AgBizLogic User Testing Findings

1. Overview

This document consolidates all qualitative and quantitative findings from Spring Term user testing for the AgBizLogic AI/UX Capstone Project. Our objective was to evaluate user interaction with the existing AgBizLogic web app and a prototype AI chatbot interface powered by three anonymized LLMs: ChatGPT, Claude, and DeepSeek. The dual goals were to identify usability challenges in the current interface and determine the most helpful, trustworthy AI responses.

2. Methodology

User testing occurred in two phases. In the first phase, participants clicked through the AgBizLogic website and described their experience while pointing out confusing elements and suggesting usability improvements. In the second phase, participants were shown Figma mockups of a chatbot interface. They asked domain-relevant questions, which were routed to the three LLMs. Responses were anonymized and randomized. Participants then rated each response on a scale from 1 to 10 and provided written qualitative feedback.

Metrics captured included numerical scores for each response, participant confidence levels, comments on tone and structure, and preferences regarding the inclusion of follow-up questions or visual elements.

3. Quantitative Summary

Results from the User Testing Analysis spreadsheet show varying preferences across participants. In one set of tests, DeepSeek received the highest average score of 9.33, followed by ChatGPT at 9.0 and Claude at 8.33. Another participant rated Claude and ChatGPT comparably, while still others showed more favorable views toward responses that included scientific precision, clear formatting, or real-world relevance.

The overall response averages across participants were:

Model	**Average Score**
DeepSeek	8.00
ChatGPT	7.11
Claude	6.56

These results suggest a slight preference for DeepSeek, although specific questions and domains influenced individual preferences considerably.

4. Qualitative Insights

Participants consistently favored answers that were structured, skimmable, and detailed without being overwhelming. Responses that included bullet points, subheadings, or logical groupings were especially well received. They appreciated when answers were contextualized to reflect agricultural or financial perspectives.

Skepticism often arose when responses included vague language or filler content. Participants were more likely to trust answers that included numerical estimates, citations, or suggestions grounded in domain knowledge. For instance, content referencing Brix levels for apples or spacing calculations for tree planting were cited as more credible and useful.

The design of the chatbot interface also influenced perceptions. Follow-up prompts such as "Would you like help calculating that?" were received positively. Participants noted that conversational tone and humility helped boost trustworthiness, whereas overly polished or marketing-heavy language sometimes reduced it.

5. Observational Themes

Some participants highlighted the value of practical application and numerical clarity. Others emphasized scientific accuracy and conciseness, particularly favoring structured formats over verbose explanations. Responses that mimicked human empathy while maintaining precision were preferred. Users also indicated a desire for clean sectioning when dealing with financial topics, along with relatable examples like budgeting for a home or planning for retirement.

6. Key Recommendations

Future chatbot outputs should integrate structured and skimmable formats while incorporating follow-up suggestions to simulate a natural conversation. Including numeric or visual elements, such as calculators or diagrams, will further assist comprehension. Whenever possible, responses should cite sources or reference data to enhance user trust. Additionally, refining the chatbot's tone to balance warmth, clarity, and modesty is essential. Language that feels artificial or promotional should be avoided.

7. Conclusion

This testing round validated strong user interest in both AgBizLogic's core tools and its potential for AI-powered assistance. However, improvements are needed in interface clarity and response trustworthiness. Future design iterations should focus on simplicity, precision, and credibility to better serve a diverse grower audience.