People + AI Guidebook
pair.withgoogle.com/guidebook

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 / Existing Evidence
Oct 10, 2025	User Feedback Surveys & App Reviews	63% of surveyed Goodreads users mentioned that recommendations are repetitive or too focused on bestsellers. Users expressed a need for more diverse and personalized suggestions aligned with their niche interests (e.g., indie authors, specific subgenres).
Oct 12, 2025	Platform Usage Data (Engagement Metrics)	Clickstream and reading history analysis revealed that over 70% of user interactions are with top 10% of most popular titles, indicating a strong popularity bias in the current recommendation model. Lesser-known books have minimal visibility.
Oct 14, 2025	Community Forums & Goodreads Groups	Qualitative analysis of user discussions shows dissatisfaction with the lack of adaptability — readers report the system doesn't evolve with their changing reading preferences or moods over time.
Oct 16, 2025	Academic Literature Review	Prior research on recommender systems (e.g., "Bias in Collaborative Filtering," ACM 2023) shows that traditional models reinforce mainstream preferences and underrepresent minority or niche content, validating similar issues observed on Goodreads.
Oct 18, 2025	Competitor Analysis (Amazon, StoryGraph)	Competing platforms integrate AI-driven hybrid recommenders combining user embeddings, NLP-based text similarity, and demographic/contextual signals to improve recommendation diversity and accuracy. These serve as benchmarks for potential improvements.
Oct 20, 2025	Internal Data Audit	Internal review identified uneven distribution in training data — high-frequency users dominate the dataset, while occasional readers are underrepresented, leading to potential personalization imbalance.
Oct 25, 2025	Support Tickets / Customer Queries	Several recurring complaints mention recommendations not reflecting recent reads or ratings, showing a lag in model updates and limited responsiveness to new user behaviour.

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 *deliver book recommendations that are dynamically personalized, context-aware, and equitable, allowing users to discover books matching both their current interests and long term reading tastes?*

Can AI solve this problem in a unique way?
Yes — AI can uniquely address the challenge by:

- *Predicting future user interests from reading, rating, and interaction history.*
- *Detecting and mitigating biases toward popular books and mainstream authors.*
- *Leveraging NLP to analyse book descriptions, reviews, and user-generated content for subtle patterns in preferences.*
- *Adapting recommendations over time to reflect changing user tastes, moods, or contextual signals.*
- *Generating serendipitous and diverse suggestions that traditional static or rule-based systems cannot achieve.*

AI Probably Better	AI Probably Not Better
✅ The core experience requires recommending different content to different users, which is inherently personalized.	❌ If recommendations must remain static or predictable (e.g., a curated list of fixed categories), AI adds unnecessary complexity.
✅ The system needs to predict future user interests based on reading and rating history, which aligns well with AI's predictive capabilities.	❌ If users only need a simple genre-based filtering (e.g., "Show me mystery books"), rule-based filtering may suffice.

✅ Personalization will improve satisfaction and engagement, as users have diverse tastes and reading patterns.	❌ If all users prefer identical content lists (unlikely for Goodreads), AI personalization isn't necessary.
✅ AI can detect subtle, evolving trends in user behaviour such as shifting interest from thrillers to sci-fi are better than hardcoded logic.	❌ High transparency requirements (if users demand to know exactly why each book was recommended) might challenge AI explainability.
✅ AI enables bias detection and mitigation, identifying unfair exposure toward gender, popular books and rebalancing visibility	❌ The cost of algorithmic errors could be high if recommendations become irrelevant or promote inappropriate content.
✅ NLP-based AI can analyse book descriptions, reviews, and user-generated text to better capture content similarity.	❌ If speed to market is the only priority, a basic collaborative filter may be faster to deploy.

We think **AI *can* help solve** the user need of delivering **fair, adaptive, and personalized book recommendations** on Goodreads, because:

1. AI can **learn from user behaviour patterns** and predict future interests, enabling dynamic and evolving recommendations.
2. AI models can **detect and mitigate bias**, ensuring equitable visibility across popular and niche titles.
3. By combining **collaborative filtering, NLP, and fairness-aware algorithms**, AI can generate recommendations that feel relevant. Personalization powered by AI will **increase user satisfaction, discovery diversity, and platform engagement**.

While AI introduces challenges in transparency and development complexity, its potential to transform the recommendation experience through personalization and fairness makes it a **strong fit for addressing this user need**.

To maintain user trust in AI-driven recommendations, we will surface lightweight explanations such as "Recommended because you liked X" or "Trending in your favourite genre."
These explanations help users understand the reasoning behind suggestions and increase confidence in AI 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.

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

Research protocol questions

Research Methods

1. 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?
    - **Interviews or Surveys:** Ask users about their reading habits, how they pick books, and how they'd like AI to help.
    - **Prototype Testing:** Show users early AI features (like "mood-based recommendations" or "discover hidden gems") and observe reactions.
    - **Contextual Inquiry:** Observe users navigating Goodreads, seeing where they struggle or spend the most time.

Sample Research Questions

1. Understanding User Tasks

- If you were helping a friend find books you like, what steps would you teach them?
- Which of these actions do you repeat often when choosing books?
    - Hourly / Daily / Weekly / Monthly / Rarely

2. Human Assistant Perspective

- If you had a human assistant helping with book selection, what would you ask them to do?
  Recommend books similar to recent reads, highlight hidden gems, sort by subgenre

3. Reactions to AI Features

- Describe your first impression of this feature.

- How often do you face the problem of generic or repetitive recommendations?
  - Daily / Often / Sometimes / Rarely / Never
- How important is it to address this problem?”
  - Not at all / Somewhat / Moderately / Very / Extremely

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 helpful and tailored to my reading preferences. I like that it shows books I wouldn’t have discovered otherwise.
I’m not sure how accurate it will be at predicting my tastes, but it’s interesting to see suggestions based on my recent reads.

- **How often do you encounter the following problem: generic or repetitive recommendations?**

☐ Daily
☑ Often (a few times a week)
☐ Sometimes (a few times a month)
☐ Rarely (a few times a year)
☐ Never

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

☐ Not at all important
☐ Somewhat important
☐ Moderately important
☑ Very important
☐ Extremely important

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

Dimension	Examples for Goodreads Recommendation System
True Positive (TP)	• User clicks on a recommended book and adds it to their "to-read" shelf. • User reads a recommended book and gives it a high rating. • User reviews a recommended book positively. • User finishes a recommended book within 1 month.
True Negative (TN)	• AI does not recommend books the user has already marked as "disliked." • AI avoids recommending books outside the user's preferred genres or subgenres. • AI does not show books the user has explicitly skipped multiple times.
False Positive (FP)	• User ignores a recommended book. • User clicks on a recommendation but quickly removes it from "to-read" shelf. • User gives a low rating to a recommended book.- User flags a recommended book as irrelevant or inappropriate.
False Negative (FN)	• AI fails to recommend a book that aligns with the user's taste based on past reads. • User discovers a book they would have loved, but it was not suggested. • Niche or indie book that fits user interest is never surfaced. • Highly rated book by similar users is missing from the recommendations.

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:

A balance of precision and recall

Because: Users benefit most when the system **reliably recommends relevant books while still surfacing a diverse range of titles**, including hidden gems. By balancing precision and recall, the AI ensures that most suggestions are enjoyable while giving users opportunities to discover new authors, genres, or niche content, maximizing satisfaction and engagement.

We understand that the trade-off for choosing this method means our model will:
occasionally recommend a few less relevant books (false positives) and may still miss a small number of potentially interesting books (false negatives). This trade-off allows users to receive a mix of highly relevant recommendations along with some exploratory suggestions, supporting both trust in the system and ongoing discovery.

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 trade-offs 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:

Version 1

If the click-through rate on recommended books for the AI recommendation feature drops below 15%, we will review and adjust the recommendation model, retraining with updated user data and tweaking personalization parameters.

Version 2

If the average user rating of recommended books for the AI recommendation feature falls below 3.5 stars, we will audit recommendation quality, analyse false positives, and implement content-based filtering adjustments.

Version 3

If the percentage of users engaging with hidden gem / niche book recommendations falls below 10%, we will reevaluate recommendation diversity and adjust the balance between popular and niche books in the algorithm.

Version 4

If the fairness equity index across author gender or popularity groups drops more than 10% compared to baseline, the team will pause deployment, retrain the model with adjusted bias-mitigation strength (λ), and run targeted A/B fairness checks.

Version 5

If more than 0.5% of users receive an empty recommendation list in a given week, we will trigger the hybrid fallback recommender and prompt onboarding preference collection to ensure continuous coverage for new users.

Statement iteration

Take each version through this checklist:

✅ Is this metric meaningful for all of our users?
Yes, each metric reflects engagement, satisfaction, or discovery—core goals of the feature.

✅ How might this metric negatively impact some of our users?
Over-optimizing for one metric (e.g., click-through) may reduce discovery of niche books. Need balance.

✅ Is this what success means for our feature on day 1?
Day 1 success: Early engagement and relevance are key; initial thresholds may be lower.

✅ What about day 1,000?
Day 1,000 success: Sustained engagement, diversity, and satisfaction; thresholds may be tightened as AI improves.

Over time, success will also include the system's ability to self-adapt through online learning — automatically re-weighting features and incorporating new feedback signals from user behaviour, reviews, and engagement data.

Final version

1. If **click-through rate on recommended books** drops below **15%**, we will **analyze recommendation relevance and retrain the model with updated data**.
2. If **average user rating of recommended books** falls below **3.5 stars**, we will **audit recommendation quality and refine personalization strategies**.
3. If **engagement with niche or hidden gem recommendations** falls below **10%**, we will **adjust the diversity balance to ensure users continue discovering new content**.

Operational Monitoring & Feedback Loops:
Model performance and success metrics will be tracked through the goodreads_recommendation_pipeline in Airflow.
Logs and metrics (CTR, fairness index, latency) will be stored in BigQuery and visualized via dashboards.
Quarterly user surveys and qualitative interviews will complement quantitative metrics, ensuring that "success" reflects real reader satisfaction.

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: Every alternate day in a week
Attendees: Ananya Asthana, Arpita Wagulde, Karan Goyal, Purva Agarwal, Shivam Sah, Shivani Sharma
Purpose: Evaluate progress, identify failing metrics, decide on interventions, iterate on AI model