Milestone 4 - Usability Testing

ARIA: an AI Based Requirements Prioritization Service

The Crackers Team

Survey Structure

First of all, we constructed a comprehensive survey to cover all parts of user interactions with our application following best practices. To achieve that we divided the survey into parts with the structure as follows:

- Part 1: ORID Structured Questions
  - Objective
  - Reflective
    - Gibson's Affordances
    - Norman's Affordances
  - Decisional
- Part 2: ARIA-Specific Usability Assessment
  - SUS calculation
- Part 3: SEQ
  - Task Difficulty

Survey

Part 1: ORID Structured Questions

Objective

1. Which features did you successfully use?

- [ ] File upload (CSV/Excel)
- [ ] Manual requirement entry
- [ ] AI prioritization analysis
- [ ] View results/charts
- [ ] Export functionality
- [ ] Other: ____

2. For each feature, indicate whether you discovered it through exploration or it was immediately clear:

Feature	Discovered by Exploring	Immediately Clear	Did Not Use
File upload button	☐	☐	☐
Manual requirement entry option	☐	☐	☐
Requirement input fields (title, description)	☐	☐	☐
Scoring criteria (Business Value, Cost, Risk, etc.)	☐	☐	☐
Navigation between screens	☐	☐	☐
AI prioritization analyze button	☐	☐	☐
Priority results display	☐	☐	☐
Charts and visualizations	☐	☐	☐
Export to CSV option	☐	☐	☐
Export to HTML option	☐	☐	☐
Requirement ranking list	☐	☐	☐
Priority scores interpretation	☐	☐	☐

Reflective

3. Rate your emotional response during different stages (1=Very Negative, 5=Very Positive):

Stage	1	2	3	4	5
First impression when opening ARIA	☐	☐	☐	☐	☐
While inputting requirements data	☐	☐	☐	☐	☐
Waiting for AI prioritization results	☐	☐	☐	☐	☐
Viewing and interpreting results	☐	☐	☐	☐	☐

4. How well does ARIA solve the problem of requirements prioritization compared to manual methods?

Aspect	Much Worse	Worse	Same	Better	Much Better
Speed/Efficiency	☐	☐	☐	☐	☐
Objectivity/Fairness	☐	☐	☐	☐	☐

Aspect	Much Worse	Worse	Same	Better	Much Better
Insight Quality	☐	☐	☐	☐	☐
Confidence in Results	☐	☐	☐	☐	☐

5. Did the system behave as you expected?

- ☐ Always
- ☐ Usually
- ☐ Sometimes
- ☐ Rarely
- ☐ Never

6. Rate your agreement with the following statements (1=Strongly Disagree, 5=Strongly Agree):

Gibson's Affordances:

Statement	1	2	3	4	5
I had to experiment and explore to discover features	☐	☐	☐	☐	☐
I learned by trial and error what actions were possible	☐	☐	☐	☐	☐
I discovered useful features by accident through exploration	☐	☐	☐	☐	☐
The system encouraged me to try different interactions	☐	☐	☐	☐	☐

Norman's Affordances:

Statement	1	2	3	4	5
The interface clearly indicated what actions were possible	☐	☐	☐	☐	☐
Visual cues guided me on how to use features correctly	☐	☐	☐	☐	☐
Button labels and icons immediately communicated their purpose	☐	☐	☐	☐	☐
I understood how to use features without needing to experiment	☐	☐	☐	☐	☐

Decisional

7. If you could change ONE thing, what would it be?

---

8. Would you use this system in your actual work?

- [ ] Definitely Yes
- [ ] Probably Yes
- [ ] Maybe
- [ ] Probably Not
- [ ] Definitely Not

Part 2: ARIA-Specific Usability Assessment

9. Rate your agreement with each statement (1=Strongly Disagree, 5=Strongly Agree):

#	Statement	1	2	3	4	5
1	The file upload process was straightforward and worked as expected	☐	☐	☐	☐	☐
2	The manual requirement entry form had too many fields and was overwhelming	☐	☐	☐	☐	☐
3	The scoring criteria (Business Value, Cost, Risk, etc.) were clearly explained	☐	☐	☐	☐	☐
4	The AI prioritization took too long to process my requirements	☐	☐	☐	☐	☐
5	The priority scores and rankings were presented in a clear, understandable way	☐	☐	☐	☐	☐
6	The visualizations (charts) were confusing and didn't add value	☐	☐	☐	☐	☐
7	I trust the AI's prioritization recommendations for my requirements	☐	☐	☐	☐	☐
8	The export functionality was difficult to find and use	☐	☐	☐	☐	☐
9	ARIA would save me significant time compared to manual prioritization	☐	☐	☐	☐	☐
10	The overall workflow from input to results felt disjointed and illogical	☐	☐	☐	☐	☐

Part 3: Single Ease Question (SEQ)

10. Rate how easy it was to complete each task (1=Very Difficult, 5=Very Easy):

#	Task	1	2	3	4	5	N/A
1	Upload a requirements file (CSV/Excel)	☐	☐	☐	☐	☐	☐
2	Create a requirement manually	☐	☐	☐	☐	☐	☐
3	Understand the scoring criteria (Business Value, Cost, etc.)	☐	☐	☐	☐	☐	☐
4	Run the AI prioritization analysis	☐	☐	☐	☐	☐	☐
5	Understand the prioritization results	☐	☐	☐	☐	☐	☐
6	Interpret the priority scores and rankings	☐	☐	☐	☐	☐	☐
7	Use the visualizations (charts)	☐	☐	☐	☐	☐	☐
8	Export results as CSV	☐	☐	☐	☐	☐	☐
9	Export results as HTML report	☐	☐	☐	☐	☐	☐
10	Navigate between different screens/sections	☐	☐	☐	☐	☐	☐

Responses

We conducted the survey with a 10 respondents with different roles including developers, system analysts, managers, e.t.c. The raw responses are available at ARIAsurveyRes.xslx that is attached.

Results

Overall Scores

Metric	Score	Status
System Usability Scale (SUS)	**81.5 / 100**	Excellent (>68)
ARIA Usability Score	**81.5 / 100**	Excellent (>68)
Average Task Ease (SEQ)	**4.31 / 5.0**	Very Good (>3.5)
Gibson's Affordances	**3.67 / 5.0**	Moderate exploration
Norman's Affordances	**4.15 / 5.0**	Excellent signifiers
Future Adoption Intent	**70%**	Strong (7 of 10)

Individual Participant Scores

Name	Role	Usability	SEQ	Gibson	Norman
Respondent1	Manager	97.5	5.00	5.00	5.00
Respondent2	Developer	82.5	4.80	4.25	4.25
Respondent3	Methodist	90.0	5.00	4.50	4.75
Respondent4	Assistent	92.5	5.00	4.00	4.00
Respondent5	Developer	80.0	3.80	3.25	4.50
Respondent6	System Analyst	87.5	4.30	3.50	3.75
Respondent7	Developer	85.0	4.60	3.50	5.00
Respondent8	Backend developer	55.0	2.50	3.50	2.75
Respondent9	Developer	67.5	4.00	2.50	3.50
Respondent10	Admin	77.5	4.10	2.75	4.00

Key Findings

Top Issues to Fix:

1. **Export CSV Button** (71% explored) - Make prominent
2. **Priority Interpretation** (62% explored) - Add help
3. **Scoring Criteria** (60% explored) - Add tooltips
4. **Analyze Button** (56% explored) - Stronger design

Top Strengths:

1. Input fields (75% immediately clear)
2. Ranking list (75% immediately clear)
3. File upload (70% immediately clear)
4. Speed advantage (100% rated better/much better)
5. Strong emotional response (4.50 first impression)

Emotional Journey (1-5 scale)

- First impression: **4.50**
- Inputting data: **4.00**
- Waiting for results: **4.40**
- Viewing results: **4.40**

ARIA vs Manual Methods

- Speed/Efficiency: 100% Better/Much Better

- Objectivity: 60% Better/Much Better
- Insight Quality: 80% Better/Much Better
- Confidence: 60% Better/Much Better

Requested Changes By Users:

- "Ranking criteria interpretability (more elaborate explanations for scores and rankings)"
- "UI design in terms of colors and alignment"
- "add hints"
- "PDF export"

Analysis Insights

Gibson vs. Norman Balance:

- **Current:** Gibson 3.67 + Norman 4.15 = Good but not optimal
- **Target:** Gibson < 3.0 + Norman > 4.0 = Optimal discoverability
- **Gap:** 0.67 points → Main issue is feature discoverability, not clarity

Reflection on ORID Method vs. Traditional Interviews:
The ORID method seems significantly more effective than traditional interviews because it settles a structured progression: Objective facts (feature usage), Reflective emotions (satisfaction), Interpretive comparisons (affordances), Decisional outcomes (adoption intent). This makes users refrain from going with vague opinions. This layered approach makes possible a convergent validation: the export CSV issue appeared in objective data (71% explored), interpretive scores (Gibson 3.67 vs. Norman 4.15), and decisional feedback (user requests), quantifying exact problem severity and root cause (discoverability, not clarity). Traditional interviews would yield vague "hard to find" complaints, but ORID separated feelings from facts - high satisfaction (4.50, 100% speed approval) coexisted with specific usability gaps and generated actionable insights : 5 concrete improvements with effort estimates and measurable baselines, rather than generic "improve UX" recommendations.

Improvements to be made:
1. Make export button prominent (Low effort, High impact - 71% explored)

- **Effort:** ~15-20 lines edited in `results_screen.dart`
- **Changes:** Modify button styling (size, color, icon), reposition in layout hierarchy
- **Why Low:** Single component styling change. No logic modification

2. Add priority interpretation help/hints (Low effort, High impact - 62% explored)

- **Effort:** ~25-30 lines added across `results_screen.dart` and new tooltip widget
- **Changes:** Add info icons with tooltips/popups explaining priority scores

- **Why Low:** Flutter has built-in Tooltip widget. Simple integration

3. Add tooltips to scoring criteria (Low effort, High impact - 60% explored)

- **Effort:** ~20-25 lines added in `requirements_form_screen.dart`
- **Changes:** Wrap each criteria label with Tooltip widget explaining Business Value, Cost, Risk, etc.
- **Why Low:** Reusable Tooltip pattern. No backend changes needed

4. Strengthen analyze button design (Low effort, High impact - 56% explored)

- **Effort:** ~10-15 lines edited in `requirements_form_screen.dart` or `upload_screen.dart`
- **Changes:** Increase button size, add contrasting color, add analysis icon
- **Why Low:** Pure styling/visual design change

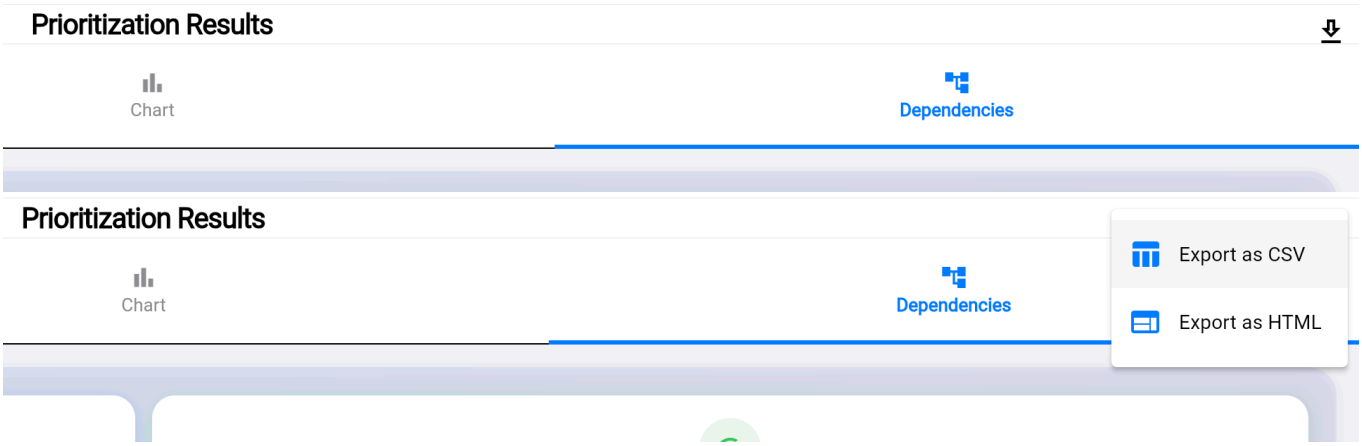
5. Consider PDF export feature (Medium effort, User request)

- **Effort:** ~80-120 lines added across frontend/backend
  - Frontend: ~40 lines (new button, PDF generation logic in `results_screen.dart`)
  - Backend: ~60 lines (new PDF export endpoint in `export_service.py`)
  - Dependencies: Add PDF library (e.g., `pdf` package for Flutter)
- **Changes:** New export button, PDF formatting logic, backend service method
- **Why Medium:** Requires new library integration, backend API endpoint, PDF templating

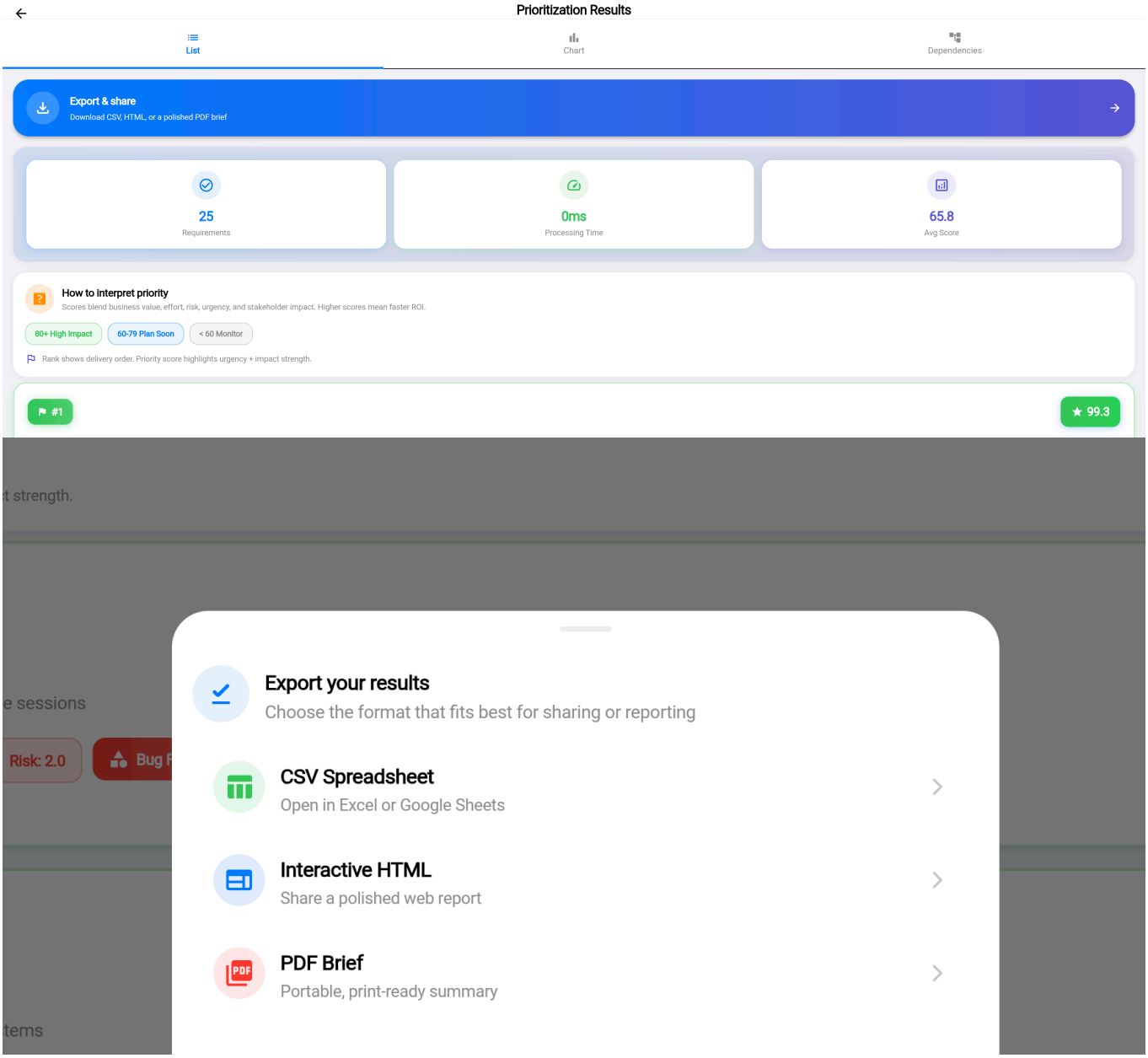
Changes were made

1. Make export button prominent

Before

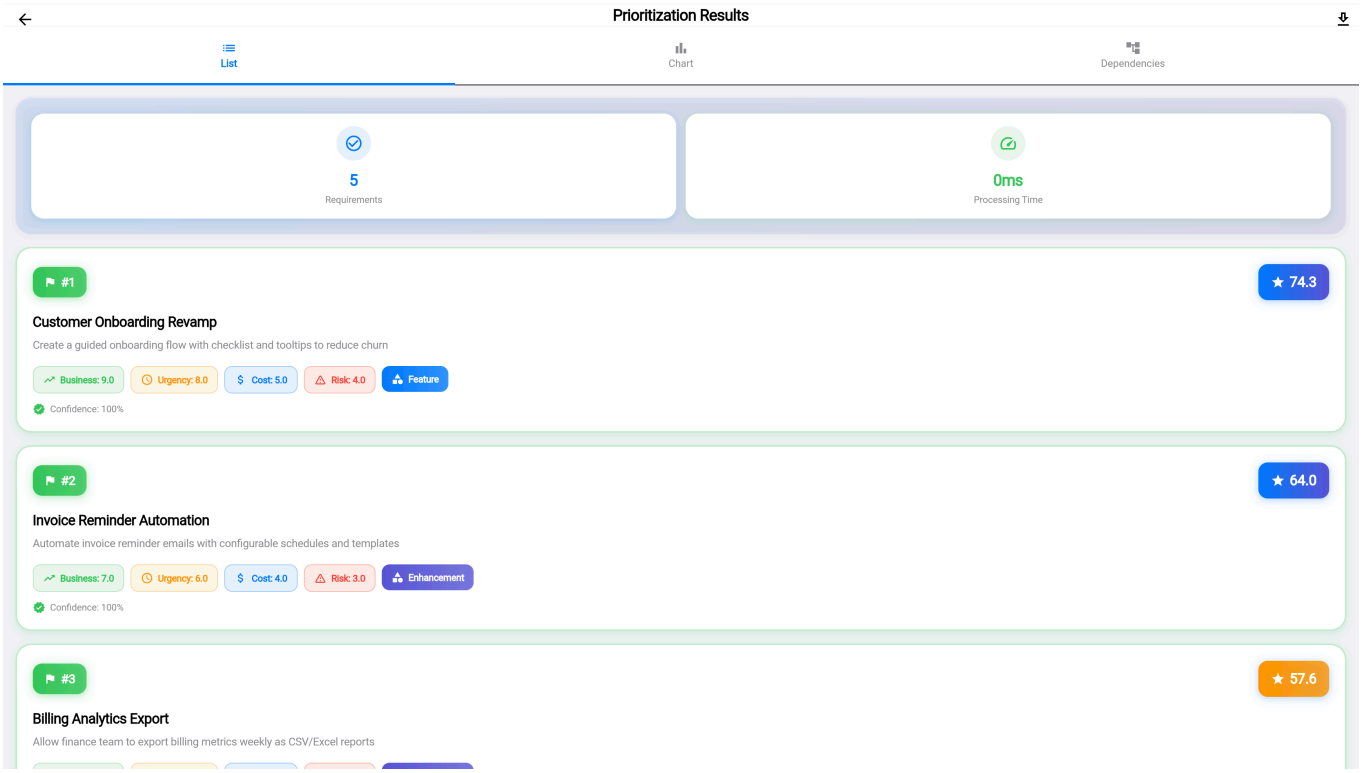


After

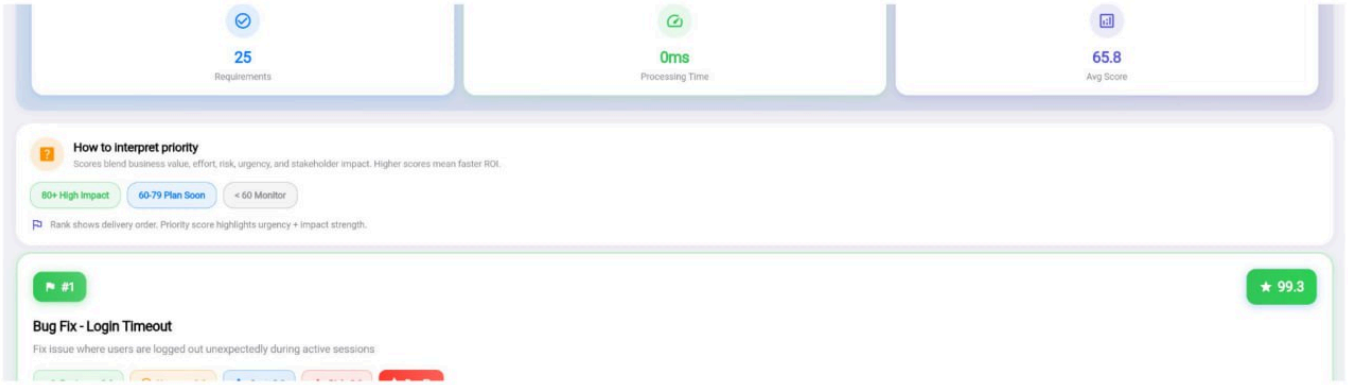


2. Add priority interpretation help/hints

Before

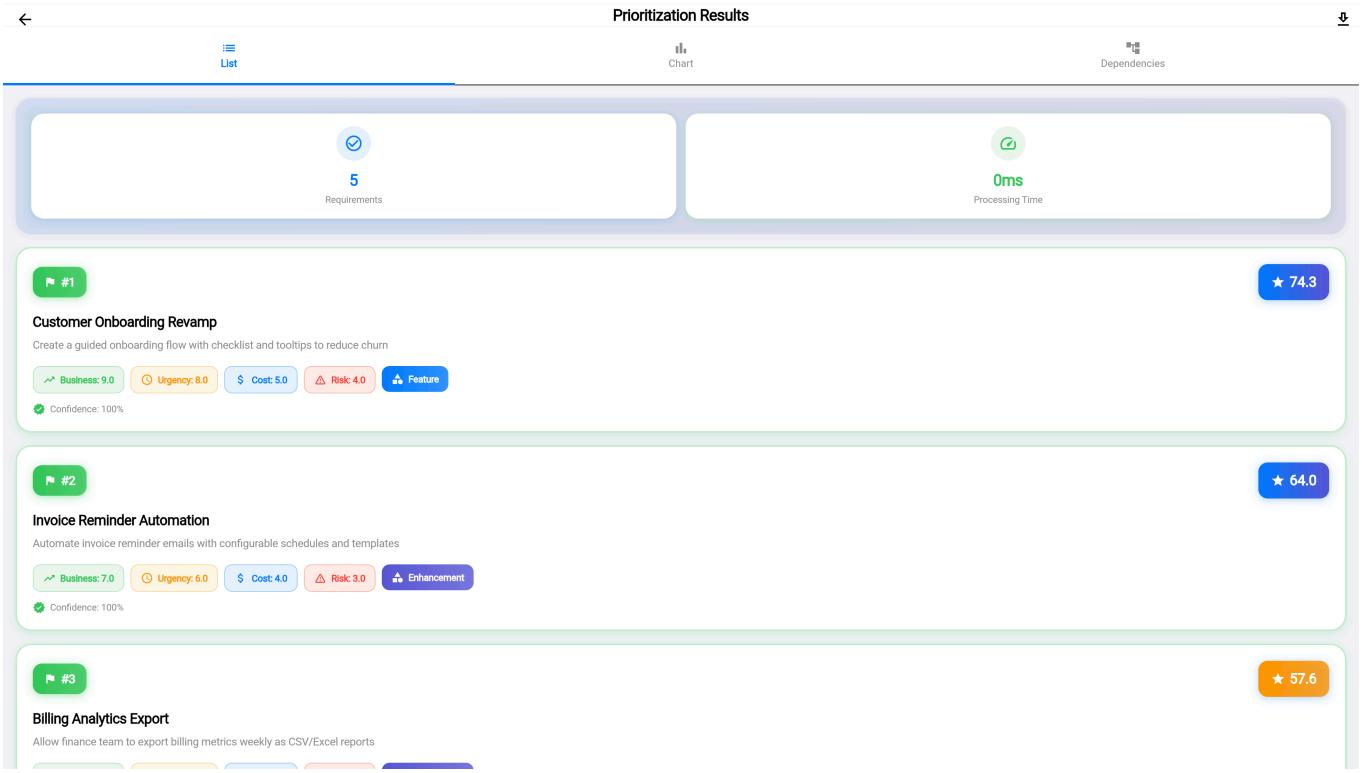


After

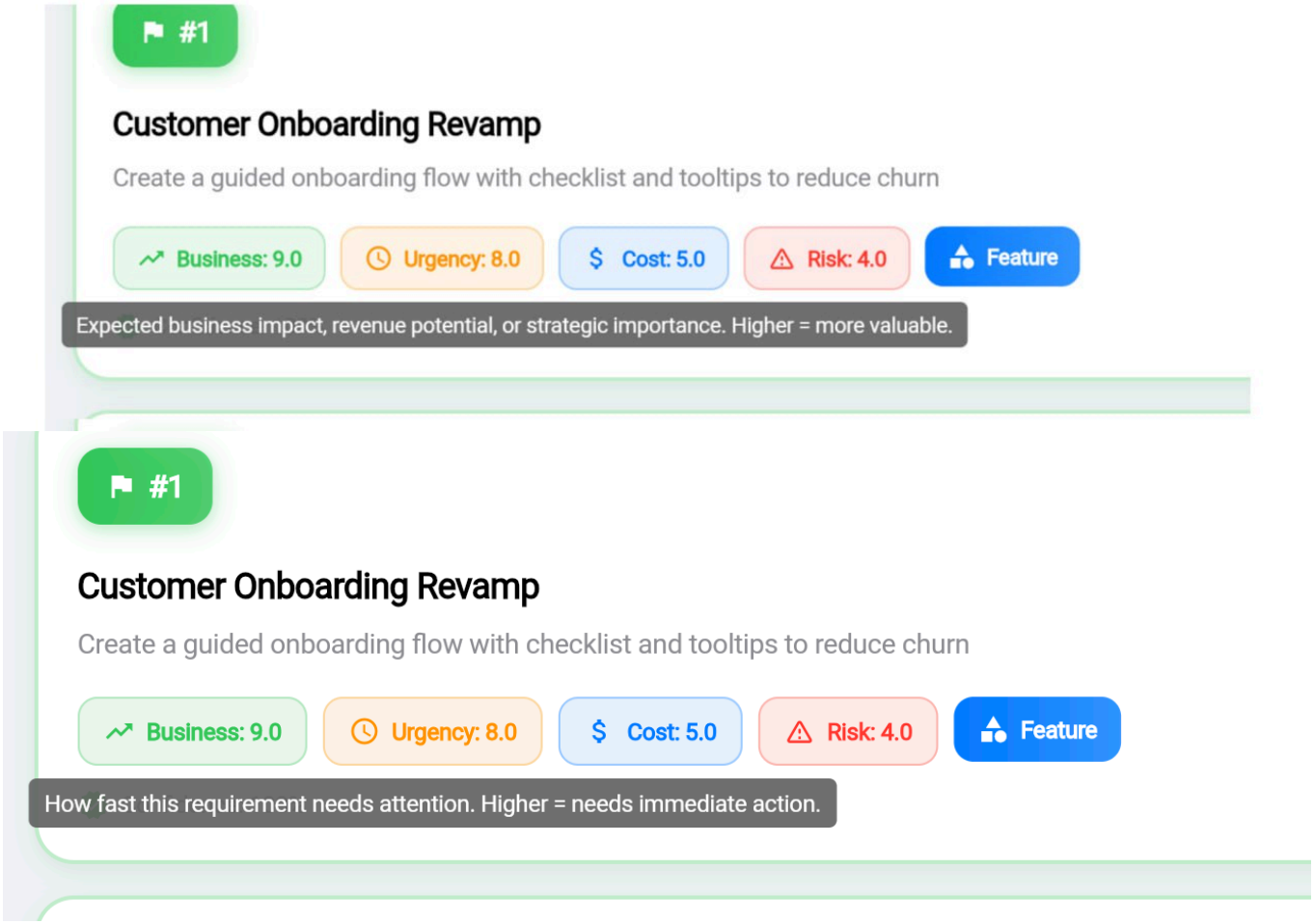


3. Add tooltips to scoring criteria

Before

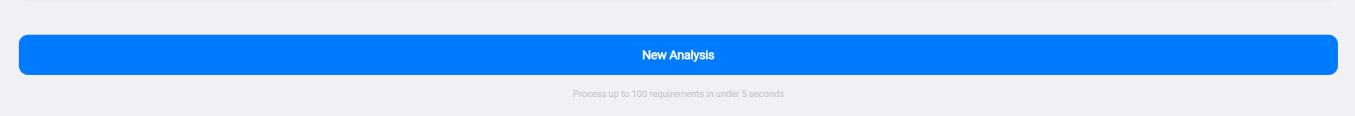


After

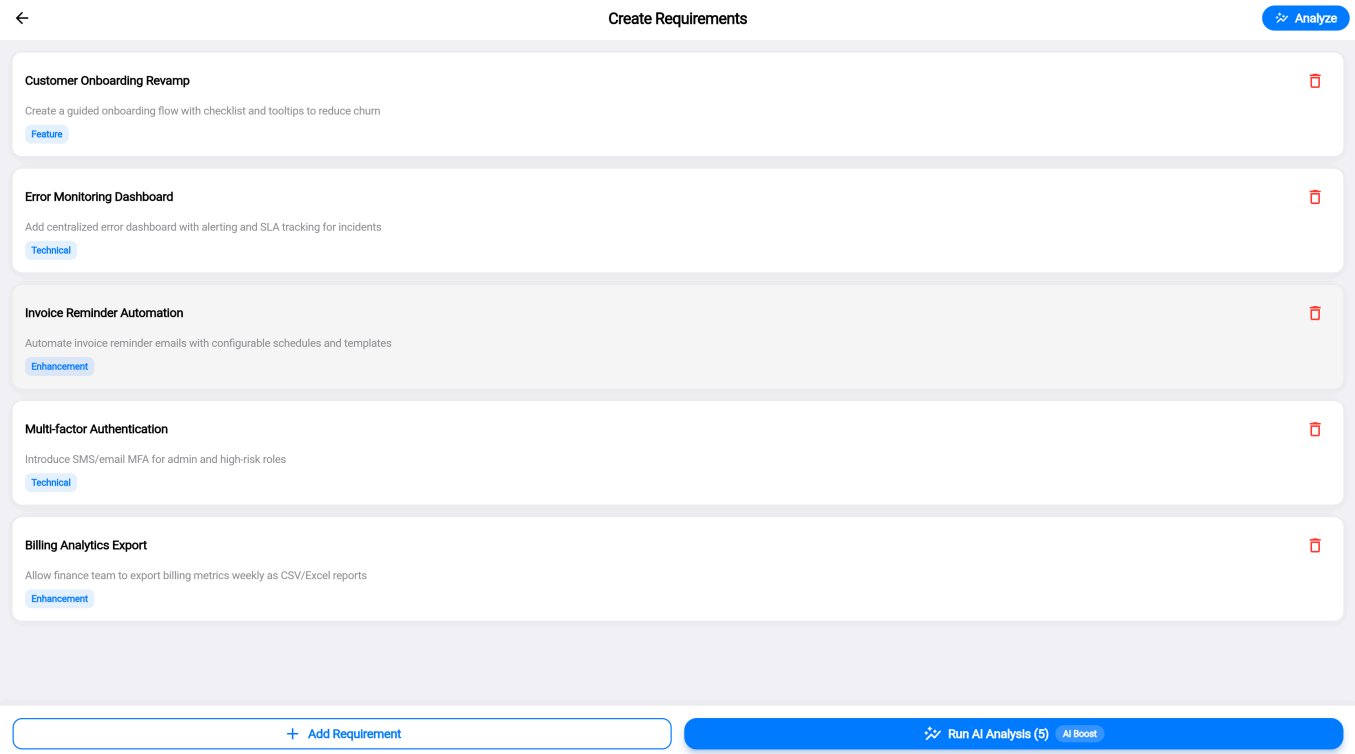


4. Strengthen analyze button design

Before

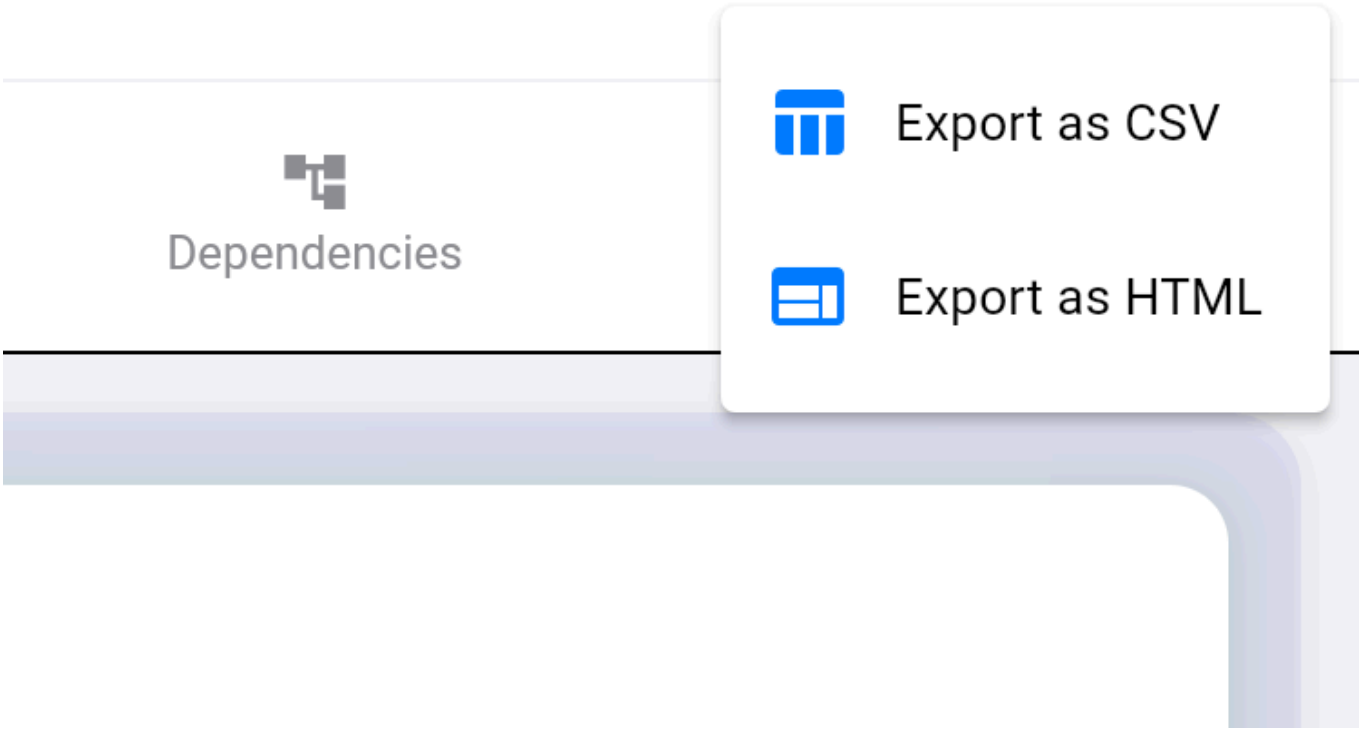


After

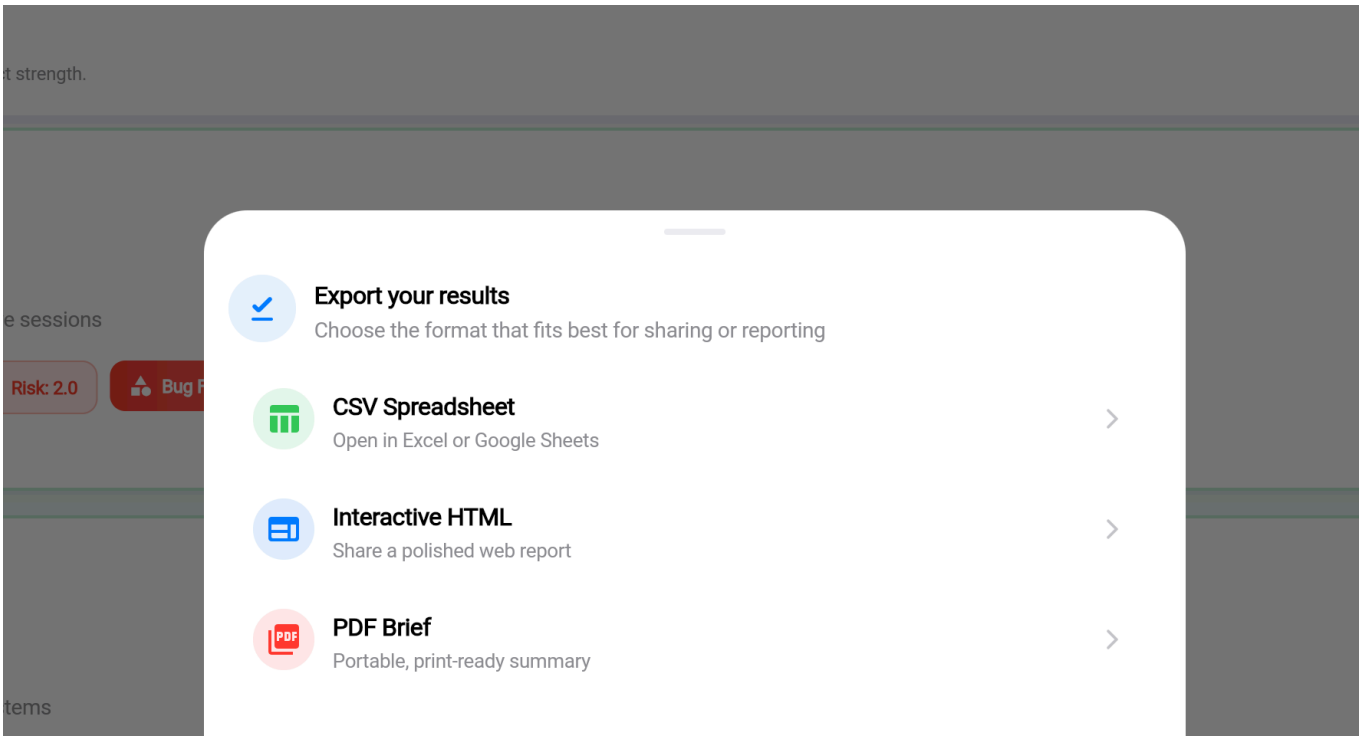


5. Consider PDF export feature

Before


Dependencies
Export as CSV
Export as HTML

After


t strength.
e sessions
Risk: 2.0
tems
Export your results
Choose the format that fits best for sharing or reporting
CSV Spreadsheet
Open in Excel or Google Sheets
Interactive HTML
Share a polished web report
PDF Brief
Portable, print-ready summary

ARIA Prioritization Report

Session ID: 9dcabc3e-96d5-4a8e-935e-7d30c2530c96

Generated: November 19, 2025 at 09:01 AM

Total Requirements	5
Average Priority Score	62.1
Average Confidence	96%

Prioritized Requirements

Title	Category	Score	Confidence	Business	Cost	Risk	Ur
Customer Onboarding Revamp	FEATURE	81.2	90%	9.0	5.0	4.0	8.0
Billing Analytics Export	ENHANCEMENT	58.5	92%	6.0	3.0	2.0	5.0
Invoice Reminder Automation	ENHANCEMENT	58.2	99%	7.0	4.0	3.0	6.0
Error Monitoring Dashboard	TECHNICAL	56.5	100%	8.0	6.0	5.0	7.0
Multi-factor Authentication	TECHNICAL	56.0	100%	9.0	7.0	6.0	9.0

Github Link

The project is available at: https://github.com/Legolass322/ms-re-r256