

Here is a compact requirements-only document without implementation phases.

---

## AI Agent for GoodFirms Reviews – Requirements

### 1. Objective

Build an **AI agent** that reviews, enhances, and prepares client reviews for the GoodFirms profile while preserving authenticity and complying with GoodFirms' review and reputation guidelines.goodfirms+1

### 2. Scope

- In-scope:
  - Intake of raw client feedback from forms, emails, and CRM notes.
  - Automated quality, sentiment, and compliance checks.
  - Language and structure enhancement of reviews without altering facts.
  - Risk detection for fake, biased, or non-compliant reviews.
  - Recommendation workflow: AI suggestion → human approval/edit → final review text for GoodFirms.
- Out-of-scope (current version):
  - Direct publishing to GoodFirms (manual posting assumed).
  - Non-text review assets (images, videos, large attachments).

### 3. Inputs & Data Model

- Mandatory fields: reviewer name, role, company, project name/summary, engagement type, budget range, duration, technologies, rating (e.g., 1–5), free-text feedback.goodfirms+1

- Optional fields: NPS/liability to recommend, industry, geography, project URL/portfolio link.[\[goodfirms\]](#)
- Internal metadata: review source (form/email/CRM), collection date, internal owner, GoodFirms status (draft/submitted/published).[\[goodfirms\]](#)

## 4. Functional Requirements

### 1. Ingestion & Validation

- Accept inputs via API and simple internal form.
- Validate mandatory fields, sanitize text, normalise to a standard JSON schema.bit+1

### 2. Content Analysis

- Perform sentiment analysis and topic extraction (communication, timeliness, quality, support, ROI).getstream+1
- Detect spam, templated content, self-written reviews by internal team, and duplicate/near-duplicate reviews.
- Score each review on completeness and clarity (e.g., 0–100).

### 3. Enhancement (Text Generation)

- Improve grammar, coherence, and narrative flow while preserving meaning and sentiment.
- Ensure GoodFirms-friendly structure: context (client & project), problem, solution, delivery, outcomes, recommendation.goodfirms+1
- Provide a side-by-side original vs enhanced version and a short explanation of key edits.

### 4. Compliance & Safety Moderation

- Check for: hate/abuse, threats, NSFW, slurs, personal data leakage (emails, phones, addresses), defamation, and legal risk terms.analytics+2

- Enforce platform-style rules: no fake testimonials, no unverifiable claims, no explicit competitor-bashing.[[goodfirms](#)]
- Classify each review into:
  - “Safe – Ready for approval”
  - “Borderline – needs human review”
  - “High risk – recommended reject”

## 5. Authenticity & Risk Signals

- Heuristics for suspicious reviews: identical language patterns, mismatched project details, extreme ratings with vague text, known bad domains/emails.opportunity+1
- Expose an “authenticity confidence” score (0–1) with reasons.

## 6. Reviewer Guidance (Internal UX)

- Present a compact card for each review: key fields, sentiment, risk flags, enhanced text, and recommendation.
- Allow internal user to: approve enhanced text, edit then approve, or reject with a reason.
- Log every decision and edit for audit and later tuning.

## 7. Explainability & Logging

- For each moderation/enhancement decision, store labels (e.g., “PII detected”, “toxic language”) and a 1–2 line justification.permit+1
- Maintain logs with timestamps, user IDs, and model versions used.

# 5. Non-Functional Requirements

- Accuracy & Quality
  - Target ≥ 90% of “Safe – Ready for approval” items requiring no or minor human edits after calibration.getstream+1

- Limit false positives for “High risk” flagging to < 5% on a validation set.

- **Performance**

- Process a single review (analysis + enhancement + checks) in ≤ 3 seconds under normal load.

- **Scalability**

- Support at least 500 reviews/day with linear scale-out capability for higher volumes.

- **Security & Privacy**

- Encrypt data in transit (TLS) and at rest; restrict access via role-based access control.<sup>anolytics+1</sup>
- No sharing of review data with third parties beyond configured model providers; comply with GoodFirms’ and internal data policies.[\[goodfirms\]](#)

- **Reliability**

- System should be highly available during working hours with graceful degradation (e.g., fallback to “analysis only” if enhancement model unavailable).