AI is a good fit for your user need and identified the tradeoffs of your AI's reward function, it's time to meet as a team to define success criteria for your feature. Your team may come up with multiple metrics for success by the end of this exercise.

By the end of this exercise, everyone on the team should feel aligned on what success looks like for your feature, and how to alert the team if there is evidence that your feature is failing to meet the success criteria.

Success metrics framework

Start with this template and try a few different versions:

> If __**{ specific success metric }**__
>
> for __ **{ your team's specific AI driven feature }**__
>
> **{ drops below/goes above }**__ **{ meaningful threshold }**__
>
> we will __**{ take a specific action }**__.

Version 1

If the data consistency between extracted numeric metrics and official XBRL tags
for our quantitative data extraction pipeline
drops below 95% field match accuracy,
we will re-engineer our parsing layer and implement cross-verification using EDGAR's schema references.

Version 2

If the average query latency
for our RAG-based financial question-answering feature
goes above 5 seconds per user query,
we will optimize the vector database configuration and enable response caching for frequently queried filings.

Version 3

If the user satisfaction score (measured via feedback or post-analysis survey)
for our AI-driven 10-K analysis tool
drops below 4.0/5,
we will conduct user interviews and refine output explanations for better interpretability and trust.

Statement iteration

Take each version through this checklist:

- ☐ Is this metric meaningful for all of our users?
  - ☐ How might this metric negatively impact some of our users?
- ☐ Is this what success means for our feature on day 1?
  - ☐ What about day 1,000?

Final version

- **Meaningful for all users?**
  Yes — both analysts and executives rely on accurate extraction; errors directly affect trust.
- **Potential negative impact?**
  High accuracy focus might slow deployment of new features or reject borderline-valuable results due to rigid validation thresholds.
- **Day 1 meaning of success:**
  Achieving 90%+ accuracy validates that the extraction logic works and users can trust early outputs.
- **Day 1,000 meaning:**
  Accuracy remains essential, but **interpretability, explainability, and adaptability** (e.g., handling new filing formats) may matter more than raw accuracy.

Schedule regular reviews

Once you've agreed upon your success metric(s), put time on the calendar to hold your team accountable to regularly evaluate whether your feature is progressing towards and meeting your defined criteria.

Success metric review

Date:

Attendees: