

Step 15: Creating a Rigorous Evaluation Protocol

Eval Protocol:

Description of the product:

WorkMate is an AI-driven emotional support and HR assistant designed to integrate seamlessly into communication tools like Slack and Teams. It offers employees a discreet, judgment-free space to access emotional support, document workplace concerns, and manage their work-life balance. The product includes personalized assistance through chat-based interactions, access to company resources on stress and emotional support, and integration with the user's daily planner. By using conversational AI and sentiment analysis, WorkMate aims to create a supportive work environment and help employees manage stress and productivity challenges.

The goals of our evaluation:

- We would like to assess the usability, efficiency, and user satisfaction of WorkMate, particularly in how well it helps users navigate workplace stress and manage HR-related concerns. Specifically, we aim to acquire knowledge regarding:
 - ease of completing key tasks within the WorkMate prototype
 - frequency with which participants successfully complete these tasks
 - time taken by participants to complete specific tasks
 - participants' subjective user experience and satisfaction with WorkMate
- Key Measures:
 - System Usability Scale (1-5): Measures general user satisfaction and usability
 - Task Completion Time: Measures how long participants take to complete tasks
 - Success Rate: Tracks the number of tasks participants successfully completed
 - Error Rate: Tracks the number of mistakes made during interactions
 - Error Type: Tracks the type of mistakes made during interactions
 - Subjective feedback from participants regarding their experience, difficulties, and suggestions for improvement

Protocol Script:

- Session Duration: approximately 1 hour
- Introduction: Overview of the session and informed consent
 - "Hi, I'm [Researcher Name] and I'm currently working on a new tool called WorkMate. Thank you for coming in today.
 - The purpose of this session is to get your feedback on an early version of the WorkMate tool. This is not a test of your abilities, but rather a way for us to understand what works well and what could be improved in the design. We want to make WorkMate as helpful and easy to use as possible.
 - The session will involve a few parts: First, I'll give you a brief overview and go over this consent form. Then, I'll ask you a few questions about your background. After that, I'll ask you to perform a few tasks using the WorkMate prototype. Finally, we'll

finish with some questions about your experience and gather any additional thoughts you have. The whole session should take about one hour.

- We will be audio and video recording the session, and also capturing your screen activity. This is solely for our research purposes to help us remember your feedback and how you interacted with the prototype. All the data we collect will be kept confidential and anonymized in any reporting.
- Your participation is voluntary, and you can stop the session at any time for any reason. You are also free to skip any tasks or questions you are not comfortable with. Your decision to participate or not will not affect your relationship with [mention relevant entity, e.g., employer if applicable].
- Please review this consent form. It explains the study in more detail and your rights as a participant. Feel free to ask any questions you have before signing.”
- Interview to collect demographic and background data about the participant
 - “Before we start with the prototype, I’d like to ask you a few background questions. To help us understand who is using WorkMate, could you please tell me:”
 - Age
 - Gender (optional)
 - Current occupation or field of work
 - Do you currently use any tools or applications for workplace communication, scheduling, or well-being support? If so, which ones and how often?
 - How familiar are you with chat-based virtual assistants or digital HR support platforms?
 - How comfortable are you generally with using new technology, on a scale of 1 to 5, where 1 is not at all comfortable and 5 is extremely comfortable?
- Participant completes tasks with the prototype
 - Introduction:

“Now we’ll move on to interacting with the WorkMate prototype. Please think aloud as you are performing the tasks – tell us what you are seeing, thinking, and feeling. This is very helpful for us to understand your process. Don’t worry about making mistakes; this is a prototype, and we are testing it, not you.”
 - Training:

“Before you start the tasks, let me just quickly show you the basic layout of the WorkMate interface...”

 - We will provide a brief, neutral overview without explaining how to complete the specific tasks or interact with the specific features.
 - Tasks:

“Here is your first task.”

 - We will present tasks one by one. Observe participant interaction, listen to think-aloud, and take notes. Avoid providing assistance unless absolutely necessary, and take note when assistance is given.
 - Task 1: Emotional Check-in

“Imagine you’ve had a frustrating meeting. Use WorkMate to check in with how you’re feeling.”

- Task 2: Find Company Resource
“You heard about a free yoga session offered by the company. Use WorkMate to find information about company resources and locate the yoga session.”
 - Task 3: Reserve a Session
“The yoga session on April 10th looks interesting. Use WorkMate to reserve a spot in that session.”
 - Task 4: View Updated Schedule
“After reserving the yoga session, check your schedule in WorkMate to see how it has been updated.”
 - Task 5: Explore HR Support
“You have a workplace concern you’d like to document. Explore the HR support features in WorkMate.”
- Interview to collect subjective user experience data
 - “Thank you for completing those tasks. Now I’d like to ask you about your overall experience using the WorkMate prototype:
 - What were your initial impressions of WorkMate?
 - Did you feel comfortable using WorkMate to complete your tasks? Why or why not?
 - What did you find most easy to use or understand about WorkMate?
 - What did you find most difficult or confusing about WorkMate? Can you tell me more about why?
 - How did you feel while using WorkMate, particularly during [mention a specific task or interaction observed based on note]?
 - On a scale of 1 to 5, how easy was it to complete the tasks today?
 - Based on your experience today, how likely are you to use a tool like WorkMate in the future, if it were available to you?
 - Do you have any suggestions for how WorkMate could be improved?
 - Do you have any other comments or questions?”
- Conclusion: Thanks and farewell
 - “That brings us to the end of the session. Thank you again for your time and valuable feedback. Your input is very important to us as we develop WorkMate.
 - As a token of our appreciation, here is your [compensation].
 - Do you have any questions for me? Or if you have any further thoughts or questions after the session, feel free to reach out. Have a great day!”

Data Collection Method:

Data to Collect:

- Quantitative Data:
 - System Usability Scale (1-5): Measures general user satisfaction and usability
 - Task Completion Time: Measures how long participants take to complete tasks
 - Success Rate: Tracks the number of tasks participants successfully completed
 - Error Rate: Tracks the number of mistakes made during interactions
- Qualitative Data:
 - Error Type: Categorizes and describes the types of errors observed during interactions
 - Researcher Notes: Handwritten or typed notes taken during the session, capturing unexpected behaviors, usability issues, and observations not fully evident in recordings.
 - Video and Audio Recording: Captures participants' interactions with WorkMate, including think-aloud comments and emotional reactions, and responses to open-ended questions after the session.
 - Screen Recording: Documents navigation paths, clicks, inputs, and other actions to trace participant interaction patterns and identify specific pain points.

How to Collect Data:

- Quantitative Data:
 - System Usability Scale (1-5): After completing all tasks, participants will rate how easy it was to complete the tasks, where 1 means "Very Difficult" and 5 means "Very Easy"
 - Task Completion Time: Use a stopwatch to record the time from the moment a task is introduced to when the participant indicates completion. Log times separately for each task
 - Success Rate: Record whether each task was completed successfully (yes/no)
 - Success criteria for each task:
 - Task 1: Emotional Check-in
Participant completes an emotional check-in by selecting or entering a feeling through WorkMate's interface and submits it successfully.
 - Task 2: Find Company Resource
Participant navigates to the company resources section and locates information about the yoga session without external help.
 - Task 3: Reserve a Session
Participants successfully reserves a spot for the yoga session on April 10th through WorkMate's interface, with confirmation shown.
 - Task 4: View Updated Schedule
Participants access their personal schedule in WorkMate and confirm that the yoga session appears in the updated schedule.

- Task 5: Explore HR Support
 - Participant finds and enters the HR support section in WorkMate and initiates or drafts a workplace concern report, then submits it.
 - Error Rate: During each task, count the number of observable errors, e.g., wrong clicks, failed attempts, misinterpretations
 - An error is recorded every time a participant deviates from the optimal or expected behavior.
- Qualitative Data:
 - Error Type: Classify each observed error into categories such as navigation error, input error, and misunderstanding
 - Researcher Notes: Take real-time handwritten or typed notes during the session, capturing participants' behavior, comments, hesitations, emotional responses, and anything unusual that recordings may not fully capture
 - Video and Audio Recording: Use a camera to record both video and audio during the task session; for the post-task interview, use an audio recorder to capture participants' responses
 - Screen Recording: Use screen capture software to record clicks, navigation paths, text inputs, error messages, and any interface elements participants interact with