

REFINERY

Where Prose Becomes Perfect

REQUIREMENTS DEFINITION DOCUMENT

All Three Product Lines

Indie Author | Academic | Enterprise Publishing

Version 1.0 — February 2026

CONFIDENTIAL

EXECUTIVE SUMMARY

Refinery is an AI-powered manuscript analysis and refinement platform that operates at the scale of the full document — not the sentence, not the paragraph, but the entire 80,000-word manuscript held in memory simultaneously. It is the first tool of its kind to combine manuscript-level comprehension with craft-level analysis and revision-level implementation in a single, integrated environment.

This Requirements Definition Document specifies the functional, technical, and experiential requirements for all three Refinery product lines:

| Product Line | Primary Customer | Pricing Tier | Core Value Proposition |
|-----------------------|---|----------------------------|--|
| Indie Author | Self-published novelists, genre fiction writers, (Espionage authors) | \$10/mo (\$20/mo Pro) | Professional-grade editorial intelligence without the expense of a traditional editor. |
| Academic | Graduate students, doctoral candidates, faculty members, and institutional administrators | \$20/mo (\$50/mo Advanced) | Advanced analysis and argument coherence for long-form academic manuscripts. |
| Enterprise Publishing | Acquisition editors, literary agencies, academic publishers | \$500s-\$2,000+/mo | Manuscript screening at scale, reducing slush pile review times. |

All three product lines share a common AI infrastructure — the Refinery Intelligence Engine — but present distinct interfaces, feature sets, and workflows designed for each customer segment's unique needs and professional context.

SHARED PLATFORM ARCHITECTURE

The Refinery Intelligence Engine

All product lines operate on a shared backend infrastructure built around long-context AI that processes the full manuscript in a single pass. This is the core technical differentiator: existing tools analyze sentences (Grammarly), organize files (Scrivener), or generate text (ChatGPT). Refinery reads the entire work and cross-references every element against every other element simultaneously.

The Six Core Modules

All three product lines share the same six-module architecture, with feature access gated by product tier:

| Module | Core Function | Indie | Academic | Enterprise |
|-----------------------------------|--|--------------|----------|------------|
| 1. Manuscript Intelligence Engine | Full structural scan: duplication detection, character census, punctuation mapping, lexical fingerprinting, metaphor density heatmap | Promotion | Priority | Priority |
| 2. Voice Isolation Lab | Character dialogue extraction, voice fingerprinting, jargon detection | Detect | Priority | Adaptive |
| 3. Pacing Architect | Chapter-level action/emotion beat mapping, tension curve modeling, rhyme realization | Limited | Priority | Replaced |
| 4. Character Arc Workshop | Arc tracking, alternative arc generation, Want/Fear/Belief/Behavior mapping | Priority | Priority | Full |
| 5. Prose Refinery | Tic detection, filter word analysis, metaphor frequency, show vs tell flagged | Show vs tell | Priority | Full |
| 6. Revision Command Center | Prioritized edit queue, batch implementation, diff preview, versioned branching, regression detection | Priority | Priority | Priority |

* Academic tier replaces Pacing Architect with the Argument Coherence Engine.

Technical Stack

Backend: Flask/FastAPI with Python processing pipeline. **AI Layer:** Claude API (long-context, full-document processing). **Manuscript Graph:** Neo4j for cross-reference and relationship mapping. **Frontend:** React with real-time collaborative editing. **File Support:** .docx, .txt, .rtf, .pdf (import); .docx, .txt (export). **Security:** TLS 1.3 in transit, AES-256 at rest, SOC 2 Type II roadmap.

PRODUCT LINE 1: INDIE AUTHOR

Refinery Pro | Where Solo Authors Write Like They Have a Team

1.1 Market Context & Target Customer

The indie author market comprises over 2 million active self-published authors in the English-language market, with approximately 300,000 actively publishing on Amazon KDP and other platforms each year. The primary pain points are access to professional-quality editorial feedback (typically \$3,000–\$8,000 per manuscript), the inability to see structural problems invisible at the sentence level, and the iterative cost of multiple revision passes without clear diagnostic guidance.

Target customer profile: genre fiction writers (romance, science fiction, fantasy, mystery, thriller) who have completed at least one draft, are self-publishing or seeking traditional publication, and are technically comfortable enough to use web-based tools. Secondary audience includes literary fiction writers and aspiring authors in MFA or creative writing programs.

1.2 Pricing & Access Tiers

| Tier | Price | Included Features |
|-------------------------|--------------------------|--|
| Free | \$0/month | Manuscript upload up to 50,000 words; Prose Refinery basic (tic detection, filter words); Pacing Architecture analysis |
| Pro | \$29/month or \$249/year | Unlimited manuscript uploads; all six modules at full depth; Voice Isolation Lab with Blind Test; Character Development analysis |
| Pro + Manuscript Bundle | \$99 one-time | Single manuscript full analysis — all six modules, one time, no subscription required. Ideal for authors |

1.3 User Journey & Interaction Model

Onboarding Flow

- User creates account and uploads manuscript (.docx, .txt, .rtf, or .pdf supported)
- System performs 90-second X-ray diagnostic — the Manuscript Intelligence scan
- Dashboard populates with health scores across five dimensions: Structure, Voice Consistency, Pacing, Character Development, Prose Clarity
- User selects first module to enter based on highest-priority issues flagged by the scan
- Guided first-run walkthrough contextualizes each finding to the user's specific manuscript

Core Workflow Loop

The Indie workflow is self-directed and exploratory. Users move between modules based on what the dashboard surfaces. The Revision Command Center maintains a master edit queue that aggregates recommendations from all modules, prioritized by estimated impact. Users can implement changes individually, in batches, or export the full annotated manuscript for manual editing in Word or Scrivener.

1.4 Module Requirements — Indie Author Tier

Module 1: Manuscript Intelligence Engine

- Duplication detection: flags chapters, scenes, or paragraphs with >70% content similarity
- Character census: lists all named characters with frequency count and first/last appearance by chapter
- Timeline anomaly detection: flags chronological inconsistencies in stated or implied time references
- Lexical fingerprint: identifies the top 50 distinctive words for the manuscript as a whole
- Metaphor density heatmap: identifies dominant metaphor families; flags chapters with >3x average density
- Output: one-page manuscript health dashboard exportable as PDF

Module 2: Voice Isolation Lab

- Extracts all dialogue for each named character and renders side-by-side in dedicated view
- Voice fingerprint: unique vocabulary, sentence length average, punctuation habits, register (formal/casual/slang)
- Jargon bleed detection: flags when specialized language appears in characters who would not plausibly use it
- Blind Test: strips speaker tags from 10 randomly selected dialogue blocks; user identifies which character is speaking
- Similarity score: 0–100 index measuring how distinct each character's voice is from the others; flags characters with score <40

Module 3: Pacing Architect

- Chapter-level waveform: visual pulse graph plotting action density vs. emotional depth by chapter
- Tension curve: models rising/falling tension across the full manuscript arc
- Breathing space detector: flags sequences of 3+ consecutive high-action chapters with no emotional decompression
- Slow zone detector: flags sequences of 3+ consecutive low-action chapters with no active threat or escalation
- Scene resequencing sandbox: drag-and-drop interface for experimenting with chapter reordering; system updates tension curve in real time

Module 4: Character Arc Workshop

- Tracks Want / Fear / Belief / Behavior for each primary character across chapters
- Flags chapters where a character's behavior is inconsistent with their established Belief or Fear without narrative justification
- Transformation validation: identifies whether the arc's climactic change is earned by prior scene-level development
- Alternative arc generation: for any primary character, generates two alternative development paths with scene-level intervention suggestions
- Decision tree: for characters at major decision points, models the logical alternatives and their downstream consequences

Module 5: Prose Refinery

- Tic tracker: identifies top 20 recurring words or phrases unique to the manuscript; flags chapters with 2x+ average frequency
- Filter word detector: flags hedging language (seemed, appeared, felt, noticed, realized, wondered) with chapter-by-chapter count
- Show vs. tell analyzer: identifies passages describing emotion directly and suggests scene-level alternatives

- Sentence rhythm profiler: measures average sentence length, variance, and consecutive-short or consecutive-long sentence runs
- Metaphor frequency manager: tracks recurring metaphor families; flags overuse within any 5,000-word window

Module 6: Revision Command Center

- Master edit queue: aggregates all recommendations from all modules, scored by estimated impact (High / Medium / Low)
- Batch implementation: user can select a group of edits and implement with a single action
- Diff preview: before/after comparison for every proposed change; user accepts or rejects individually
- Version branching: user can fork the manuscript at any point to explore Arc A vs. Arc B without overwriting the original
- Regression detection: after each revision pass, system re-runs affected modules and flags any new issues introduced by the edits
- Export: clean .docx with all accepted changes incorporated; or annotated .docx with tracked changes for manual review

1.5 Non-Functional Requirements — Indie Tier

- Manuscript X-ray diagnostic completes within 90 seconds for manuscripts up to 120,000 words
- All module analyses complete within 3 minutes for standard manuscript length
- UI must be operable without prior training; new users reach first meaningful insight within 10 minutes of upload
- Mobile-responsive design; full functionality on tablet; read-only dashboard on mobile
- GDPR and CCPA compliant; user data and manuscript content never used for model training without explicit opt-in
- Auto-save every 60 seconds; 30-day version history retained

PRODUCT LINE 2: ACADEMIC

Refinery Academic | Structural Intelligence for Serious Scholarly Writing

2.1 Market Context & Target Customer

Over 180,000 doctoral dissertations are completed in the United States annually, plus approximately 2 million master's theses and hundreds of thousands of journal articles submitted by faculty and researchers. The academic writing market is underserved by AI tools: existing products focus on grammar correction (Grammarly) or citation management (Zotero, Mendeley) but do not address the structural and argumentative coherence problems that represent the primary reason academic manuscripts are rejected by journals, dissertation committees, or book acquisitions editors.

Primary target customers: (a) doctoral students in humanities, social sciences, and professional programs; (b) faculty preparing book manuscripts or journal articles; (c) dissertation advisors and program directors; (d) writing centers and academic support programs. Institutional pilot targets include Columbus State University and Chattahoochee Valley Community College.

2.2 Pricing & Access Tiers

| Tier | Price | Included Features |
|-----------------------|----------------------------|---|
| Individual Student | \$29/month | Full access for single user; up to 3 active manuscripts; all six modules; dissertation-specific analysis |
| Advisor+ | \$79/month | Up to 10 student manuscripts in shared advisor dashboard; advisor annotation layer; progress tracking |
| Institutional License | Custom / from \$2,500/year | Department or college-wide access; LMS integration (Canvas, Blackboard, Moodle); dedicated onboarding support |

2.3 Academic-Specific Feature Adaptations

Argument Coherence Engine (replaces Pacing Architect)

- Thesis statement extraction: identifies the primary claim and maps where it is restated, developed, or contested across the manuscript
- Evidence-to-claim ratio: measures whether each major argument is supported by sufficient evidentiary weight
- Logical progression map: visualizes the argumentative structure chapter-by-chapter; flags where the argument stalls, loops, or contradicts itself
- Counterargument coverage: identifies whether acknowledged counterarguments are adequately addressed or simply dismissed
- Chapter coherence score: measures how well each chapter's argument connects to the preceding and following chapters

Citation & Source Architecture (new module for academic tier)

- Citation frequency heatmap: identifies chapters over-reliant on a single source (>30% of citations to one author or work)

- Source recency analysis: flags when a field with active recent literature relies primarily on sources older than 10 years
- Primary vs. secondary source balance: measures ratio of direct primary source engagement vs. reliance on secondary literature
- Citation gap detection: flags substantive claims made without evidentiary citation
- Supports APA, MLA, Chicago, and AMA citation formats

Academic Voice Calibration (replaces Voice Isolation Lab)

- Register consistency: flags shifts between formal academic register and informal or colloquial language
- Hedge appropriateness: identifies both over-hedging (weakens the argument) and under-hedging (overclaiming without qualification)
- Passive voice density: chapter-level tracking; flags chapters with >40% passive constructions
- Discipline-specific vocabulary validation: compares lexical profile against corpus of published work in the user's discipline

Dissertation Committee Report Generator

- Generates a structured 3–5 page committee-ready progress report summarizing manuscript health across all dimensions
- Customizable report templates: proposal defense, chapter review, full draft review, final defense preparation
- Advisor annotation layer: advisor can add inline comments and overall assessment within the shared dashboard
- Institutional report template option for programs with specific formatting requirements

2.4 Advisor Dashboard Requirements

- Single view shows all assigned students, each with a manuscript health snapshot (current score, last activity, flagged issues count)
- Advisor can open any student's manuscript in read-only mode and access all module outputs
- Annotation tool: advisor can add comments anchored to specific manuscript passages, visible in student's view
- Progress tracking: longitudinal view showing how manuscript health scores change over revision cycles
- Batch report generation: generate committee-ready reports for all students in one action
- Email notification to advisor when a student completes a full analysis pass

2.5 Institutional Integration Requirements

- LMS integration via LTI 1.3 standard (Canvas, Blackboard, Moodle, D2L)
- Single sign-on (SSO) via SAML 2.0 and OAuth 2.0 for institutional identity providers
- FERPA compliance: all student manuscript data is institution-controlled; Refinery staff have no access to student content without institutional administrator approval
- Usage analytics dashboard for writing center directors: active users, analyses completed, most common issues flagged
- Custom institutional branding option: university logo and color scheme on student-facing interface
- IRB data handling compliance documentation available upon request

2.6 Non-Functional Requirements — Academic Tier

- Full analysis of 100,000-word dissertation completes within 5 minutes
- Advisor dashboard loads within 3 seconds for up to 10 active student manuscripts
- 99.5% uptime SLA during academic calendar periods (September–May, August–December)
- Data residency options: US-only servers available for institutions with data sovereignty requirements
- Audit trail: all access and edits logged for compliance purposes; log retention 7 years

PRODUCT LINE 3: ENTERPRISE PUBLISHING

Refinery Enterprise | Manuscript Intelligence at Acquisition Scale

3.1 Market Context & Target Customer

Traditional publishing houses receive between 10,000 and 100,000+ manuscript submissions per year (depending on size) and employ acquisition editors, editorial assistants, and readers to triage this volume. Literary agencies face a similar challenge: top agencies receive 20,000–50,000 query letters annually. The slush pile problem — enormous submission volume, limited qualified reader hours — is the central operational challenge for both.

Academic publishers (university presses) face a different version of the same problem: peer review pipelines are slow, editorial staff evaluate manuscripts that may not survive double-blind review, and the cost of acquiring and then rejecting a manuscript is significant.

Primary enterprise customers: (a) Big Five and independent trade publishers; (b) literary agencies with active submission programs; (c) academic presses and university publishing operations; (d) hybrid and assisted publishing companies at scale.

3.2 Pricing & Contracting Model

| Package | Price Range | Scope |
|----------------------|--------------------------|---|
| Agency Starter | \$500/month | Up to 50 full manuscript analyses/month; 3 editor seats; genre-specific scoring; query letter pre-scr... |
| Publisher Standard | \$1,200/month | Up to 200 analyses/month; 10 editor seats; acquisition workflow integration; custom scoring rubrics... |
| Publisher Enterprise | \$2,500+/month | Unlimited analyses; unlimited seats; dedicated instance option (local hosting available); full API acc... |
| Academic Press | Custom / annual contract | Peer review triage support; peer reviewer matching; revision tracking across submission rounds; ed... |

3.3 Enterprise-Specific Feature Set

Acquisition Triage Dashboard

- Batch upload: accepts 1–500 manuscript submissions simultaneously via drag-and-drop or API
- Auto-scoring: each manuscript receives a composite Acquisition Score (0–100) based on configurable rubric across structure, voice, pacing, market fit, and originality
- Priority queue: manuscripts ranked by Acquisition Score; editors begin with the highest-ranked submissions
- One-click reader report: generates a 1–2 page acquisition summary for each manuscript; format mirrors the house's existing reader report template
- Comparative analytics: shows how the current submissions batch compares to prior successful acquisitions from the same house

Custom Scoring Rubrics

- Editorial directors can configure scoring weights by genre (literary fiction, commercial fiction, narrative nonfiction, category romance, etc.)

- House voice calibration: upload 5–20 previously acquired and published titles; system calibrates its scoring to align with the house's editorial sensibility
- Genre comparables: system identifies the closest comp titles within the past 5 years for each submission
- Market fit scoring: estimates commercial viability based on current market trends in the genre (optional add-on)

Multi-Editor Collaborative Workflow

- Seat management: unlimited reader/editor accounts with role-based access (Reader, Editor, Acquisitions Director, Admin)
- Annotation layer: all editors can annotate any manuscript in the shared queue; annotations visible to all with appropriate permissions
- Editorial conference mode: side-by-side view for two editors reviewing the same manuscript simultaneously with real-time annotation sync
- Decision workflow: Reader submits → Editor reviews → Acquisitions Director approves/rejects; each stage triggers notification and logs timestamp
- Rejection letter drafting: system generates a personalized, professional rejection letter based on the manuscript's specific strengths and weaknesses

Local Hosting Option (Enterprise & Law Firm Extension)

- On-premise deployment available for publishers with strict data confidentiality requirements (pre-publication manuscript security)
- Container-based deployment (Docker/Kubernetes) on the client's own infrastructure
- Air-gapped option: fully offline operation with no data leaving the client's network — critical for major house pre-publication security protocols
- This same local hosting architecture supports law firm, medical, and financial services verticals where document confidentiality is legally mandated
- Dedicated implementation team provides setup, training, and ongoing support for local instances

API & Integration Layer

- Full REST API for integration with existing submission management systems (Submittable, QueryManager, PubTrack)
- Webhook support: trigger analyses automatically when new submissions arrive via API
- CRM integration: Salesforce and HubSpot connectors for acquisition pipeline management
- Export formats: PDF reader reports, CSV scoring data, Word annotated manuscripts, JSON data export for custom downstream use

Literary Agency Extensions

- Query letter pre-screening: before requesting a full manuscript, editors can submit the query letter and first 10 pages for an AI-assisted viability assessment
- Submission history tracking: searchable record of every manuscript evaluated, including scoring history across multiple submission rounds
- Author relationship management: links multiple submissions from the same author; flags if an author has previously submitted under a different title
- Genre trend dashboard: visualizes what genres and subgenres are appearing at highest volume in the submission pile, updated monthly

3.4 Security & Compliance Requirements — Enterprise Tier

- SOC 2 Type II certification roadmap (target: 18 months from launch)
- GDPR compliance for European publisher clients
- NDA-grade manuscript confidentiality: all manuscripts processed under enterprise contracts are subject to strict confidentiality provisions; Refinery staff are excluded from content access
- Zero-retention option: manuscripts are deleted from Refinery servers within 24 hours of analysis completion
- Penetration testing: annual third-party security audit
- Data loss prevention (DLP): no manuscript content is accessible via Refinery employee accounts; content access requires dual-authorization approval from the enterprise client

3.5 Non-Functional Requirements — Enterprise Tier

- Batch of 100 manuscripts (average 80,000 words each) fully analyzed within 4 hours
- Single manuscript Acquisition Score and reader report generated within 5 minutes
- 99.9% uptime SLA with financial remedy for violations
- Dedicated customer success manager for contracts above \$1,200/month
- 24-hour support response for P1 issues; 48-hour for P2
- Custom onboarding and editorial team training included for all contracts

DEVELOPMENT ROADMAP — ALL PRODUCT LINES

| Phase | Timeline | Deliverables | Success Criteria |
|-----------|--------------------------|--|--|
| Alpha MVP | Weeks 1–12 (Q1 2026) | Manuscript Intelligence Engine + Prose Refinery live; End-to-end API integration ready for beta users. | Initial 5 beta users in 4 weeks; NPS (24) and user satisfaction (4.5/5). |
| Beta | Weeks 13–24 (Q2 2026) | All six modules live for Indie Pro; Academic tier completed. | 500+ customers; Satisfaction rating 4.8/5; \$75K in new bookings. |
| V1 Launch | Month 7–9 (Q3 2026) | Public launch all three tiers; Enterprise pilot with 1 agency. | 100+ publishers; SALT award nomination; MR \$25K. |
| V1.5 | Month 10–15 | Local hosting option; API layer for enterprise; LMS integration. | Enterprise clients; Educational institutions; \$500K in AR. |
| V2.0 | Month 16–20 | Custom AI fine-tuning for enterprise house voice; market expansion. | 10,000+ users; enterprise AR \$500K; 5 new international markets. |

COMPETITIVE POSITIONING

Refinery competes in a market where no existing tool performs its core function. This is the key strategic insight: Refinery is not competing with Grammarly, Scrivener, or ChatGPT. It is creating a new product category — manuscript-scale analytical intelligence — that none of them address.

| Capability | Refinery | Grammarly | Scrivener | ChatGPT / Claude |
|-------------------------------------|------------------|---------------|----------------|------------------|
| Operates at full manuscript scale | Yes | No (sentence) | No (file mgmt) | Limited |
| Voice consistency across characters | Yes | No | No | Manual only |
| Pacing & structural analysis | Yes | No | No | Manual only |
| Academic argument coherence | Yes | No | No | No |
| Enterprise acquisition triage | Yes | No | No | No |
| Local hosting / air-gapped option | Yes (Enterprise) | No | Yes | Limited |
| Multi-editor collaborative workflow | Yes | Limited | No | No |

RISK REGISTER

| Risk | Likelihood | Impact | Mitigation |
|---|------------|----------|---|
| AI API cost scaling at volume | Medium | High | Implement analysis caching; optimize prompts for token efficiency; negotiate volume discounts with providers. |
| Manuscript data security breach | Low | Critical | Zero-trust architecture; zero-retention option for enterprise; third-party penetration testing. |
| Large competitor (Grammarly, Adobe) enters market | Medium | High | Accelerate enterprise and institutional sales; build switching costs through version lock. |
| Academic pilot fails to convert to paid users | Medium | Medium | Structure pilot agreements with clear conversion milestones; identify internal champions. |
| MVP quality insufficient for beta user retention | Extended | High | Founder-led beta testing with 4 committed users; weekly iteration cycles; NPS monitoring. |

APPENDIX: GLOSSARY & DEFINITIONS

Core Terms

- **Manuscript Intelligence Engine:** The foundational analysis module that performs a full structural scan of the uploaded manuscript, producing the health dashboard.
- **Voice Isolation Lab:** Module that extracts, compares, and scores the distinctiveness of each character's dialogue and narration voice.
- **Pacing Architect:** Module that measures the rhythm of action and emotional beats across chapters and models the tension curve of the full manuscript.
- **Character Arc Workshop:** Module that tracks character development across chapters and generates alternative arc options with scene-level intervention suggestions.
- **Prose Refinery:** Module that identifies writing tics, filter words, show-vs-tell passages, and sentence rhythm problems at the manuscript scale.
- **Revision Command Center:** The integrated revision environment where all module recommendations are aggregated, prioritized, and implemented.
- **Acquisition Score:** The composite 0–100 score generated for Enterprise Publishing clients representing an AI-assisted estimate of a manuscript's commercial and editorial viability.
- **House Voice Calibration:** The process by which an Enterprise client uploads previously acquired titles to calibrate Refinery's scoring rubric to their editorial sensibility.
- **Advisor Dashboard:** The shared workspace for Academic Tier advisor users, showing all assigned student manuscripts and enabling annotation, progress tracking, and report generation.
- **Local Hosting:** On-premise deployment option for Enterprise clients requiring air-gapped data security; no manuscript data leaves the client's network.

— END OF REQUIREMENTS DEFINITION DOCUMENT —

Refinery | Version 1.0 | February 2026 | CONFIDENTIAL