Report

Gather inputs Begin with qualitative and quantitative research:

User interviews and support requests – capture the jobs-to-be-done, terminology, and data sources users rely on today. Prompt sampling – list direct asks (e.g., "show my Jira board") and indirect intents ("what am I blocked on for the launch?") that should route to your app. System constraints – note any compliance requirements, offline data, or rate limits that will influence tool design later. Document the user persona, the context they are in when they reach for ChatGPT, and what success looks like in a single sentence for each scenario.

Define evaluation prompts Decision boundary tuning is easier when you have a golden set to iterate against. For each use case:

Author at least five direct prompts that explicitly reference your data, product name, or verbs you expect the user to say. Draft five indirect prompts where the user states a goal but not the tool ("I need to keep our launch tasks organized"). Add negative prompts that should not trigger your app so you can measure precision. Use these prompts later in Optimize metadata to hill-climb on recall and precision without overfitting to a single request.