

30 Test Cases for Product Manager Agent Testing

Overview

This test suite evaluates an AI Agent designed to assist product managers with market and user research across 5 product categories: **AI Agent**, **Enterprise Internal Software**, **Hardware Product**, **Software Tools (SaaS)**, and **Consumer Mobile Apps**. Each test case includes realistic user queries, standard answers, and a detailed 4-dimensional scoring rubric.

Scoring Rubric Framework

Each test case is evaluated across 4 dimensions, each scored out of 10 points (40 points total):

Dimension A: User Requirements (10 points)

- **10 points:** Accurately identifies all explicit and implicit user needs; demonstrates deep understanding of the underlying business problem; correctly interprets technical and non-technical requirements
- **7 points:** Identifies most user needs; shows good understanding of the core problem; minor gaps in interpreting implicit requirements
- **4 points:** Identifies only surface-level needs; partial understanding of the problem; misses key implicit requirements
- **1 point:** Fails to identify user needs accurately; misinterprets the core problem; focuses on irrelevant aspects

Dimension B: Target Users (10 points)

- **10 points:** Precisely defines core user segments with demographic/psychographic details; identifies 3+ critical unmet needs with supporting evidence; explains pain point severity and frequency
- **7 points:** Identifies main user segments with reasonable detail; identifies 2 critical unmet needs; explains pain points with some evidence
- **4 points:** Generic user segment definition; identifies 1 unmet need; vague pain point description
- **1 point:** Incorrect or no user segment identification; fails to identify unmet needs; no pain point analysis

Dimension C: Market Analysis (10 points)

- **10 points:** Identifies 5+ relevant competitors with accurate positioning; provides market share data and differentiation analysis; includes emerging competitors and market gaps
- **7 points:** Identifies 3-4 competitors with positioning; provides some market data; mentions key differentiators
- **4 points:** Lists 1-2 competitors with minimal detail; lacks market data; superficial differentiation
- **1 point:** No competitors identified or completely irrelevant competitors; no market understanding

Dimension D: Market Insights (10 points)

- **10 points:** Provides 3+ actionable positioning recommendations with clear rationale; identifies specific market opportunities with TAM/SAM data; suggests differentiation strategies tied to user needs
 - **7 points:** Provides 2 positioning recommendations; identifies market opportunities with general sizing; suggests differentiation strategies
 - **4 points:** Provides 1 vague recommendation; mentions opportunities without sizing; generic strategies
 - **1 point:** No actionable insights; irrelevant recommendations; lacks strategic thinking
-

Test Cases by Product Type

Category 1: AI Agent Products

Test Case 1.1: AI Customer Support Agent

Test Case Type: AI Agent

User Question: "I'm considering building an AI agent that handles tier-1 customer support queries for e-commerce companies. What are the key user pain points I should solve, and who are my main competitors?"

Standard Answer:

User Requirements Analysis: The user needs to understand the unmet needs in AI-powered customer support specifically for e-commerce, identify which tier-1 queries are most painful to handle manually, and assess the competitive landscape

to find differentiation opportunities. The implicit requirement is to validate product-market fit before building.¹

Target Users:

- **Primary:** E-commerce companies with 50-500 employees, handling 1,000-10,000 customer inquiries monthly
- **Secondary:** Customer support managers and CX directors
- **Critical Unmet Needs:**
 - a. **24/7 multilingual support** without proportional cost scaling (current solutions require separate language models or human agents)²
 - b. **Order-specific context awareness** that integrates real-time with inventory, shipping, and payment systems (most chatbots lack deep integration)
 - c. **Seamless human handoff** when queries exceed AI capability, with full context transfer (current solutions lose conversation history)

Market Competitors:

1. **Zendesk AI** - Market leader with 30%+ enterprise share, strong in ticketing integration³
2. **Intercom Fin** - AI agent with \$13B valuation, focuses on mid-market SaaS⁴
3. **Ada** - Specialized in e-commerce, serves 400+ brands including Shopify merchants
4. **Gorgias** - E-commerce focused, integrated with Shopify/WooCommerce, \$29M ARR
5. **Kustomer (Meta)** - CRM-integrated agent, enterprise positioning⁵

Market Insights: The AI customer support market is growing at 29.5% CAGR, reaching \$11.14B in 2026. Key positioning strategies:

¹<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

²<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

³<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

⁴<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁵<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

1. **Vertical specialization:** Focus exclusively on e-commerce sub-verticals (fashion, electronics, food) rather than horizontal customer support
2. **Integration-first approach:** Build native integrations with Shopify, WooCommerce, BigCommerce before launch—not as afterthought
3. **Hybrid pricing model:** Offer per-resolution pricing (not per-seat) to align with ROI, since e-commerce operates on thin margins
4. **TAM opportunity:** The mid-market e-commerce segment (50-500 employees) represents \$2.8B addressable market with lower competition than enterprise

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies both explicit need (tier-1 support automation) and implicit needs (differentiation, PMF validation); 7 = Identifies main automation need; 4 = Only mentions general support automation; 1 = Misses the e-commerce context
 - **Target Users (10 pts):** 10 = Defines e-commerce company size, role personas, and 3 specific unmet needs with evidence; 7 = Identifies company profile and 2 needs; 4 = Generic "e-commerce companies" with 1 need; 1 = Wrong target or no pain points
 - **Market Analysis (10 pts):** 10 = Lists 5+ competitors with positioning and market share; 7 = Lists 3-4 competitors with some detail; 4 = Lists 1-2 competitors; 1 = No relevant competitors
 - **Market Insights (10 pts):** 10 = Provides vertical specialization strategy, pricing model innovation, and TAM sizing; 7 = Suggests 2 strategies with rationale; 4 = Generic advice; 1 = No actionable insights
-

Test Case 1.2: AI Sales Qualification Agent

Test Case Type: AI Agent

User Question: "What are the current gaps in AI sales qualification agents, and how can I differentiate my product in the market?"

Standard Answer:

User Requirements Analysis: The user seeks to identify underserved needs in AI-powered lead qualification (BDR/SDR automation) and discover white-space opportunities for differentiation. Implicit requirements include understanding buying criteria for sales tools and typical deal sizes.

Target Users:

- **Primary:** B2B SaaS companies with 10-200 sales reps, \$5M-\$100M ARR
- **Secondary:** VP Sales, Revenue Operations leaders
- **Critical Unmet Needs:**

- a. **Multi-channel engagement consistency** (email, LinkedIn, phone) - current agents operate in silos
- b. **Real-time qualification scoring** that updates based on prospect behavior across channels, not static demographic data
- c. **CRM data enrichment accuracy** - existing solutions have 60-70% accuracy, causing rep distrust⁶

Market Competitors:

1. **Gong Revenue AI** - Market leader in revenue intelligence, \$7.25B valuation, enterprise focus
2. **Drift (Salesloft)** - Conversational AI for inbound, 5,000+ customers⁷
3. **Exceed.ai** - Outbound AI assistant, acquired by Genesys 2021
4. **Qualified.com** - Pipeline Cloud platform, raised \$107M, mid-market position
5. **Conversica** - AI sales assistant, 1,500+ customers, revenue in \$50M+ range

Market Insights: AI agents market growing at 45.8% CAGR, from \$7.63B (2025) to \$50.31B (2030). Differentiation strategies:

1. **Industry-specific playbooks:** Pre-built qualification frameworks for healthcare, fintech, manufacturing rather than generic templates
2. **Two-way CRM sync with conflict resolution:** Most solutions push data one-way; build intelligent merge logic
3. **ROI transparency dashboard:** Show cost-per-qualified-lead vs human SDR benchmarks in real-time
4. **SAM focus:** Target the 15,000+ B2B SaaS companies with \$5M-\$50M ARR—underserved by enterprise tools but outgrowing point solutions

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies gap analysis need, differentiation need, and buying criteria understanding; 7 = Identifies gap and differentiation needs; 4 = Only mentions one requirement; 1 = Misunderstands query
- **Target Users (10 pts):** 10 = Specifies company profile, ARR range, personas, 3 unmet needs with accuracy data; 7 = Identifies profile and 2 needs; 4 = Generic target; 1 = Wrong segment

⁶<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁷<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

- **Market Analysis (10 pts):** 10 = Lists 5 competitors with valuation/customer data and positioning; 7 = Lists 3-4 with data; 4 = Lists 2 with minimal info; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Provides vertical playbook strategy, technical differentiation, ROI approach, and SAM sizing; 7 = Provides 2 strategies; 4 = Vague advice; 1 = No insights
-

Test Case 1.3: AI Code Review Agent

Test Case Type: AI Agent

User Question: "I want to build an AI agent that automates code reviews for development teams. Who would benefit most from this, and what's missing in current solutions?"

Standard Answer:

User Requirements Analysis: The user needs to identify the highest-value user segment for code review automation, understand what existing tools fail to address, and assess market readiness. The implicit need is to understand the technical feasibility and integration requirements.⁸

Target Users:

- **Primary:** Engineering teams at Series A-C startups (20-200 developers) with high velocity requirements
- **Secondary:** Engineering managers, DevOps leads at high-growth companies
- **Critical Unmet Needs:**
 - a. **Context-aware security vulnerability detection** beyond pattern matching—understanding business logic flaws in specific frameworks (Django, Rails, Spring)
 - b. **Learning from team-specific style guides**—current tools enforce generic rules, not company conventions
 - c. **PR review prioritization** based on risk and business impact, not just FIFO⁹

Market Competitors:

⁸<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

⁹<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

1. **GitHub Copilot** - \$100M ARR (estimated), 1M+ paid users, integrated in GitHub¹⁰
2. **Codacy** - Automated code review platform, 1,600+ customers, \$20M funding
3. **SonarQube** - Code quality platform, 400K+ organizations, €150M revenue (2023)
4. **DeepCode (acquired by Snyk)** - AI code review, now part of Snyk's \$7.4B platform
5. **CodeRabbit** - AI code reviews, launched 2023, fast-growing startup

Market Insights: Machine learning is a +3.2% CAGR driver in SaaS adoption.¹¹

Positioning strategies:

1. **Framework-first approach:** Start with deep Django or React Native expertise rather than supporting all languages superficially
2. **Compliance as a wedge:** Target regulated industries (fintech, healthcare) where code review audit trails are mandatory
3. **Usage-based pricing:** Per-PR pricing instead of per-seat, since review volume is variable
4. **Target market:** The 50,000+ startups with 20-200 developers represent a \$450M TAM, underserved by enterprise-focused incumbents

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies target user question, missing capabilities, market readiness, and integration needs; 7 = Identifies target user and gaps; 4 = Superficial understanding; 1 = Misses the point
 - **Target Users (10 pts):** 10 = Specifies startup stage, team size, personas, 3 technical unmet needs; 7 = Identifies segment and 2 needs; 4 = Generic "developers"; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with ARR/revenue/users and positioning; 7 = Lists 3-4 with data; 4 = Lists 2 competitors; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Framework focus, compliance wedge, pricing model, TAM sizing; 7 = Two strategies with rationale; 4 = Generic advice; 1 = No insights
-

¹⁰<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

¹¹<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

Test Case 1.4: AI Personal Finance Agent

Test Case Type: AI Agent

User Question: "What are users looking for in AI-powered personal finance agents that current apps don't provide?"

Standard Answer:

User Requirements Analysis: The user needs to understand unmet consumer needs in AI personal finance management, identify friction points in existing solutions, and find opportunities for a consumer-focused AI agent. Implicit needs include understanding monetization models and regulatory constraints.

Target Users:

- **Primary:** Millennials and Gen Z (25-40 years old) with \$50K-\$150K household income, tech-savvy, interested in optimization
- **Secondary:** Financial advisors serving mass-affluent clients
- **Critical Unmet Needs:**
 - a. **Proactive financial opportunities** (refinancing, tax optimization, benefit enrollment) instead of reactive tracking—current apps only show what happened
 - b. **Multi-institution actionability**—users can see all accounts but can't act (transfer, invest, pay) without leaving the app
 - c. **Conversational financial planning** that explains tradeoffs in natural language, not spreadsheets

Market Competitors:

1. **Mint (Intuit)** - 20M+ users, free model, being shut down Dec 2023
2. **Copilot** - Premium finance app, \$2.99-\$8.99/month, Apple ecosystem focus
3. **Monarch Money** - Collaborative finance app, \$99-\$165/year, couples focus
4. **YNAB (You Need A Budget)** - \$99/year, zero-based budgeting methodology, 500K+ users
5. **Rocket Money (Truebill)** - Subscription cancellation + budgeting, acquired by Rocket Companies \$1.275B

Market Insights: Consumer mobile app spending reached \$156B in 2025, up from prior year. Differentiation strategies:

1. **Action-first architecture:** Partner with banks and brokerages for in-app transactions, not read-only aggregation
2. **Life-event triggers:** Build agents that detect life moments (job change, home purchase, marriage) and provide contextual financial planning
3. **Freemium with AI upsell:** Free basic aggregation, charge \$15-30/month for AI agent features (industry benchmark is \$8.99/month)

4. **TAM:** 80M+ US households with \$50K+ income who use digital banking = \$6B+ addressable market for premium personal finance tools

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies unmet needs, friction points, monetization models, regulations; 7 = Identifies needs and friction; 4 = Surface-level needs; 1 = Misses consumer context
 - **Target Users (10 pts):** 10 = Age range, income, psychographics, 3 specific unmet needs with proactive vs reactive distinction; 7 = Demographics and 2 needs; 4 = "Millennials" with 1 vague need; 1 = Wrong segment
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 apps; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Action architecture, life-event triggers, pricing strategy, TAM; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 1.5: AI Recruiting Agent

Test Case Type: AI Agent

User Question: "I'm building an AI agent for recruiting teams to screen candidates faster. What pain points should I prioritize, and who's already in this space?"

Standard Answer:

User Requirements Analysis: The user needs to understand recruiter pain points in candidate screening, prioritize which pain points have the highest ROI to solve, identify competitive landscape in AI recruiting, and understand buyer personas. Implicit needs include compliance (EEOC, bias) and ATS integration requirements.¹²

Target Users:

- **Primary:** In-house recruiting teams at tech companies with 100-2,000 employees (5-40 recruiters)
- **Secondary:** Recruiting operations managers, TA leaders
- **Critical Unmet Needs:**
 - a. **Skills inference from unstructured experience**—resumes list job duties, not skills; agents need to infer "Python" from "built data pipeline"
 - b. **Candidate engagement quality** (writing personalized outreach at scale)—current tools send templates that candidates ignore

¹²<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

- c. **Bias auditing transparency**—showing which factors influenced scoring to meet regulatory requirements¹³

Market Competitors:

1. **HireVue** - Video interview AI, 1,100+ customers, enterprise focus, \$93M funding
2. **Paradox (Olivia)** - Conversational AI recruiter, 1,500+ customers, \$200M valuation
3. **Fetcher** - AI-powered sourcing, focused on startups, \$14M funding
4. **SeekOut** - Diversity-focused AI sourcing, \$189M funding, Microsoft acquisition rumors
5. **Eightfold.ai** - Talent intelligence platform, \$2.1B valuation, 1,000+ customers¹⁴

Market Insights: AI adoption is accelerating in HR tech with enterprise SaaS projected to grow 17.9% CAGR.¹⁵ Positioning strategies:

1. **Compliance-first messaging:** Lead with bias auditing and EEOC compliance rather than just "speed"—critical for risk-averse HR buyers
2. **ATS-agnostic architecture:** Build connectors for Greenhouse, Lever, Workday first—not proprietary ATS
3. **Role-specific models:** Train separate models for engineering vs sales vs operations roles rather than one generic screener
4. **SAM focus:** 15,000+ tech companies with 100-2,000 employees represent \$1.2B SAM for recruiting AI tools

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies pain prioritization, competitive landscape, buyer personas, compliance, ATS needs; 7 = Pain points and competitors; 4 = Surface pain points; 1 = Misses screening context
- **Target Users (10 pts):** 10 = Company size, team size, personas, 3 unmet needs with skills/engagement/bias specifics; 7 = Profile and 2 needs; 4 = "Recruiters" with 1 need; 1 = Wrong target

¹³<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

¹⁴<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

¹⁵<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

- **Market Analysis (10 pts):** 10 = Lists 5 competitors with customers/funding/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 players; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Compliance messaging, ATS strategy, role-specific models, SAM sizing; 7 = Two strategies; 4 = Vague advice; 1 = No insights
-

Test Case 1.6: AI Meeting Assistant Agent

Test Case Type: AI Agent

User Question: "What features do users expect from AI meeting assistants in 2026, and how saturated is this market?"

Standard Answer:

User Requirements Analysis: The user needs to understand evolved user expectations for AI meeting tools (beyond basic transcription), assess market saturation and remaining white space, and identify what drives purchase decisions. Implicit needs include understanding enterprise security requirements and pricing benchmarks.

Target Users:

- **Primary:** Knowledge workers at B2B companies with 50-5,000 employees who attend 10+ meetings per week
- **Secondary:** Sales teams, customer success teams, executive assistants
- **Critical Unmet Needs:**
 - a. **Cross-meeting insights**—surfacing themes across 20 customer calls, not just transcribing each call in isolation
 - b. **Automated follow-up actions** with integration (creating tickets, updating CRM, sending emails) without manual copy-paste
 - c. **Privacy-preserving recording**—attendees concerned about always-on recording; need selective recording with explicit consent

Market Competitors:

1. **Otter.ai** - 6M+ users, freemium model, \$8.83/user/month premium, consumer + SMB focus
2. **Fireflies.ai** - 300K+ organizations, integrates with 50+ apps, \$10-\$19/user/month
3. **Fathom** - Free meeting assistant, monetizes through integrations, Series A \$17M
4. **Gong** - Enterprise revenue intelligence, \$7.25B valuation, \$150-\$300/user/year, sales focus

5. **Microsoft Copilot** - Integrated in Teams, part of \$30/user/month M365 bundle, 1M+ enterprise seats¹⁶

Market Insights: AI chatbot market at \$11.14B in 2026 growing 29.5% annually. The market is saturated for basic transcription but white space exists:

1. **Vertical specialization:** Medical (HIPAA-compliant), legal (privilege-aware), education (lecture-specific formatting) vs horizontal tools
2. **Freemium + usage cap:** Offer free tier with 10 hours/month, charge \$15-25/user for unlimited (market rate: \$10-19/user¹⁷)
3. **Privacy toggle positioning:** Market as "consent-first" meeting assistant—differentiate from always-on competitors
4. **Market opportunity:** 100M+ knowledge workers globally, but only 20M using paid meeting assistants = \$3B TAM at \$150 blended ASP

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies feature expectations, saturation assessment, white space, purchase drivers, security; 7 = Features and saturation; 4 = Basic features only; 1 = Misunderstands query
 - **Target Users (10 pts):** 10 = Company size, meeting frequency, personas, 3 unmet needs with cross-meeting/automation/privacy specifics; 7 = Profile and 2 needs; 4 = "Meeting users" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 tools; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Vertical specialization, pricing model, privacy positioning, TAM opportunity; 7 = Two strategies; 4 = Generic advice; 1 = No insights
-

Category 2: Enterprise Internal Software

Test Case 2.1: Internal Knowledge Management System

Test Case Type: Enterprise Internal Software

User Question: "Our company wants to build an internal knowledge base that's actually used. What are the common reasons employees don't use these systems, and what should we learn from successful products?"

¹⁶<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

¹⁷<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

Standard Answer:

User Requirements Analysis: The user needs to understand adoption barriers for internal knowledge management systems, identify what drives employee engagement, and learn from successful implementations. Implicit needs include change management strategies and measuring usage success metrics.¹⁸

Target Users:

- **Primary:** Enterprise companies with 500-10,000 employees, distributed teams, high information volume
- **Secondary:** Knowledge managers, IT leaders, department heads
- **Critical Unmet Needs:**
 - a. **Search that understands intent**, not just keywords— "how to submit expenses" should find the policy even if document says "reimbursement process"
 - b. **Automatic content freshness scoring**—flag outdated articles instead of surfacing 3-year-old onboarding docs
 - c. **Contextual delivery** (Slack, Teams) rather than forcing users to leave workflow to visit separate portal¹⁹

Market Competitors:

1. **Confluence (Atlassian)** - Market leader, 75,000+ customers, \$4-\$8/user/month, technical teams focus
2. **Notion** - 35M+ users, \$10-\$15/user/month, modern UX, startup/SMB focus
3. **SharePoint (Microsoft)** - 400M+ users, part of M365, enterprise standard but criticized for complexity
4. **Guru** - AI-powered knowledge management, \$15-\$20/user/month, 2,000+ customers
5. **Document360** - Knowledge base software, \$149-\$599/month, mid-market focus²⁰

Market Insights: Enterprise collaboration software market growing from \$48.9B (2025) to \$143.9B (2035). Success factors:

¹⁸<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

¹⁹<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

²⁰<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

1. **Embed in workflow tools:** Build Slack/Teams apps where employees already work—don't force portal visits (Guru's strategy²¹)
2. **Gamification + incentives:** Award points for creating/updating content and tie to performance reviews
3. **AI-assisted authoring:** Use LLMs to convert Slack threads into polished docs automatically, reducing creation friction
4. **Target market:** 120,000+ enterprises with 500-10,000 employees represent \$8.5B TAM for modern knowledge management, as SharePoint replacements accelerate

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies adoption barriers, engagement drivers, success implementations, change management, metrics; 7 = Barriers and drivers; 4 = Surface barriers only; 1 = Misses adoption focus
 - **Target Users (10 pts):** 10 = Company size, employee count, distribution, personas, 3 unmet needs with search/freshness/context specifics; 7 = Profile and 2 needs; 4 = Generic "employees"; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with customers/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 tools; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Workflow embedding, gamification, AI authoring, TAM sizing; 7 = Two strategies; 4 = Vague advice; 1 = No insights
-

Test Case 2.2: Employee Onboarding Platform

Test Case Type: Enterprise Internal Software

User Question: "What are the biggest pain points in employee onboarding that technology can solve, and which companies are doing this well?"

Standard Answer:

User Requirements Analysis: The user needs to identify technology-solvable onboarding pain points (vs process issues), understand what "good" onboarding looks like, and learn from market leaders. Implicit needs include understanding time-to-productivity metrics and integration with HRIS systems.²²

Target Users:

²¹<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

²²<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

- **Primary:** HR operations teams at companies with 200-5,000 employees and >50 new hires per year
- **Secondary:** Department managers, new hires themselves
- **Critical Unmet Needs:**
 - a. **Role-specific onboarding paths**— not one-size-fits-all; engineers need GitHub access, sales need Salesforce training
 - b. **Cross-department coordination** (IT, facilities, payroll)— current systems are siloed, causing delays
 - c. **Progress visibility for managers**— managers don't know if new hire completed setup until week 2²³

Market Competitors:

1. **BambooHR** - 35,000+ customers, \$99-\$299/month + employee pricing, SMB focus with onboarding module
2. **Workday** - Enterprise HCM leader, \$6.2B revenue (2024), expensive (\$100-\$300/employee/year), Fortune 500 focus
3. **Sapling (Kallidus)** - Onboarding specialist, 500+ customers, acquired 2021, mid-market
4. **Enboarder** - Journey-based onboarding, \$8-\$12/employee/month, 500+ customers
5. **WorkBright** - Onboarding for hourly/remote workers, 4,000+ customers, deskless workforce focus²⁴

Market Insights: Enterprise collaboration market expanding 194% to \$143.9B by 2035. Differentiation strategies:

1. **Industry-specific templates:** Pre-built onboarding for healthcare, construction, retail vs generic enterprise
2. **Manager copilot feature:** Proactive Slack notifications to managers about new hire progress and suggested check-ins
3. **Onboarding analytics:** Show time-to-productivity by role, department, manager to drive continuous improvement
4. **TAM focus:** 45,000+ companies with 200-5,000 employees in US = \$2.7B SAM for standalone onboarding platforms

Scoring Rubric:

²³<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

²⁴<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

- **User Requirements (10 pts):** 10 = Identifies tech-solvable pain points, success definition, market leaders, TTT metrics, HRIS integration; 7 = Pain points and leaders; 4 = Surface pain points; 1 = Confuses onboarding with training
 - **Target Users (10 pts):** 10 = Company size, hiring volume, personas, 3 unmet needs with role-specific/coordination/visibility specifics; 7 = Profile and 2 needs; 4 = "HR teams" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with customers/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 systems; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Industry templates, manager copilot, analytics, TAM focus; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 2.3: Internal Communication Tool

Test Case Type: Enterprise Internal Software

User Question: "We're developing an internal communication platform. What frustrates employees about tools like Slack and Teams, and is there still room for a new player?"

Standard Answer:

User Requirements Analysis: The user needs to understand employee frustrations with dominant communication tools, identify white space opportunities, assess market feasibility for a new entrant, and understand switching costs. Implicit needs include understanding network effects and integration ecosystem requirements.²⁵

Target Users:

- **Primary:** Knowledge workers at companies with 100-10,000 employees using Slack or Teams
- **Secondary:** IT administrators, department heads
- **Critical Unmet Needs:**
 - a. **Notification overload management**—users mute channels to cope, then miss important updates; need intelligent priority ranking
 - b. **Async communication structure**—real-time chat creates urgency culture; need hybrid async/sync designed for deep work
 - c. **Knowledge retrieval from chat history**—finding that discussion from 3 months ago is nearly impossible without perfect keyword recall

²⁵<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

Market Competitors:

1. **Slack (Salesforce)** - 20M+ daily active users, \$27.50-\$34.25/user/month, mid-market + enterprise leader
2. **Microsoft Teams** - 320M+ monthly active users, included in M365 (\$12.50-\$57/user/month), enterprise dominant
3. **Zoom Team Chat** - 3.3M+ customers (Zoom), \$149-\$250/year/license, video-first positioning
4. **Discord** - 200M+ monthly users, expanding to workplace via Midjourney, Anthropic adoption
5. **Twist (Doist)** - Async-first communication, \$6.25/user/month, small teams focus

Market Insights: Collaboration tools market valued at \$48.9B (2025) growing to \$143.9B (2035). Market is highly concentrated, but opportunities exist:

1. **Industry-specific safety:** Build for regulated industries (healthcare, finance) with built-in compliance features (message retention, audit logs) vs retrofitting
2. **Async-first architecture:** Position as "anti-Slack" for deep work cultures—default to threads with optional real-time
3. **Unbundled pricing:** Charge \$5-8/user/month vs Slack's \$27.50—target cost-conscious mid-market
4. **Realistic TAM:** The "Slack fatigue" market is real but switching costs are high; focus on 10,000 companies growing from email to first team chat tool = \$500M SAM

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies frustrations, white space, market feasibility, switching costs, network effects, integration needs; 7 = Frustrations and feasibility; 4 = Surface frustrations; 1 = Misses market dynamics
 - **Target Users (10 pts):** 10 = Company size, current tool usage, personas, 3 unmet needs with notification/async/retrieval specifics; 7 = Profile and 2 needs; 4 = "Employees" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 tools; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Industry safety positioning, async architecture, unbundled pricing, realistic TAM; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 2.4: IT Asset Management System

Test Case Type: Enterprise Internal Software

User Question: "What are the unmet needs in IT asset management software for mid-size companies, and who dominates this market?"

Standard Answer:

User Requirements Analysis: The user needs to understand mid-size company ITAM pain points (different from enterprise or SMB), identify capabilities missing in current solutions, and map the competitive landscape. Implicit needs include understanding integration with procurement, HRIS, and MDM systems.²⁶

Target Users:

- **Primary:** IT operations teams at companies with 500-5,000 employees and 1,000-15,000 assets
- **Secondary:** IT directors, procurement managers, finance (for depreciation tracking)
- **Critical Unmet Needs:**
 - a. **Automated asset lifecycle management**—manual tracking when laptops should be refreshed based on actual usage, not just purchase date
 - b. **Shadow IT discovery**—employees install SaaS tools; need automatic detection from SSO logs, credit card systems
 - c. **Cost allocation by department/project**—finance needs to allocate \$2M IT spend across 30 departments but current ITAM tools don't connect to accounting systems²⁷

Market Competitors:

1. **ServiceNow ITAM** - Part of \$8.9B ITSM platform, \$100-\$200/user/year, enterprise focus, market leader
2. **Lansweeper** - Network asset discovery, 25,000+ customers, \$899-\$5,999/year for 100-500 assets, SMB/mid-market
3. **Ivanti (formerly Landesk)** - Unified endpoint management, \$100-\$150/device/year, enterprise IT focus
4. **ManageEngine AssetExplorer** - Zoho product, \$595-\$2,495/year for 250-1,000 nodes, SMB focus

²⁶<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

²⁷<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

5. **Flexera One** - ITAM and FinOps, acquired Snow Software 2021, enterprise positioning²⁸

Market Insights: Enterprise software trends driven by hybrid cloud adoption and automation. Positioning strategies:

1. **Mid-market focus:** Build for 500-5,000 employee companies explicitly — not downselling from enterprise or upselling from SMB
2. **FinOps integration:** Connect ITAM data to cloud cost management (AWS, Azure, GCP) for total tech spend visibility
3. **Self-service portal:** Let employees request upgrades, swap hardware via portal instead of IT tickets
4. **TAM opportunity:** 65,000+ US companies with 500-5,000 employees × \$20K-\$80K average deal size = \$2B TAM for mid-market ITAM

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies mid-size pain points, missing capabilities, competitive landscape, integration with procurement/HRIS/MDM; 7 = Pain points and competitors; 4 = Generic ITAM needs; 1 = Misses mid-size focus
 - **Target Users (10 pts):** 10 = Company size, asset count, personas, 3 unmet needs with lifecycle/shadow IT/allocation specifics; 7 = Profile and 2 needs; 4 = "IT teams" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with customers/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 systems; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Mid-market focus, FinOps integration, self-service portal, TAM sizing; 7 = Two strategies; 4 = Vague; 1 = No insights
-

Test Case 2.5: Internal Project Management Platform

Test Case Type: Enterprise Internal Software

User Question: "Why do engineering teams hate project management tools, and what would make them actually enjoy using one?"

Standard Answer:

User Requirements Analysis: The user needs to understand why PM tools have low adoption among technical teams, identify features that engineers would value, and learn what successful engineering-first tools do differently. Implicit needs

²⁸<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

include understanding developer workflow integration and lightweight vs heavyweight trade-offs.²⁹

Target Users:

- **Primary:** Software engineering teams at product companies with 20-500 engineers
- **Secondary:** Engineering managers, technical project managers, product managers
- **Critical Unmet Needs:**
 - a. **Git integration for automatic status updates**—update ticket status based on commit messages and PR events, not manual clicking
 - b. **CLI and API-first design**—engineers prefer terminal/scripts to clicking through UIs; need full functionality via API
 - c. **Minimal process overhead**—lightweight issue tracking that doesn't enforce waterfall ceremonies (estimates, story points, sprint planning) on teams doing continuous delivery³⁰

Market Competitors:

1. **Jira (Atlassian)** - 200,000+ customers, \$8.15-\$16/user/month, market leader but criticized for complexity
2. **Linear** - Modern issue tracking, 8,000+ teams, \$8-\$12/user/month, engineering-first design, fast-growing
3. **GitHub Issues/Projects** - Included with GitHub, 100M+ developers, tight Git integration
4. **Asana** - 135,000+ customers, \$10.99-\$24.99/user/month, general-purpose PM but some eng adoption
5. **Shortcut (formerly Clubhouse)** - Engineering project management, \$8.50-\$12/user/month, 2,000+ teams³¹

Market Insights: Collaboration software market growing at \$48.9B to \$143.9B trajectory. Success factors:

1. **Developer experience obsession:** Optimize for keyboard shortcuts, dark mode, fast load times—treat like building for engineers

²⁹<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

³⁰<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

³¹<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

2. **Git-native architecture:** Make Git the source of truth, PM tool is read-only view—reverse of typical architecture
3. **Flexible methodology:** Don't force Scrum; support Kanban, continuous delivery, and hybrid approaches equally
4. **Target market:** 25,000+ product companies with 20-500 engineers = \$1.8B TAM for engineering-first PM tools at \$12K-\$60K ACV

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies low adoption reasons, valued features, successful tools, workflow integration, lightweight vs heavyweight; 7 = Adoption issues and features; 4 = Surface complaints; 1 = Generic PM needs
 - **Target Users (10 pts):** 10 = Company type, team size, personas, 3 unmet needs with Git/CLI/minimal process specifics; 7 = Profile and 2 needs; 4 = "Developers" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with customers/pricing/positioning and Jira criticism; 7 = Lists 3-4 with data; 4 = Lists 2 tools; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Dev experience obsession, Git-native architecture, flexible methodology, TAM sizing; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 2.6: Internal Analytics Dashboard

Test Case Type: Enterprise Internal Software

User Question: "We're building an internal analytics dashboard for non-technical employees. What mistakes do most internal BI tools make, and how can we avoid them?"

Standard Answer:

User Requirements Analysis: The user needs to understand common BI tool failure patterns, identify what makes dashboards actually used by non-technical users, learn from both failures and successes, and understand the technical/non-technical user balance. Implicit needs include data governance, refresh frequency, and training requirements.

Target Users:

- **Primary:** Non-technical business users (sales, marketing, ops, finance) at data-driven companies with 100-2,000 employees
- **Secondary:** Data analysts, BI team leads, department heads
- **Critical Unmet Needs:**

- a. **Natural language queries**—non-technical users struggle with SQL and filters; need "show me top customers by revenue this quarter" interface
- b. **Contextual explanations**—dashboards show metrics without explaining what's good/bad or why numbers changed
- c. **Collaboration on data**—can't annotate charts or discuss insights in-tool; must screenshot and Slack³²

Market Competitors:

1. **Tableau (Salesforce)** - Market leader, 90,000+ customers, \$70-\$42/user/month, enterprise standard
2. **Power BI (Microsoft)** - 6.5M+ users, \$10-\$20/user/month, included in some M365 plans, tight Excel integration
3. **Looker (Google)** - Part of Google Cloud, \$5,000+/month, data-modeling focus, technical users
4. **Metabase** - Open-source BI, 50,000+ deployments, free self-hosted or \$85+/month cloud, SMB focus
5. **Mode** - SQL + visualization, \$200-\$500/month, data analyst focus³³

Market Insights: Enterprise software pushing toward AI-driven insights and low-code approaches. Common mistakes and solutions:

1. **Building for analysts, not end users:** Most internal tools get built by data teams for data teams—then sales/marketing struggle to use them
2. **AI-assisted insight generation:** Use LLMs to auto-generate text explanations for why metrics changed ("Revenue up 15% driven by EMEA region")
3. **Embed in workflow:** Deliver insights via Slack/Teams/email rather than forcing dashboard visits—"push" vs "pull"
4. **Market positioning:** 35,000+ companies with 100-2,000 employees building data cultures = \$1.5B TAM for accessible BI tools at \$10K-\$50K ACV

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies common mistakes, success factors, failures vs successes, technical/non-technical balance, governance, refresh, training; 7 = Mistakes and success factors; 4 = Surface mistakes; 1 = Misses non-technical focus

³²<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

³³<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

- **Target Users (10 pts):** 10 = User roles, company size, data maturity, personas, 3 unmet needs with NLP/context/collaboration specifics; 7 = Profile and 2 needs; 4 = "Business users" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with customers/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 tools; 1 = No competitors
 - **Market Insights (10 pts):** 10 = End-user focus, AI insights, workflow embedding, TAM positioning; 7 = Two solutions; 4 = Generic advice; 1 = No insights
-

Category 3: Hardware Products

Test Case 3.1: Smart Home Security Camera

Test Case Type: Hardware Product

User Question: "I'm designing a smart home security camera. What features do users actually care about vs marketing hype, and who are the market leaders?"

Standard Answer:

User Requirements Analysis: The user needs to distinguish between must-have features and nice-to-have marketing features, understand customer purchase criteria, identify market leaders and their positioning, and understand the hardware/software/service bundle model. Implicit needs include manufacturing costs, retail vs DTC channels, and subscription revenue models.³⁴

Target Users:

- **Primary:** Homeowners aged 30-55 with \$75K+ household income, suburban/urban, tech-comfortable
- **Secondary:** Renters who can install non-permanent security, small business owners
- **Critical Unmet Needs:**
 - a. **Local storage without cloud fees**—users resent \$10-15/month cloud subscriptions; want SD card or hub-based recording
 - b. **Privacy-first design**—concerns about cameras being hacked or company selling footage; want physical privacy shutter and end-to-end encryption
 - c. **Accurate person/package detection**—false alerts from trees/cars frustrate users; need AI that distinguishes person vs motion

³⁴<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

Market Competitors:

1. **Ring (Amazon)** - Market leader, 10M+ units sold, \$99-\$249 hardware + \$4-\$20/month cloud, ecosystem lock-in
2. **Nest (Google)** - Premium positioning, \$179-\$279 hardware + \$6-\$12/month Nest Aware, Google Home integration
3. **Arlo** - 3M+ customers, \$49-\$199 hardware + \$4.99-\$14.99/month cloud, multi-camera focus
4. **Wyze** - Budget leader, \$25-\$50 hardware + \$1.99-\$9.99/month optional cloud, cost-conscious segment
5. **Eufy (Anker)** - Privacy-focused, \$129-\$349, local storage emphasis, no mandatory subscription

Market Insights: Consumer electronics market growing with smart home adoption as key driver. Positioning strategies:

1. **Subscription-optional model:** Follow Eufy's approach—charge \$149-\$199 for hardware with robust local features, offer optional \$3-5/month cloud backup
2. **Privacy certification:** Get third-party security audits and prominently display certifications—differentiate from Ring/Nest surveillance concerns
3. **Retail + DTC channels:** Launch DTC for margins (50-60% vs 30-40% retail), expand to Home Depot/Best Buy in year 2 for scale
4. **TAM:** 90M+ US households with \$75K+ income, 35% adoption = 31M addressable market × \$175 blended ASP = \$5.4B TAM

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Distinguishes must-have vs hype, purchase criteria, market leaders, bundle model, manufacturing costs, channels, subscriptions; 7 = Must-have features and leaders; 4 = Lists features without priority; 1 = Can't distinguish necessity from hype
 - **Target Users (10 pts):** 10 = Demographics, income, location, personas, 3 unmet needs with storage/privacy/accuracy specifics; 7 = Profile and 2 needs; 4 = "Homeowners" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with units/pricing/positioning and subscriptions; 7 = Lists 3-4 with data; 4 = Lists 2 brands; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Subscription-optional model, privacy certification, channel strategy, TAM sizing; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 3.2: Fitness Wearable Device

Test Case Type: Hardware Product

User Question: "The fitness tracker market feels crowded. What unmet user needs still exist, and is there space for a new entrant?"

Standard Answer:

User Requirements Analysis: The user needs to assess market saturation vs opportunity, identify underserved user segments or use cases, understand what drives sustained usage (not just purchase), and evaluate feasibility for a new brand. Implicit needs include understanding sensor technology differentiation, battery life expectations, and app ecosystem importance.

Target Users:

- **Primary:** Fitness enthusiasts aged 25-45 who already tried 1-2 trackers but stopped using them
- **Secondary:** Medical patients with chronic conditions (diabetes, heart disease) needing monitoring
- **Critical Unmet Needs:**
 - a. **Actionable insights vs raw data**—users get steps/heart rate but not "you should go to bed earlier because your HRV is declining"
 - b. **Interoperability between devices**—users want to switch brands without losing historical data; need open data standards
 - c. **Discrete medical-grade monitoring**—diabetics/cardiac patients want continuous monitoring in a fashionable form factor, not bulky medical devices

Market Competitors:

1. **Apple Watch** - 49% smartwatch market share, \$399-\$799, ecosystem lock-in, premium positioning
2. **Fitbit (Google)** - 5M+ units sold annually, \$79-\$299, mass market fitness focus, being integrated into Google ecosystem
3. **Garmin** - 10M+ wearables sold, \$199-\$999, athlete/outdoor enthusiast positioning
4. **Whoop** - Subscription-first (\$30/month, hardware free), 950K+ members, athlete recovery focus
5. **Oura Ring** - \$299-\$549, sleep tracking focus, 2M+ rings sold, discrete form factor

Market Insights: Consumer electronics market evolving toward health-focused technology. White space opportunities:

1. **Medical-grade with fashion:** Partner with fashion brands for medical-certified continuous glucose monitor or cardiac monitor that looks like jewelry—not Fitbit
2. **Open data platform:** Build device that exports to Apple Health, Google Fit, Strava simultaneously—compete on data freedom vs lock-in
3. **Subscription for insights:** Free device (\$99 cost), charge \$15-25/month for AI coaching based on biometric trends (Whoop model)
4. **SAM focus:** 15M people with diabetes + 30M with heart disease in US who want discrete monitoring = \$6.7B SAM at \$150 blended ASP

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Assesses saturation vs opportunity, underserved segments, sustained usage drivers, new entrant feasibility, sensor tech, battery, app ecosystem; 7 = Opportunity and usage drivers; 4 = Lists missing features; 1 = Misses market dynamics
 - **Target Users (10 pts):** 10 = Age, prior device usage, medical segment, personas, 3 unmet needs with insights/interoperability/medical specifics; 7 = Profile and 2 needs; 4 = "Fitness users" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with market share/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 brands; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Medical-fashion hybrid, open data platform, subscription model, SAM focus; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 3.3: Wireless Earbuds

Test Case Type: Hardware Product

User Question: "What differentiates successful wireless earbuds brands from the hundreds of failed ones, and what features are table stakes vs premium?"

Standard Answer:

User Requirements Analysis: The user needs to understand success factors in a commoditized market, distinguish between table-stakes features and differentiation, identify brand positioning strategies, and understand margin structure. Implicit needs include audio quality expectations across price tiers, retail shelf space competition, and consumer review influence.

Target Users:

- **Primary:** Consumer aged 18-45 who use earbuds daily for music, calls, content consumption

- **Secondary:** Fitness users (sweat resistance), professionals (call quality), commuters (ANC)
- **Critical Unmet Needs:**
 - a. **Multi-device switching**—work laptop, personal phone, tablet; switching is clunky with most brands except Apple
 - b. **All-day comfort**—earbuds hurt after 2-3 hours; need better ergonomics and multiple tip sizes
 - c. **Sustainable design**—batteries die after 2 years, entire earbud becomes e-waste; want replaceable batteries

Market Competitors:

1. **Apple AirPods** - 28% global market share, \$129-\$549, ecosystem integration, premium positioning
2. **Samsung Galaxy Buds** - 10% market share, \$149-\$229, Android optimization, good value
3. **Sony WF series** - Premium audio, \$198-\$298, audiophile positioning, industry-leading ANC
4. **Jabra Elite** - Business/fitness focus, \$79-\$229, call quality emphasis
5. **Anker Soundcore** - Budget leader, \$39-\$149, Amazon-first distribution, cost-conscious segment

Market Insights: Consumer electronics market for audio devices driven by mobile content consumption. Success factors:

1. **Table stakes (all price points):** 6+ hour battery, IPX4 water resistance, touch controls, Bluetooth 5.0+, USB-C charging
2. **Premium differentiators (\$150+):** Active noise cancellation, spatial audio, wireless charging, multi-device pairing, hearing health features
3. **Brand positioning strategies:**
 - **Premium (\$200+):** Audio quality, ANC, brand prestige (Sony, Apple)
 - **Value (\$80-150):** Good features at fair price (Samsung, Jabra)
 - **Budget (\$40-80):** Basic functionality, price-first (Anker, Wyze)
4. **Distribution is critical:** Amazon for budget/value, Apple Stores + carriers for premium = \$28B global market

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies success factors, table stakes vs premium, brand positioning, margins, audio quality tiers, retail competition, review influence; 7 = Success factors and feature tiers; 4 = Lists features without context; 1 = Misses market structure

- **Target Users (10 pts):** 10 = Age, daily usage, use cases, personas, 3 unmet needs with switching/comfort/sustainability specifics; 7 = Profile and 2 needs; 4 = "Audio users" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with market share/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 brands; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Table stakes list, premium differentiators, brand positioning strategies, distribution importance; 7 = Feature tiers and positioning; 4 = Generic; 1 = No insights
-

Test Case 3.4: Smart Thermostat

Test Case Type: Hardware Product

User Question: "Nest and Ecobee dominate smart thermostats. What would a third competitor need to offer to succeed, and who would be their target customer?"

Standard Answer:

User Requirements Analysis: The user needs to understand the competitive moat of incumbents, identify customer segments underserved by Nest/Ecobee, find white space in features or business model, and assess go-to-market feasibility. Implicit needs include HVAC installer relationships, smart home platform integration, and energy utility partnerships.

Target Users:

- **Primary:** Homeowners in 1,500-3,500 sq ft homes with multi-zone HVAC systems
- **Secondary:** Property management companies (rental properties), energy-conscious early adopters
- **Critical Unmet Needs:**
 - a. **Multi-zone optimization**—Nest/Ecobee focus on single-zone; users with zoned systems need per-zone scheduling and balancing
 - b. **HVAC system diagnostics**—warn when filter needs changing, detect inefficiency, predict failures before breakdown
 - c. **Energy tariff optimization**—automatically run AC during off-peak electricity hours for time-of-use plans

Market Competitors:

1. **Nest (Google)** - Market leader, \$129-\$249, learning algorithm, Google Home integration, retail + HVAC installer channels
2. **Ecobee** - #2 player, \$189-\$249, room sensor focus, HomeKit/Alexa support, HVAC partnership focus

3. **Honeywell Home** - Legacy HVAC brand, \$99-\$199, installer relationships, older demographic
4. **Emerson Sensi** - Budget option, \$89-\$169, contractor-friendly installation, traditional HVAC channel
5. **Wyze Thermostat** - Budget disruptor, \$59, basic features, online-only distribution

Market Insights: Smart home device adoption continues with energy costs driving thermostat upgrades. Competitive strategies:

1. **Multi-zone specialist:** Focus marketing on 2-3 zone systems that Nest handles poorly—target 10M+ US homes with zoned HVAC
2. **Installer co-marketing:** Revenue share with HVAC contractors who install/recommend (30% of margin) vs DIY-only—Nest is weak in contractor channel
3. **Utility partnership GTM:** Partner with power companies to subsidize hardware (they pay \$50-100 per unit) in exchange for demand response capabilities
4. **Target market:** Multi-zone homeowners (10M) + property management (5M units) = 15M SAM × \$180 ASP = \$2.7B opportunity

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies incumbent moats, underserved segments, white space features/model, GTM feasibility, installer relationships, platform integration, utility partnerships; 7 = Moats and underserved segments; 4 = Generic "better features"; 1 = Doesn't address competition
 - **Target Users (10 pts):** 10 = Home size, HVAC type, property mgmt segment, personas, 3 unmet needs with multi-zone/diagnostics/tariff specifics; 7 = Profile and 2 needs; 4 = "Homeowners" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with positioning/channels and explains Nest/Ecobee dominance; 7 = Lists 3-4 with data; 4 = Lists 2 brands; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Multi-zone specialist, installer co-marketing, utility partnerships, SAM sizing; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 3.5: Portable Power Station

Test Case Type: Hardware Product

User Question: "What use cases drive purchases of portable power stations, and how do leading brands position themselves?"

Standard Answer:

User Requirements Analysis: The user needs to understand the diverse use cases (emergency, camping, outdoor events, off-grid living), identify which use case segments are largest/most profitable, learn how brands position differently, and understand technical specs that matter. Implicit needs include battery chemistry trade-offs, solar compatibility, and retail vs outdoor specialty channels.

Target Users:

- **Primary:** Homeowners in areas with frequent power outages, outdoor enthusiasts, RV/van life community
- **Secondary:** Contractors/mobile workers, event organizers, preppers
- **Critical Unmet Needs:**
 - a. **Right-sizing guidance**—users don't know if they need 300Wh or 2,000Wh; need interactive calculator for their use case
 - b. **Expandable capacity**—buying 1,500Wh then needing more means buying second unit; want modular battery expansion
 - c. **Smart home integration**—automatically power critical circuits during outage (fridge, wifi) vs running extension cords

Market Competitors:

1. **Goal Zero Yeti** - Premium outdoor brand, \$299-\$2,999, REI/outdoor retail distribution, explorer/camping positioning
2. **Jackery** - Market leader, \$199-\$2,799, Amazon-first, balanced features/price, general-purpose positioning
3. **EcoFlow** - Innovation focus, \$299-\$3,999, fast charging differentiation (80% in 1 hour), tech-forward branding
4. **Bluetti** - Value positioning, \$249-\$3,599, higher capacity per dollar, off-grid/prepper community focus
5. **Anker** - Budget entry, \$149-\$1,099, smaller capacities, camping/tailgating focus

Market Insights: Consumer electronics market growth driven by sustainability and outdoor recreation trends. Positioning strategies by use case:

1. **Emergency backup (largest segment):** Emphasize reliability, UPS-speed switching, solar charging for multi-day outages—partner with electricians for smart panel integration
2. **Outdoor recreation:** Emphasize portability (weight per Wh), ruggedness, outdoor retailer distribution (REI, Bass Pro)
3. **Off-grid living:** Emphasize expandability, long lifespan (3,000+ cycles), high capacity (2,000Wh+), solar ecosystem

4. **Market sizing:** 50M US households in outage-prone areas × 15% adoption × \$800 ASP = \$6B emergency backup TAM

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies use case diversity, segment size/profitability, brand positioning differences, technical specs that matter, battery chemistry, solar, channels; 7 = Use cases and positioning; 4 = Lists use cases without depth; 1 = Single use case focus
 - **Target Users (10 pts):** 10 = Geography, lifestyle segments, personas, 3 unmet needs with sizing/expandability/integration specifics; 7 = Profile and 2 needs; 4 = "Outdoor users" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with pricing/distribution/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 brands; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Use-case-specific positioning strategies (emergency/outdoor/off-grid) with TAM sizing; 7 = Two positioning approaches; 4 = Generic; 1 = No insights
-

Test Case 3.6: E-Ink Tablet for Reading/Writing

Test Case Type: Hardware Product

User Question: "E-ink tablets are niche compared to iPads. What specific user segments find e-ink valuable, and who are the key players?"

Standard Answer:

User Requirements Analysis: The user needs to understand why e-ink remains niche vs LCD/OLED, identify user segments who prioritize e-ink benefits, understand the value proposition vs iPads, and assess market size. Implicit needs include understanding e-ink technology limitations (refresh rate, color), app ecosystem requirements, and price sensitivity.

Target Users:

- **Primary:** Students and academics who read/annotate research papers extensively, authors/writers who want distraction-free writing
- **Secondary:** Business professionals who review contracts/reports, book lovers who want library portability
- **Critical Unmet Needs:**
 - a. **App ecosystem parity**—users want Notability/GoodNotes quality apps, not basic note apps that come with devices
 - b. **Responsive writing latency**—early e-ink styluses had 100ms+ lag; need <20ms for natural writing feel

- c. **Hybrid use cases**—users want to read/annotate PDFs AND sketch diagrams, not just one or other

Market Competitors:

1. **reMarkable** - Premium writing focus, \$399-\$579 + \$7.99/month subscription, distraction-free positioning, 100K+ units sold
2. **Kobo Elipsa** - Reading + writing, \$399, Rakuten/bookstore integration, book lover positioning
3. **Onyx Boox** - Android-based, \$379-\$899, app flexibility, power user focus, largest e-ink display options
4. **Kindle Scribe (Amazon)** - \$339-\$419, Kindle ecosystem integration, reader-first positioning, newest entrant (2022)
5. **SuperNote** - \$398-\$598, long battery life, ceramic nib stylus, note-taking purist focus

Market Insights: E-ink tablets address eye strain and distraction concerns that LCD/OLED can't solve. Segment strategies:

1. **Academic/research segment:** Position as "research workflow device" with reference managers (Zotero, Mendeley), export to LaTeX—target 4M+ graduate students and academics
2. **Subscription-optional model:** ReMarkable's forced subscription frustrates users; offer cloud sync for \$5/month as option, not requirement
3. **Hybrid positioning:** Don't pick "reader" or "writer"—market as best device for "reading + thinking" workflow (read → annotate → synthesize)
4. **Realistic TAM:** E-ink will remain niche; 15M users globally who prioritize eye health and focus × \$450 ASP = \$6.75B total market (vs \$50B+ tablet market)

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Explains niche positioning, identifies value segments, compares to iPad value prop, assesses market size, technology limitations, app ecosystem, price sensitivity; 7 = Segments and value prop; 4 = "E-ink is better" without segments; 1 = Misses niche nature
- **Target Users (10 pts):** 10 = Profession/lifestyle, use case specifics, personas, 3 unmet needs with app/latency/hybrid specifics; 7 = Profile and 2 needs; 4 = "Readers" with 1 need; 1 = Wrong target
- **Market Analysis (10 pts):** 10 = Lists 5 competitors with pricing/positioning/unit estimates; 7 = Lists 3-4 with data; 4 = Lists 2 brands; 1 = No competitors
- **Market Insights (10 pts):** 10 = Academic segment positioning, subscription model, hybrid positioning, realistic TAM with niche acknowledgment; 7 = Two strategies; 4 = Generic; 1 = No insights

Category 4: Software Tools (SaaS)

Test Case 4.1: Design Collaboration Tool

Test Case Type: Software Tools (SaaS)

User Question: "Figma dominates design collaboration. Is there room for competitors, and what would they need to focus on?"

Standard Answer:

User Requirements Analysis: The user needs to assess Figma's competitive moat, identify underserved design workflows or verticals, find differentiation opportunities, and evaluate market entry feasibility. Implicit needs include understanding network effects, pricing pressure, and Adobe's acquisition impact.^{35 36}

Target Users:

- **Primary:** Design teams at product companies with 5-50 designers
- **Secondary:** Marketing teams doing social/web design, agencies serving clients
- **Critical Unmet Needs:**
 - a. **Design systems at scale**—large enterprises struggle to maintain consistency across 100+ designers; need better governance tools
 - b. **Developer handoff quality**—designers make specs, but engineers still need to translate; want code generation that actually works
 - c. **Specialized workflows** (motion design, 3D, print) that Figma handles poorly³⁷

Market Competitors:

1. **Figma (Adobe)** - Market leader, 4M+ users, \$15-\$75/editor/month, collaborative design standard, pending \$20B Adobe acquisition
2. **Sketch** - Legacy Mac-only tool, \$10-\$20/editor/month, 2M+ users, losing share to Figma
3. **Adobe XD** - Part of Creative Cloud, included in \$54.99/month bundle, enterprise Adobe customers

³⁵<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

³⁶<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

³⁷<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

4. **Framer** - Design + code, \$15-\$30/editor/month, interactive prototype focus, developer-friendly
5. **Penpot** - Open source alternative, free self-hosted or \$8/editor/month cloud, privacy-focused teams³⁸

Market Insights: SaaS market growing at 17.9% CAGR with AI integration as key driver.³⁹ Opportunities exist:

1. **Enterprise design systems:** Build for 50+ designer organizations with advanced governance, version control, design API—charge \$100-\$200/editor (vs Figma's \$75)
2. **Vertical specialization:** Focus on motion design (competitor to After Effects) or print design (competitor to InDesign) where Figma is weak
3. **Open-source positioning:** Fork/compete with Penpot for privacy-conscious European market concerned about Adobe ownership
4. **Realistic market:** Design tools = \$8B market, but Figma + Adobe control 70%; focus on \$2.4B underserved segments (enterprise, motion, privacy)

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Assesses Figma moat, identifies underserved workflows/verticals, differentiation opportunities, entry feasibility, network effects, pricing, Adobe impact; 7 = Moat and opportunities; 4 = "Build better Figma"; 1 = Ignores Figma dominance
- **Target Users (10 pts):** 10 = Team size, company type, use cases, personas, 3 unmet needs with systems/handoff/specialized specifics; 7 = Profile and 2 needs; 4 = "Designers" with 1 need; 1 = Wrong target
- **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/positioning and Adobe context; 7 = Lists 3-4 with data; 4 = Lists 2 tools; 1 = No competitors
- **Market Insights (10 pts):** 10 = Enterprise systems focus, vertical specialization, open-source, realistic market sizing; 7 = Two strategies; 4 = Generic; 1 = No insights

Test Case 4.2: API Development Platform

Test Case Type: Software Tools (SaaS)

³⁸<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

³⁹<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

User Question: "What pain points do developers experience when building and testing APIs, and which tools address these best?"

Standard Answer:

User Requirements Analysis: The user needs to understand the API development lifecycle pain points, identify which stages have the most acute problems, map tool landscape by development stage, and understand integration patterns. Implicit needs include understanding REST vs GraphQL vs gRPC workflows, team collaboration requirements, and CI/CD integration.^{40 41}

Target Users:

- **Primary:** Backend developers and API engineers at SaaS companies with 10-200 engineers
- **Secondary:** DevOps engineers, platform teams, technical product managers
- **Critical Unmet Needs:**
 - a. **API design collaboration**—frontend and backend engineers design APIs in isolation, causing misalignment; need shared spec-first workflow
 - b. **Realistic testing environments**—mocking doesn't catch real issues; need production-like staging with realistic data
 - c. **Breaking change detection**—API changes break client applications; need automated compatibility testing⁴²

Market Competitors:

1. **Postman** - Market leader, 25M+ users, free + \$14-\$49/user/month, API testing/documentation standard
2. **Insomnia** - Kong product, 1M+ users, free + \$5-\$15/user/month, lightweight Postman alternative
3. **Swagger/OpenAPI** - Open standard + tools, Smartbear stewarded, free + enterprise tooling
4. **Apollo GraphQL** - GraphQL platform, \$150-\$400/month + usage, specialized for GraphQL workflows

⁴⁰<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁴¹<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

⁴²<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

5. **RapidAPI** - API marketplace + testing, 4M+ developers, free + \$50-\$500/month, API discovery focus⁴³

Market Insights: SaaS spending on developer tools growing as API economy expands.⁴⁴ Positioning strategies:

1. **Spec-first workflow:** Build collaborative API design tool (OpenAPI editor) before any code—compete with async design process
2. **Contract testing platform:** Focus on breaking change detection with automated compatibility tests—complement Postman rather than replace
3. **GraphQL vs REST positioning:** Most tools default REST; position as "GraphQL-native" platform for teams adopting GraphQL
4. **Market sizing:** 30M+ developers building APIs × 20% paid tool adoption × \$300 ARPU = \$1.8B TAM for API development platforms

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies lifecycle pain points, acute problem stages, tool landscape mapping, integration patterns, REST/GraphQL/gRPC, collaboration, CI/CD; 7 = Pain points and tools; 4 = Generic API testing needs; 1 = Surface-level understanding
 - **Target Users (10 pts):** 10 = Company size, team size, personas, 3 unmet needs with collaboration/testing/breaking change specifics; 7 = Profile and 2 needs; 4 = "Developers" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 tools; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Spec-first workflow, contract testing, GraphQL positioning, TAM sizing; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 4.3: Customer Data Platform (CDP)

Test Case Type: Software Tools (SaaS)

User Question: "CDPs promise unified customer data. Why do many implementations fail, and what differentiates successful platforms?"

Standard Answer:

⁴³<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

⁴⁴<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

User Requirements Analysis: The user needs to understand common CDP implementation failure patterns, identify success factors that separate winners from losers, learn what "unified customer data" actually means in practice, and understand organizational readiness requirements. Implicit needs include data governance, identity resolution quality, and activation capabilities.^{45 46}

Target Users:

- **Primary:** Marketing operations teams at B2C companies with \$50M-\$500M revenue and multiple customer touchpoints
- **Secondary:** Data engineers, CRM managers, growth teams
- **Critical Unmet Needs:**
 - a. **Pre-built identity resolution**—matching "john.smith@gmail" with "j.smith@gmail" across 10 systems is hard; need ML-powered resolution out-of-box
 - b. **Activation integrations**—collecting data is 30% of value; need 100+ integrations with ad platforms, email tools, personalization engines
 - c. **Data governance tools**—GDPR/CCPA require deletion workflows; most CDPs make this manual⁴⁷

Market Competitors:

1. **Segment (Twilio)** - Market leader, 20,000+ customers, \$120-\$2,000+/month, developer-friendly, Twilio ecosystem
2. **mParticle** - Enterprise focus, 500+ customers, \$30K-\$500K/year, mobile app focus, strict data quality
3. **Rudderstack** - Open-source alternative, free self-hosted + \$500+/month cloud, developer control, warehouse-native
4. **Lytics** - Activation focus, 300+ customers, \$2,000-\$10,000/month, marketer-friendly UI
5. **Treasure Data (Arm)** - Enterprise CDP, \$100K+/year contracts, Japan strong, global expansion⁴⁸

⁴⁵<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁴⁶<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

⁴⁷<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁴⁸<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

Market Insights: AI integration and real-time data processing driving CDP evolution.⁴⁹ Success vs failure factors:

1. **Implementation failures stem from:** Treating CDP as "data warehouse replacement" instead of activation layer; poor data hygiene before CDP deployment; lack of use case definition
2. **Start with activation:** Pick 2-3 activation use cases (email personalization, ad suppression) before collecting all data—prove value fast
3. **Warehouse-native architecture:** Build on customer's Snowflake/BigQuery instead of proprietary database—reduces vendor lock-in fear
4. **Market positioning:** B2C companies with \$50M-\$500M revenue (75,000 globally) × \$50K ACV = \$3.75B SAM for mid-market CDP

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies failure patterns, success factors, unified data definition, organizational readiness, data governance, identity resolution, activation; 7 = Failures and success factors; 4 = "Integration is hard"; 1 = Surface understanding
 - **Target Users (10 pts):** 10 = Company type, revenue range, touchpoint complexity, personas, 3 unmet needs with identity/activation/governance specifics; 7 = Profile and 2 needs; 4 = "Marketing teams" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with customers/pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists 2 platforms; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Failure causes, activation-first approach, warehouse-native, SAM positioning; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 4.4: Video Conferencing Software

Test Case Type: Software Tools (SaaS)

User Question: "Zoom and Teams dominate video conferencing. What niche use cases or verticals could a new entrant target?"

Standard Answer:

User Requirements Analysis: The user needs to assess horizontal platform dominance, identify underserved vertical markets or specialized use cases,

⁴⁹<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

understand why vertical-specific solutions can compete with horizontal giants, and evaluate GTM approaches. Implicit needs include understanding video quality requirements by use case, compliance needs, and enterprise sales dynamics.

Target Users (by vertical):

- **Telemedicine:** Healthcare providers needing HIPAA-compliant video with EHR integration
- **Online education:** Teachers needing engagement tools (whiteboards, breakout rooms, attention tracking) beyond Zoom
- **Customer support:** Support teams needing video + screen control + CRM integration
- **Critical Unmet Needs:**
 - a. **Compliance-first design**—HIPAA/GDPR by default, not afterthought; Zoom/Teams require enterprise plans for compliance
 - b. **Vertical workflow integration**—telemedicine needs Epic/Cerner integration; education needs LMS integration; support needs Zendesk integration
 - c. **Specialized features**—telemedicine needs exam tools; education needs engagement analytics; support needs session recording/transcription

Market Competitors:

1. **Zoom** - 300M+ daily users, \$149-\$250/user/year, horizontal leader
2. **Microsoft Teams** - 320M+ monthly active users, included in M365, enterprise standard
3. **Doxy.me** - Telemedicine specialist, 750K+ providers, \$35-\$99/provider/month, HIPAA focus
4. **Whereby** - Embeddable video, \$9.99-\$59.99/host/month, developer/product embedding use case
5. **Around** - Designed for remote teams, \$8-\$12/user/month, "camera on" fatigue solution

Market Insights: Video conferencing market matured during pandemic but vertical opportunities remain. Vertical strategies:

1. **Telemedicine positioning:** Partner with EHR vendors (Epic, Cerner, Athenahealth) for native integration—become default video layer in EHR workflow
2. **Education positioning:** Build for K-12 specifically (not higher ed)—integrate with Google Classroom, Canvas—sell through school district procurement
3. **Pricing strategy:** Charge 2-3× horizontal platforms (\$25-40/user/month) but justify with vertical features and compliance

4. Vertical TAM examples:

- Telemedicine: 1M+ US healthcare providers × \$600 ARPU = \$600M
- K-12 education: 3.7M teachers × \$300 ARPU = \$1.1B
- Customer support: 2M support agents × \$400 ARPU = \$800M

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Assesses horizontal dominance, identifies underserved verticals/use cases, explains vertical competitive advantage, GTM approaches, video quality by use case, compliance, enterprise sales; 7 = Dominance and verticals; 4 = "Better features"; 1 = Ignores Zoom/Teams
 - **Target Users (10 pts):** 10 = Specifies 3 vertical markets, personas in each, 3 unmet needs with compliance/integration/specialized features; 7 = 2 verticals and 2 needs; 4 = "Healthcare" with 1 need; 1 = Generic users
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors including vertical specialists with users/pricing; 7 = Lists 3-4 with data; 4 = Lists Zoom/Teams only; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Vertical positioning strategies for each (telemedicine/education/support) with TAM sizing; 7 = Two vertical strategies; 4 = Generic; 1 = No insights
-

Test Case 4.5: No-Code Automation Platform

Test Case Type: Software Tools (SaaS)

User Question: "No-code automation tools like Zapier are popular. What features do power users wish these tools had, and who are the emerging competitors?"

Standard Answer:

User Requirements Analysis: The user needs to understand power user limitations with current no-code tools, identify the "ceiling" where users outgrow no-code, find opportunities to raise that ceiling, and map the competitive landscape including emerging players. Implicit needs include understanding when users need to migrate to code, collaboration features for teams, and pricing model evolution.^{50 51}

Target Users:

⁵⁰<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁵¹<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

- **Primary:** Operations/RevOps professionals at growth-stage companies (50-500 employees) managing complex workflows
- **Secondary:** Product managers, growth teams, citizen developers
- **Critical Unmet Needs:**
 - a. **Conditional logic complexity**—nested if/then/else is clunky in visual UI; need scripting escape hatch within no-code tool
 - b. **Team collaboration**—workflows are creator-owned, not team-owned; lack version control, testing environments, approval workflows
 - c. **Error handling**—automations fail silently or flood email; need monitoring dashboard with alerting and retry logic⁵²

Market Competitors:

1. **Zapier** - Market leader, 2M+ businesses, \$19.99-\$799+/month, 6,000+ app integrations, SMB focus
2. **Make (Integromat)** - Visual workflow, \$9-\$299+/month, 1,500+ apps, European strong
3. **n8n** - Open-source automation, free self-hosted + \$20-\$100/month cloud, developer-friendly, extensible
4. **Workato** - Enterprise automation, \$10K-\$100K+/year, complex workflows, IT/Enterprise focus
5. **Tray.io** - Enterprise iPaaS, \$30K+/year, API-first, technical teams⁵³

Market Insights: Low-code/no-code driving SaaS adoption among non-technical users.⁵⁴ Power user needs:

1. **Low-code hybrid:** Offer JavaScript/Python code steps within visual workflow—bridge no-code and code (n8n model)
2. **Team features:** Add version control (Git-like), workflow testing environments, approval gates for production deploys—treat automation like engineering
3. **Observability built-in:** Dashboard showing all workflow runs, success rates, latency, errors with Slack/PagerDuty alerting
4. **Market positioning:** 500K+ operations professionals at growth companies × \$2,400 ARPU = \$1.2B TAM for power-user-focused automation tools

Scoring Rubric:

⁵²<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁵³<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

⁵⁴<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

- **User Requirements (10 pts):** 10 = Identifies power user limits, no-code ceiling, ceiling-raising opportunities, emerging competitors, code migration point, collaboration, pricing evolution; 7 = Power user limits and opportunities; 4 = Generic automation needs; 1 = Beginner focus only
 - **Target Users (10 pts):** 10 = Role, company stage, workflow complexity, personas, 3 unmet needs with logic/collaboration/error handling specifics; 7 = Profile and 2 needs; 4 = "No-code users" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors including emerging players with pricing/positioning; 7 = Lists 3-4 with data; 4 = Lists Zapier + 1; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Low-code hybrid, team features, observability, TAM positioning; 7 = Two enhancements; 4 = Generic; 1 = No insights
-

Test Case 4.6: Freelancer/Agency Project Management

Test Case Type: Software Tools (SaaS)

User Question: "Project management tools are everywhere, but freelancers and small agencies have unique needs. What are they, and which tools serve this market?"

Standard Answer:

User Requirements Analysis: The user needs to understand how freelancer/agency needs differ from product company PM tools, identify specific pain points for client services businesses, map tools built for this market, and understand the buying process. Implicit needs include client collaboration requirements, time tracking/invoicing integration, and portfolio vs project thinking.⁵⁵

Target Users:

- **Primary:** Freelancers and agencies (2-20 people) doing client services work (design, development, marketing, consulting)
- **Secondary:** Solo consultants, creative studios
- **Critical Unmet Needs:**
 - a. **Client collaboration without revealing internal work**—clients need visibility but not into internal tasks, time tracking, profitability
 - b. **Integrated time tracking → invoicing**—manually copying tracked hours to invoices is error-prone; need automatic invoice generation

⁵⁵<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

- c. **Multi-project resource allocation**—agencies juggle 5-15 client projects simultaneously; need capacity planning across projects⁵⁶

Market Competitors:

1. **Monday.com** - 180K+ customers, \$8-\$16+/user/month, general PM tool with agency use
2. **ClickUp** - 2M+ teams, \$5-\$12/user/month, feature-rich, some agency adoption
3. **Productive.io** - Agency-specific, \$9-\$24/user/month, 10K+ agencies, time/budget/resource focus
4. **Teamwork** - Agency positioning, \$10-\$18/user/month, 20K+ customers, client portal emphasis
5. **Bonsai** - Freelancer focus, \$17-\$39/month (flat), 1M+ freelancers, contracts + PM + invoicing⁵⁷

Market Insights: SaaS market expanding in SMB segment with affordable tools.⁵⁸

⁵⁹ Freelancer/agency-specific strategies:

1. **Client portal built-in:** Let clients see their project without agency workspace access—Monday/ClickUp make this add-on, build it core
2. **Services business metrics:** Track utilization rate, project profitability, billable vs non-billable time—not story points or velocity
3. **Flat pricing for small teams:** Charge \$49-\$99/month for 1-5 users instead of per-seat—align with small agency budgets
4. **Market sizing:**
 - US freelancers: 60M (2023) × 30% using PM tools × \$300 ARPU = \$5.4B
 - US agencies <20 people: 250K × \$2,000 ARPU = \$500M
 - Combined TAM: \$5.9B for freelancer/small agency PM tools

Scoring Rubric:

⁵⁶<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁵⁷<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁵⁸<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

⁵⁹<https://www.mordorintelligence.com/industry-reports/software-as-a-service-market>

- **User Requirements (10 pts):** 10 = Identifies how agency needs differ from product companies, client services pain points, tools for market, buying process, client collaboration, time/invoicing, portfolio vs project; 7 = Different needs and pain points; 4 = Generic PM needs; 1 = Doesn't distinguish from product companies
 - **Target Users (10 pts):** 10 = Business type, team size, service types, personas, 3 unmet needs with client visibility/invoicing/resource allocation specifics; 7 = Profile and 2 needs; 4 = "Freelancers" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with customers/pricing/positioning mix of general + specialized; 7 = Lists 3-4 with data; 4 = Lists 2 tools; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Client portal, services metrics, flat pricing, TAM sizing (freelancer + agency); 7 = Two strategies; 4 = Generic; 1 = No insights
-

Category 5: Consumer Mobile Apps

Test Case 5.1: Language Learning App

Test Case Type: Consumer Mobile App

User Question: "Duolingo dominates language learning apps. What do users complain about, and where are the opportunities for a new app?"

Standard Answer:

User Requirements Analysis: The user needs to understand Duolingo's weaknesses despite market dominance, identify frustrated user segments, find feature gaps or methodology differences, and assess new entrant viability. Implicit needs include understanding freemium conversion dynamics, streak psychology, and proficiency measurement.⁶⁰

Target Users:

- **Primary:** Adults aged 25-45 learning languages for career advancement or relocation (not casual/hobby learners)
- **Secondary:** Students preparing for language proficiency tests (TOEFL, DELE, JLPT)
- **Critical Unmet Needs:**

⁶⁰<https://www.thebusinessresearchcompany.com/report/software-as-a-service-global-market-report>

- a. **Conversational practice with real humans**—Duolingo is great for vocabulary/grammar but users plateau without speaking practice
- b. **Business/professional vocabulary**—Duolingo teaches "the cat drinks milk" but not "let's schedule a meeting" or industry jargon
- c. **Structured curriculum for tests**—users preparing for DELE/JLPT need test-specific prep, not general fluency

Market Competitors:

1. **Duolingo** - 80M+ monthly active users, freemium + \$12.99/month premium, gamification leader, IPO valuation \$3.7B
2. **Babbel** - 10M+ subscriptions, \$13.95/month, structured lessons, European focus, premium-only
3. **Rosetta Stone** - 25M+ lifetime users, \$11.99-\$15.99/month, immersive method, legacy brand
4. **Busuu** - 120M+ users, \$9.99-\$13.99/month, social learning, McGraw-Hill partnership
5. **Pimsleur** - Audio-first method, \$14.99-\$20.99/month, driving/commute focus

Market Insights: Mobile app market at \$156B consumer spending with subscription models thriving. Opportunities:

1. **Hybrid AI + human tutors:** Offer unlimited AI practice (Duolingo model) + 4 live tutor sessions/month for \$29.99—Duolingo lacks human interaction
2. **Vertical language learning:** Build for healthcare Spanish, tech English, business Mandarin vs general fluency—target professionals with \$40-60/month willingness
3. **Test prep positioning:** Partner with test bodies (DELE, JLPT, TOEFL) for official prep app—premium pricing \$49.99/month for 3-month prep
4. **Target market:** 1B+ people learning language for career × 2% paid app adoption × \$25 ARPU = \$500M TAM for professional-focused language apps

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies Duolingo weaknesses despite dominance, frustrated segments, feature gaps/methodology, new entrant viability, freemium dynamics, streak psychology, proficiency measurement; 7 = Weaknesses and opportunities; 4 = "Add more features"; 1 = Praises Duolingo without critique
- **Target Users (10 pts):** 10 = Age, motivation (career/relocation), test prep segment, personas, 3 unmet needs with conversation/professional/curriculum specifics; 7 = Profile and 2 needs; 4 = "Language learners" with 1 need; 1 = Wrong target

- **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/methodology and Duolingo's dominance context; 7 = Lists 3-4 with data; 4 = Lists 2 apps; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Hybrid AI+human, vertical language, test prep partnership, TAM for professional segment; 7 = Two strategies; 4 = Generic; 1 = No insights
-

Test Case 5.2: Personal Finance Budgeting App

Test Case Type: Consumer Mobile App

User Question: "Many budgeting apps exist but users stop using them after a few months. What causes abandonment, and how can a new app drive sustained engagement?"

Standard Answer:

User Requirements Analysis: The user needs to understand app abandonment root causes, differentiate between activation and retention challenges, identify what drives long-term engagement, and learn from both failed and successful apps. Implicit needs include understanding budgeting methodology differences (zero-based vs envelope vs tracking), data aggregation reliability, and behavioral psychology.

Target Users:

- **Primary:** Millennials/Gen Z aged 25-35 with \$40K-\$80K income trying to save for specific goals (house, debt payoff, emergency fund)
- **Secondary:** Couples managing joint finances, parents tracking family budgets
- **Critical Unmet Needs:**
 - a. **Goal-driven design**—existing apps show spending categories but users don't connect that to "I want to buy a house in 2027"
 - b. **Partner collaboration**—couples fight about money but most apps are single-user; need shared budgets with communication tools
 - c. **Forgiveness for imperfection**—users blow their budget and abandon app in shame; need positive reinforcement for getting back on track

Market Competitors:

1. **Mint (Intuit)** - 20M+ users, free ad-supported, being shut down Dec 2023 and migrated to Credit Karma
2. **YNAB (You Need A Budget)** - \$99/year, zero-based budgeting methodology, 500K+ users, high retention

3. **Monarch Money** - \$99-\$165/year, couples focus, 100K+ users, Mint replacement positioning
4. **PocketGuard** - Freemium + \$7.99/month, "in my pocket" concept, automated budgeting
5. **EveryDollar** - Ramsey Solutions product, freemium + \$79.99/year, Dave Ramsey methodology tie-in

Market Insights: Consumer spending on finance apps growing but retention is challenging. Retention strategies:

1. **Goal visualization:** Start with "what do you want to achieve?" not "link your bank" —show progress toward goal daily (saved \$1,250 of \$30,000 down payment)
2. **Relationship features:** Shared budgets, spending transparency, financial goal discussions—market to couples explicitly (\$80-120/year for 2-person plan)
3. **Behavioral nudges:** Send encouragement when user goes over budget, celebrate small wins, share "got back on track" stories from community
4. **Methodology flexibility:** Don't force zero-based budgeting on everyone—support multiple approaches with quiz to recommend best fit
5. **Market sizing:** 80M US households with \$40K-\$150K income × 15% paid app adoption × \$100 ARPU = \$1.2B TAM

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies abandonment causes, activation vs retention, long-term engagement drivers, success/failure examples, methodology differences, aggregation reliability, behavioral psychology; 7 = Abandonment and engagement; 4 = "Make it easier to use"; 1 = Surface understanding
- **Target Users (10 pts):** 10 = Age, income, specific goals, couples segment, personas, 3 unmet needs with goal-driven/collaboration/forgiveness specifics; 7 = Profile and 2 needs; 4 = "Budget users" with 1 need; 1 = Wrong target
- **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/business model and Mint shutdown context; 7 = Lists 3-4 with data; 4 = Lists 2 apps; 1 = No competitors
- **Market Insights (10 pts):** 10 = Goal visualization, relationship features, behavioral nudges, methodology flexibility, TAM sizing; 7 = Three strategies; 4 = Generic; 1 = No insights

Test Case 5.3: Meditation and Wellness App

Test Case Type: Consumer Mobile App

User Question: "The meditation app market is crowded with Calm and Headspace. What wellness areas are underserved, and what do users want that these apps don't offer?"

Standard Answer:

User Requirements Analysis: The user needs to assess meditation app market saturation, identify underserved wellness sub-categories, understand user unmet needs within meditation, and explore adjacent wellness opportunities. Implicit needs include content creation costs, instructor/creator economics, and measuring wellness outcomes.

Target Users:

- **Primary:** Stressed professionals aged 28-45 with anxiety/burnout symptoms, interested in holistic wellness beyond meditation
- **Secondary:** People with clinical anxiety/depression (adjunct to therapy), parents managing stress
- **Critical Unmet Needs:**
 - a. **Personalized content paths**—Calm/Headspace offer libraries but not adaptive paths based on user progress and mood patterns
 - b. **Integration with therapy**—users in therapy want to share app insights with therapist; need therapist dashboard and progress reports
 - c. **Beyond meditation**—users want breathwork, somatic exercises, sleep, nutrition in one app rather than subscribing to 4 apps

Market Competitors:

1. **Calm** - 100M+ downloads, \$14.99/month or \$69.99/year, content library focus, \$2B+ valuation
2. **Headspace** - 70M+ downloads, \$12.99/month or \$69.99/year, guided meditation, merged with Ginger (mental health)
3. **Insight Timer** - 25M+ users, freemium + \$9.99/month, teacher marketplace, community focus
4. **Balance** - Personalized meditation, \$11.99/month, AI-driven content selection
5. **Ten Percent Happier** - \$99.99/year, meditation for skeptics positioning, podcast tie-in

Market Insights: Mobile wellness apps growing as mental health awareness increases. Underserved opportunities:

1. **Therapy integration positioning:** Partner with therapists to offer app as "homework between sessions"—B2B2C model where therapists prescribe app (\$15/month user, \$5 commission to therapist)

2. **Holistic wellness bundle:** Combine meditation + breathwork + sleep + micro-workouts + nutrition guidance—charge \$24.99/month vs \$14.99 meditation-only
3. **Men's mental health:** Most wellness apps skew 70% female; build for men explicitly with different language, aesthetics, and content (stoicism, purpose, performance)
4. **Clinical outcomes tracking:** Add PHQ-9/GAD-7 screeners, show anxiety score improvements over time—differentiate from "feel-good" consumer apps
5. **Market sizing:** 100M+ US adults with anxiety/depression × 5% paid app adoption × \$150 ARPU = \$750M TAM for clinical-adjacent wellness apps

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Assesses saturation, identifies underserved sub-categories, unmet needs in meditation, adjacent opportunities, content costs, creator economics, outcome measurement; 7 = Saturation and opportunities; 4 = "Better meditation content"; 1 = Ignores crowded market
 - **Target Users (10 pts):** 10 = Age, stress profile, clinical vs non-clinical segment, personas, 3 unmet needs with personalization/therapy/holistic specifics; 7 = Profile and 2 needs; 4 = "Stressed people" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with downloads/pricing/positioning and market concentration; 7 = Lists 3-4 with data; 4 = Lists Calm/Headspace only; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Therapy integration, holistic bundle, men's positioning, clinical outcomes, TAM sizing; 7 = Three strategies; 4 = Generic; 1 = No insights
-

Test Case 5.4: Recipe and Meal Planning App

Test Case Type: Consumer Mobile App

User Question: "What frustrates users about existing recipe and meal planning apps, and what would make them worth paying for?"

Standard Answer:

User Requirements Analysis: The user needs to understand user pain points with current recipe apps, identify the gap between free and paid value propositions, understand what drives monetization in this category, and learn willingness-to-pay drivers. Implicit needs include understanding content sourcing (user-generated vs editorial vs creator), grocery integration, and dietary restriction handling.

Target Users:

- **Primary:** Home cooks aged 30-50 cooking 5-7 dinners/week for families, interested in healthy eating but time-constrained
- **Secondary:** Health-conscious individuals with dietary restrictions (keto, vegan, gluten-free), meal preppers
- **Critical Unmet Needs:**
 - a. **Meal planning that considers household reality**—apps suggest recipes but don't account for ingredients already in fridge, schedule constraints, kid preferences
 - b. **Accurate grocery list generation**—lists include wrong quantities or assume pantry staples; users waste time editing lists
 - c. **Recipe quality variability**—user-generated recipes lack testing; editorial recipes are too complex; need middle ground

Market Competitors:

1. **Yummly** - 25M+ users, free + \$4.99/month premium, recipe aggregation, personalization focus
2. **Mealime** - 4M+ downloads, free + \$6.99/month, meal plan + grocery list automation
3. **Paprika** - \$4.99 one-time purchase, recipe manager + grocery list, organization focus, no subscription
4. **PlateJoy** - \$12.99/month, custom meal plans, nutritionist-designed, weight loss/health focus
5. **Whisk** - Free recipe manager, grocery delivery integrations (Instacart, Amazon Fresh), Samsung partnership

Market Insights: Consumer app monetization challenging in crowded categories; must deliver clear paid value. Monetization drivers:

1. **Household context awareness:** Ask about family size, dietary restrictions, cooking skill, schedule—then suggest meals that fit all constraints (not just search results)
2. **Smart grocery integration:** Don't just list ingredients—integrate with Instacart/Amazon Fresh for 1-click ordering, or Kroger/Safeway for pickup
3. **Recipe quality curation:** Commission food bloggers for tested family-friendly recipes—charge \$9.99/month for curated library vs free user-generated chaos
4. **Pricing tiers:**
 - Free: Basic recipe saving
 - \$6.99/month: Meal planning + grocery lists
 - \$14.99/month: Curated recipes + nutrition tracking + grocery delivery integration

5. **Market sizing:** 80M US households cooking regularly × 10% paid adoption × \$100 ARPU = \$800M TAM

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies pain points, free vs paid value gap, monetization drivers, WTP drivers, content sourcing, grocery integration, dietary restrictions; 7 = Pain points and monetization; 4 = "Better recipes"; 1 = Doesn't address WTP
 - **Target Users (10 pts):** 10 = Age, cooking frequency, family situation, dietary segment, personas, 3 unmet needs with household reality/grocery lists/quality specifics; 7 = Profile and 2 needs; 4 = "Home cooks" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/monetization models; 7 = Lists 3-4 with data; 4 = Lists 2 apps; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Household context, grocery integration, quality curation, pricing tiers, TAM sizing; 7 = Three strategies; 4 = Generic; 1 = No insights
-

Test Case 5.5: Social Fitness App

Test Case Type: Consumer Mobile App

User Question: "Fitness apps often lack social features beyond leaderboards. What social mechanics would actually motivate users to work out more consistently?"

Standard Answer:

User Requirements Analysis: The user needs to understand limitations of current fitness app social features, identify what social mechanics drive behavior change (not just engagement), learn from gaming/social app psychology, and understand accountability vs competition dynamics. Implicit needs include understanding workout tracking data structure, privacy concerns in fitness sharing, and monetization of social features.

Target Users:

- **Primary:** Fitness beginners/intermediates aged 25-40 who struggle with consistency and need external motivation
- **Secondary:** People who lost workout partners post-pandemic, parents with limited gym time
- **Critical Unmet Needs:**

- a. **Accountability partnerships**—not leaderboards with strangers, but paired commitment with friend/family where both parties check in on each other
- b. **Micro-social interactions**—current apps require posting full workout; users want quick "finished my run!" with emoji reactions, not social media pressure
- c. **Group challenges with meaningful stakes**—leaderboards don't motivate; need team-based challenges where letting down the team matters

Market Competitors:

1. **Strava** - 100M+ users, freemium + \$11.99/month, athlete social network, segment leaderboards, cycling/running focus
2. **Peloton App** - 6.6M+ subscribers, \$12.99/month, classes + social features, declining from pandemic peak
3. **Nike Run Club** - 100M+ downloads, free, run tracking + challenges, Nike ecosystem
4. **MapMyRun (Under Armour)** - 60M+ users, freemium + \$5.99/month, route tracking, casual runner focus
5. **Fitbit Premium** - Part of Fitbit, \$9.99/month, device data + challenges, Google ecosystem

Market Insights: Social features can drive retention but must align with user psychology. Effective social mechanics:

1. **Accountability pods (2-4 people):** Let users form small pods with commitments ("work out 3x/week"); pod members get notifications if someone misses goal—intimate accountability vs public leaderboard
2. **Asynchronous challenges:** Not "who runs fastest" but "complete these 5 workouts this week"—team succeeds together, accommodates different fitness levels and schedules
3. **Celebration over competition:** Default to celebrating others' workouts (easy emoji reactions) rather than comparing performance—reduce intimidation for beginners
4. **Privacy-first sharing:** Let users share "I worked out" without revealing specific metrics unless they choose—many users don't want to share weight, pace, etc.
5. **Monetization:** Free social features + \$9.99/month for AI-generated workout plans and nutrition guidance
6. **Market positioning:** 150M+ people with gym memberships/fitness intent × 15% app adoption × \$80 ARPU = \$1.8B TAM for social fitness apps

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Identifies current social limitations, behavior-change mechanics vs engagement, gaming psychology application, accountability vs competition, workout data structure, privacy, monetization; 7 = Limitations and mechanics; 4 = "Add leaderboards"; 1 = Surface understanding
 - **Target Users (10 pts):** 10 = Fitness level, age, consistency challenges, post-pandemic context, personas, 3 unmet needs with accountability/micro-social/group stakes specifics; 7 = Profile and 2 needs; 4 = "Fitness users" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/social features; 7 = Lists 3-4 with data; 4 = Lists 2 apps; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Accountability pods, asynchronous challenges, celebration focus, privacy-first, monetization, TAM positioning; 7 = Three mechanics; 4 = Generic; 1 = No insights
-

Test Case 5.6: Habit Tracking App

Test Case Type: Consumer Mobile App

User Question: "Dozens of habit tracking apps exist. What causes users to abandon them, and how can a new app help users actually form lasting habits?"

Standard Answer:

User Requirements Analysis: The user needs to understand the gap between tracking habits and forming habits, identify abandonment patterns (when and why), learn what behavioral science says about habit formation, and differentiate between features and outcomes. Implicit needs include understanding streak psychology, notification fatigue, and when apps should "graduate" users.

Target Users:

- **Primary:** Self-improvement enthusiasts aged 25-40 who've tried 2-3 habit apps before but stopped using them
- **Secondary:** People with ADHD/executive function challenges who need external structure
- **Critical Unmet Needs:**
 - a. **Habit stacking and triggers**—apps track individual habits in isolation; don't help users attach new habits to existing routines ("after coffee, meditate")
 - b. **Gradual difficulty progression**—users try to meditate 30 min/day on day 1 and fail; need smart ramping (start with 2 min, increase 1 min/week)

- c. **Identity reinforcement over tracking**—apps focus on "you did it!" not "you're becoming a person who does this"—identity shift is key to lasting change

Market Competitors:

1. **Habitica** - 8M+ users, freemium + \$4.99/month, gamification via RPG mechanics, quest-based
2. **Streaks** - \$4.99 one-time purchase, iOS-only, 24-habit limit, simple design, no subscription
3. **Habitify** - 2M+ downloads, free + \$4.99/month, routine focus, time-based habits
4. **Loop Habit Tracker** - Open source, free, Android-only, simple tracking, privacy-focused
5. **Done** - Freemium + \$2.99/month, streak-focused, minimal design, Apple platform exclusive

Market Insights: Habit apps have high initial adoption but 90% abandon within 3 months. Lasting habit formation strategies:

1. **Implementation intentions:** Force users to define "when/where/how" not just "what"—research shows implementation intentions 2-3× effectiveness
2. **Adaptive coaching:** Use AI to detect when user is struggling (3 missed days) and suggest making habit easier or changing trigger time—not just "try harder"
3. **Identity milestones:** After 30 days of meditation, unlock "meditator" identity badge and shift language from "do meditation" to "you're a meditator"—self-perception shift
4. **Graduation mechanism:** After 90 days of successful habit, offer to "graduate" habit to autopilot and free up focus for new habit—signal success
5. **Community accountability:** Small groups (5-10 people) working on same habit category with weekly reflection threads
6. **Pricing:** Free for 3 habits, \$6.99/month unlimited + coaching features
7. **Market sizing:** 200M+ people interested in habit formation globally × 5% paid adoption × \$60 ARPU = \$600M TAM

Scoring Rubric:

- **User Requirements (10 pts):** 10 = Distinguishes tracking from forming, identifies abandonment patterns, applies behavioral science, features vs outcomes, streak psychology, notification fatigue, graduation concept; 7 = Abandonment and science; 4 = "Better reminders"; 1 = Focuses only on tracking

- **Target Users (10 pts):** 10 = Age, prior app usage, ADHD segment, personas, 3 unmet needs with stacking/progression/identity specifics; 7 = Profile and 2 needs; 4 = "Self-improvement people" with 1 need; 1 = Wrong target
 - **Market Analysis (10 pts):** 10 = Lists 5 competitors with users/pricing/core mechanic and abandonment context; 7 = Lists 3-4 with data; 4 = Lists 2 apps; 1 = No competitors
 - **Market Insights (10 pts):** 10 = Implementation intentions, adaptive coaching, identity milestones, graduation, community, pricing, TAM sizing; 7 = Four strategies; 4 = Generic; 1 = No insights
-

Document Summary

This test suite provides 30 comprehensive test cases across 5 product categories, each with:

- Realistic product manager queries
- Detailed standard answers demonstrating ideal Agent output
- 4-dimensional scoring rubrics (User Requirements, Target Users, Market Analysis, Market Insights)
- Specific criteria for 10, 7, 4, and 1-point scores in each dimension

The test cases draw on current market data (2025-2026), include specific competitors with market positioning, and provide actionable insights that a product manager would need for market research.