STRATEGY DOCUMENT · AI CONTENT OPERATIONS

AI Content Production System

A custom n8n-powered content automation architecture replacing a $1,500-per-month vendor stack with an in-house system of sixteen workflows, seven integrations, and a five-stage human approval layer across seven websites.

COST REDUCTION	WORKFLOWS	MONTHLY VISITS	CLIENT
75%	**16**	**315K+**	**Cross Even**

Nimra Shakil

Executive Summary

This document describes the AI content production system built to support a seven-site publishing operation across the ESA, cannabis, telehealth, and weight-loss verticals. The system replaced a $1,500-per-month third-party content vendor (Tely.ai) with a custom n8n-powered orchestration layer, sixteen production workflows, and a five-stage human approval architecture. The result was a **75% reduction in per-article cost**, **$8,448 to $13,944 in annual savings**, and an estimated **80% reduction in manual process overhead**, while supporting 315K+ combined monthly visits across the portfolio.

The design brief was not “generate more content.” It was to hold quality steady at scale while removing the vendor dependency that had become the cost bottleneck. Every architectural decision in this document traces back to that framing.

The Build-vs-Buy Decision

The operation had been running on Tely.ai, a content-generation SaaS producing 240 articles a month across four brands (60 per brand, two per day). The headline problem wasn’t output volume. It was two structural constraints that neither more budget nor more articles could solve.

Constraint 1: Quality control was outsourced

Tely.ai operated as a black-box pipeline. Content arrived finished; the editorial team could approve or reject, but could not intervene at the stages where voice, structure, or SEO decisions were being made. In multi-brand operations, brand voice drift is the single largest failure mode. A system that surfaces completed work downstream of the decisions that matter cannot fix that failure mode; it can only flag it after the fact.

Constraint 2: The economics were inverted

At $18,000 per year for 240 articles a month, the per-article cost landed at $6.25. Adding brands to the Tely.ai plan cost an additional $375 per brand per month. In-house infrastructure runs at $50 to $100 per brand per month. The vendor model was cheapest at two or three brands and progressively more expensive with each additional site. The portfolio was moving toward seven.

The decision

A custom build on n8n with a paid API layer pencilled out at $4,056 to $9,552 annually, against the $18,000 baseline. More importantly, the model inverted: adding brands became cheap rather than expensive, and quality control moved inside the pipeline rather than after it. Content-generation SaaS optimises for output volume. A multi-site operation is constrained by QC bandwidth, not generation capacity. The two don’t share a solution.

Financial impact

- **Annual cost:** $18,000 (Tely.ai) vs. $4,056 to $9,552 (in-house). Savings: $8,448 to $13,944.
- **Cost per article:** $6.25 (Tely.ai) vs. $1.41 to $3.32 (in-house). Reduction: 74 to 77 percent.
- **Scaling cost:** $375 per additional brand per month (Tely.ai) vs. $50 to $100 per brand per month (in-house). The gap widens linearly with portfolio size.

- **Setup cost:** Zero. The build used existing in-house engineering capacity rather than external contractors.

Integration Architecture

The orchestration layer runs on n8n, selected for its visual workflow model (non-technical operators can read and modify the pipelines), node library (pre-built connectors for every platform in the stack), and self-hostable footprint ($18 per month on a single server). Seven external services connect into n8n, each chosen for a specific function.

Google Analytics 4

Purpose: user-behaviour and conversion signal for content prioritisation. Primary integration: GA4 Data API for dimensional reports and real-time reporting for current-traffic patterns. Metrics consumed: page views and unique visitors, session duration and bounce rate, conversion rates and goal completions, user engagement signals, and traffic sources and referral data. GA4 feeds the traffic component of the opportunity-scoring algorithm and supplies the denominator for post-publication performance evaluation.

Google Search Console

Purpose: search performance and ranking intelligence. Primary integration: the Search Analytics API for query performance, URL inspection for indexing status, and performance reports for ranking trends. Core Web Vitals feed the technical-SEO checks that run as part of the pre-publication pipeline. The GSC layer is where quick-win identification happens: pages ranking in positions 4 to 20 with high impressions but sub-category CTR get flagged for optimisation rather than net-new content.

DataForSEO

Purpose: competitive intelligence and keyword research at scale. Services consumed: Backlinks API for competitor backlink profiling, Keywords API for search volume and keyword difficulty, SERP API for ranking analysis, and Domain Analytics for authority metrics. DataForSEO replaced a combination of Semrush and Ahrefs seat costs with a consumption-priced API at $100 to $500 per month depending on analysis volume.

Crawl4AI

Purpose: self-hosted content extraction and technical assessment. Metrics produced: content quality and structure, internal linking patterns, technical SEO factors, content freshness and relevance, and media and visual content assessment. Self-hosting kept per-crawl costs near zero at portfolio scale, which mattered because competitive-intelligence crawling ran weekly across every competitor set.

OpenRouter

Purpose: multi-model LLM access through a single API surface. The routing logic is what matters: cheaper models (lower latency, lower cost per token) run the screening and triage steps; premium models run the generation and optimisation steps where quality cost matters more than per-call price. A single vendor lock-in on LLMs is the fragility failure mode for every content-gen system built today. OpenRouter removes it.

NocoDB

Purpose: structured storage layer, selected as the self-hosted replacement for Airtable. Holds the brand configuration table (domain, GA4 property ID, GSC site URL, competitor list, target keywords, analysis frequency), the analysis_results table (scored opportunities per brand per page), and the campaigns table (outreach and optimisation tracking). NocoDB on PostgreSQL gave the relational guarantees the content pipeline needed without per-seat licensing.

ClickUp

Purpose: human approval surface. Every one of the five approval stages terminates in a ClickUp task assigned to the reviewer responsible for that stage, with the content artifact, scoring context, and required approval criteria attached. Task comments become the audit trail. This was the single design choice that made the five-stage approval model operationally viable rather than theoretical.

WordPress REST API

Purpose: multi-brand publishing endpoint. Once content clears the final approval stage, the n8n workflow pushes to the brand’s WordPress instance with meta title, meta description, canonical, category, tags, featured image, and internal link assignments set programmatically. Seven sites, one publishing integration, no manual copy-paste step.

Opportunity Scoring Methodology

The system’s prioritisation layer is a composite scoring algorithm that lets editorial effort concentrate on the pages where incremental work will produce the largest traffic return. The composite runs four factors, each weighted to reflect its leverage over ranking and conversion outcomes. The weights are not symmetrical on purpose: backlink gap carries the most weight because it is the slowest and most expensive to close, and content quality carries the least because it is the fastest to improve.

Composite opportunity score = (Traffic Potential × 0.25) + (Ranking Opportunity × 0.25) + (Backlink Gap × 0.30) + (Content Quality × 0.20). Each factor resolves to a 0-100 score built from its own weighted sub-components.

Traffic Potential (25%)

Monthly organic sessions (40%), search volume for target keywords (35%), and seasonal trends or growth potential (25%). The pattern this surfaces: pages already receiving traffic but ranking on low-volume terms, and pages ranking on high-volume terms but receiving disproportionately little traffic. Both signal unrealised upside.

Ranking Opportunity (25%)

Current average position scored inversely (50%), keyword difficulty gap (30%), and CTR improvement potential (20%). The inverse position scoring is deliberate: a page ranking position 12 on a competitive term is a higher-leverage target than a page ranking position 3 on the same term, because the next ranking move creates more incremental traffic. Position 4 to 20 is where quick wins live.

Backlink Gap (30%)

Competitor backlink advantage (60%), domain authority differential (25%), and linking domain diversity gap (15%). This is the heaviest factor because closing a backlink gap is the slowest work in SEO and the hardest to reverse once gained. The sub-weighting reflects that concentrated authority gaps are worse than distributed ones: a competitor with 60% more backlinks on a single page is a larger threat than a competitor with 20% more backlinks across ten.

Content Quality (20%)

Content depth and comprehensiveness (40%), technical SEO factors (30%), user engagement signals (20%), and content freshness (10%). Lightest weighting because content quality is the fastest lever to move. A depth gap can be closed in a week; a backlink gap takes a quarter. Weighting reflects actual time-to-impact.

Output categorisation

The composite score buckets pages into four action categories: **Quick Wins** (high traffic potential, low competition), **Strategic Targets** (high business value regardless of difficulty), **Competitive Gaps** (backlink opportunities on pages where three or more competitors hold links), and **Content Enhancement** (pages with existing ranking but quality gaps). Each bucket routes to a different workflow downstream, which is what makes the categorisation more than taxonomy.

Workflow Catalog

Sixteen production workflows were built on the orchestration layer. Three are detailed below because they carry most of the system’s strategic weight. The remainder are named in the appendix at the end of this section.

1. Backlink Opportunity Identification

The most computationally expensive workflow in the catalog. A weekly cron trigger initiates a four-source data pull: GA4 (top 100 pages by traffic with engagement metrics), GSC (pages ranking in positions 4 to 20 with CTR gaps), DataForSEO (top 10 competitors’ backlink profiles with common linking domains), and Crawl4AI (content analysis for linkability and quality scoring).

A custom scoring function produces a composite opportunity score per page, weighted across four factors: traffic potential (25%), ranking opportunity (25%), backlink gap (30%), and content quality (20%). The output is a prioritised list bucketed into four categories: Quick Wins (high traffic, low competition), Strategic Targets (high business value), Competitive Gaps (competitor backlink opportunities), and Content Enhancement (pages needing content improvement). The top 20 priority pages flow into the editorial calendar automatically.

2. Content Production Pipeline

The core content manufacturing workflow. Stages execute in sequence: keyword research via DataForSEO (with brand-specific filtering and opportunity scoring), competitive content analysis via Crawl4AI (structure, angle, coverage gaps), outline generation via OpenRouter (premium model), human brief approval, first-draft generation via OpenRouter, human draft approval, optimisation pass (meta content, internal links, schema), final human approval, and publish via WordPress REST API. The pipeline includes multiple API redundancies and fallback systems so that a single LLM provider outage does not halt production.

3. Content Quality Assessment

A monitoring workflow rather than a production workflow. Crawl4AI sweeps the live portfolio, scoring every published page against four dimensions: content depth and comprehensiveness (40%), technical SEO factors (30%), user engagement signals (20%), and content freshness (10%). Pages scoring below threshold enter the refresh queue; pages scoring above threshold with declining performance enter the investigation queue. The workflow surfaces content decay before it shows up in the traffic numbers.

4. Internal Link Architecture Automation

Internal linking is where most in-house SEO operations lose discipline at scale. This workflow enforces it. On every publish event, the pipeline runs a semantic match between the new article's target keyword cluster and the existing site's topical inventory. Matching anchor opportunities are surfaced as a ranked list with the target URL, suggested anchor text, and insertion context preserved. The editor approves or rejects each link inline. The result: every published piece ships with its internal link architecture already decided, rather than being back-filled weeks later or skipped entirely. Across seven sites at scale, this is the difference between a site that compounds topical authority and one that leaks it.

5. Brand Voice Validation

The failure mode that content-generation SaaS never solves is voice drift across brands. This workflow addresses it structurally. Each brand in the NocoDB configuration table carries a voice profile: tone-of-voice descriptors, banned phrasings, preferred vocabulary, sentence-length distribution, and a reference corpus of approved past content. On every draft generation, the workflow scores the output against the profile and flags deviations above threshold before the draft surfaces for human review. Editors spend their time on judgment calls, not on catching obvious voice mismatches.

Workflow catalog (full list)

- Backlink Opportunity Identification
- Content Production Pipeline (topic → publish)
- Content Quality Assessment
- Keyword Research and Clustering
- Competitive Content Gap Analysis
- Internal Link Architecture Automation
- Meta Content and Schema Generation
- Competitor Backlink Monitoring
- Ranking and Position Change Alerts
- Core Web Vitals Monitoring
- Content Refresh Queue Management
- Brand Voice Validation
- Image Generation and Optimisation
- WordPress Publishing with SEO Metadata
- Performance Reporting and Dashboard Sync

- Editorial Calendar and ClickUp Task Automation

Five-Stage Human Approval Architecture

The approval layer is the system's defining feature. Rather than running generation end-to-end and reviewing the output, humans intervene at the five points where the decisions that shape quality are being made. Each stage surfaces through a ClickUp task with the artifact attached and a defined approval criterion.

Stage 1: Topic selection and keyword strategy

The first gate. Before any generation happens, the proposed topic, target keyword cluster, and opportunity score pass to the editorial lead. Criterion: does this topic match the brand's content architecture and the quarter's strategic priorities? Rejected topics go back into the backlog; approved topics proceed to brief generation.

Stage 2: Content brief and outline

The structural gate. Heading outline, section-level talking points, target word count, internal linking anchor list, and citation strategy all land for review. Criterion: does this outline answer the search intent more completely than the top three ranking competitors? Brief-stage rejection is where the sharpest quality leverage sits, because fixing structure here costs minutes; fixing it post-draft costs hours.

Stage 3: Draft content review

The voice and accuracy gate. The full first draft arrives for editorial review. Criterion: brand voice, factual accuracy, and tone-of-voice fidelity. Edits are returned in-line; the workflow re-generates affected sections rather than rewriting manually.

Stage 4: Meta content and SEO optimisation

The on-page SEO gate. Meta title, meta description, schema markup, internal link placements, and image alt text all land together. Criterion: every ranking factor the team controls at the page level has been set deliberately, not defaulted. This stage catches the errors that kill search performance long after publication.

Stage 5: Final publication approval

The last sign-off. Full page preview, readability score, Semrush readability target (50+), final internal link check, and publish target (draft / scheduled / live). Criterion: ready to publish. On approval, the workflow pushes to WordPress via REST API and closes the task.

Production Metrics

Six weeks after production rollout, the system hit the benchmarks that had been set during the build-vs-buy decision.

- **End-to-end content creation in under 4 hours.** From topic approval to live publication, including all five human approval stages. Tely.ai's equivalent cycle ran 24 to 48 hours with less editorial intervention.
- **95%+ workflow success rate.** Measured as the share of initiated workflows completing without manual intervention. API redundancy and retry logic did most of this work.

- **90%+ human approval efficiency.** Measured as the share of artifacts clearing each approval stage on first review. The number was deliberately optimised for: rejections are signal, not noise, and high first-pass rates confirm the upstream generation was well-targeted.
- **All four brands automated by Week 6.** Matched the accelerated six-week implementation plan. The remaining three sites onboarded on the same template over the following quarter.
- **Infrastructure cost: $18 per month server + $320 to $778 per month in API calls.** Against the $1,500 per month vendor cost replaced.

Implementation Timeline

The build shipped on a six-week accelerated timeline, compressed from the 22 to 24 weeks a conventional build would have required. Compression was possible because scope was held tight: every week had to ship a working component, and nothing was built that did not have an immediate consumer downstream.

Weeks 1 to 2: Core MVP

Server environment on $18/month hosting, NocoDB schema design, Crawl4AI self-hosted server, API credentials and rate-limit testing for DataForSEO, OpenRouter, GA4, and GSC. Keyword research workflow built and validated end-to-end. Content generation workflow built with multi-model fallback logic. ClickUp approval integration live. First brand onboarded and producing content through the full pipeline by end of Week 2.

Week 3: Intelligence layer

Competitive analysis workflow, trend detection, and the opportunity scoring algorithm shipped. The workflow started surfacing prioritised editorial targets rather than relying on editor intuition for topic selection. This was the inflection point where the system began outperforming the Tely.ai baseline on strategic signal, not just cost.

Week 4: Advanced features

On-page SEO automation, internal link architecture workflow, schema generation, and meta content workflows all shipped. The publishing layer went from “push content to WordPress” to “push fully optimised content to WordPress with every ranking factor the team controls at page level already set.”

Week 5: Analytics and monitoring

Performance reporting workflows, dashboard sync, Core Web Vitals monitoring, and ranking-change alerting all went live. This is where the observability work lived: the system needed to be legible before it could be trusted in production.

Week 6: Production rollout and team enablement

All four initial brands fully onboarded. Team training on ClickUp approval stages, voice profile management, and workflow troubleshooting. Documentation shipped alongside the system so operational knowledge didn’t concentrate in the build team.

Observability and Resilience

A production content operation cannot tolerate workflow failures that go undetected. The system was built with three layers of failure protection, each addressing a different failure mode.

API redundancy and fallback

Every external API integration carries a fallback path. The OpenRouter layer routes across multiple LLM providers; a single model outage triggers automatic failover to a secondary model with equivalent capability. DataForSEO queries fall back to cached results when the live API is rate-limited. The GSC and GA4 integrations queue requests during quota exhaustion rather than failing the workflow outright.

Retry logic and error handling

Every node that touches an external service carries exponential-backoff retry logic. Transient failures, which dominate the failure distribution in API-heavy workflows, resolve themselves without operator intervention. Non-transient failures surface as ClickUp tasks for the operations owner rather than as silent workflow aborts.

Monitoring alerts and audit trail

Workflow success rates, API cost burn, and per-brand throughput feed a monitoring dashboard with alert thresholds on each. Regular backups of the NocoDB configuration layer run nightly. Every approval stage writes to ClickUp as a permanent audit trail, which means any content decision made in the system can be traced back to a named reviewer and a timestamp. This matters for brand governance and matters more when the operation scales beyond the build team’s direct oversight.

Data Architecture

The storage layer was designed around three core relational tables rather than a document store. Relational guarantees matter here: per-brand configuration, per-page scoring, and per-campaign progress tracking are all naturally tabular, and the query patterns the workflows depend on (filtering pages by brand, joining scores to campaigns, aggregating performance over time) map cleanly to SQL. NocoDB sits on top of PostgreSQL to surface a spreadsheet-style UI for non-technical operators without compromising the underlying schema.

Brands configuration table

The system’s entry point. Each row is a brand and carries domain, GA4 property ID, GSC site URL, competitor domain list, target keyword set, analysis frequency, and brand voice profile. Adding a new brand to the system is a row insert plus API credential configuration. No workflow modifications are required because every production workflow reads its per-brand parameters from this table at runtime.

Analysis results table

The scoring layer’s output log. Each row captures a single analysis run for a single page: page URL, analysis type (backlink, content, technical), composite opportunity score, the four sub-component scores, recommendations as structured JSON, and analysis timestamp. The table is append-only, which preserves the scoring history for every page across time and enables longitudinal analysis of how editorial work moves the scores.

Campaigns tracking table

The execution layer. Each row represents a backlink outreach campaign, content refresh initiative, or optimisation sprint, keyed to a brand and a set of target pages. Status, start date, and performance metrics persist as the campaign runs. This is the table that answers the question *did the work we prioritised actually move the numbers*, which is the question every content operation eventually gets asked.

Per-brand onboarding

The operational consequence of this architecture is that onboarding a new brand into the system takes hours, not weeks. Brand row insert, API credentials configured, voice profile captured from a reference corpus of approved past content, competitor set populated, target keyword set seeded from the existing content audit. From that point, the full workflow catalog runs against the new brand without any pipeline changes. This is what inverted the scaling economics against the vendor model.

Security and Compliance

Production content systems touch three categories of sensitive material: API credentials for paid services, analytics data that includes user behaviour, and draft content that represents brand positioning before publication. The security layer was designed around all three, not just the first.

API authentication and credential management

All Google service integrations (GA4, GSC, and any Google API touched by the workflows) authenticate via OAuth 2.0 with automated token refresh running as a scheduled workflow. DataForSEO and OpenRouter credentials live in n8n’s encrypted credential store rather than in workflow parameters. Secret rotation is an operational procedure with ownership and schedule, not an aspiration.

Rate limiting and quota management

Every external API integration carries rate-limit awareness built into the workflow node. The system tracks quota consumption per provider and throttles proactively rather than waiting for rate-limit errors. DataForSEO usage, in particular, is cost-sensitive enough that a runaway workflow could materially affect monthly spend; monitoring and hard caps prevent that failure mode.

Data privacy and retention

Analytics data consumed from GA4 is processed in aggregate form for the scoring algorithm; no individual user sessions or personally identifiable information persist in the NocoDB layer. Minimal data retention policies mean the analysis_results table keeps scoring metadata but not the underlying traffic samples. All data transmission between n8n and external APIs runs over HTTPS with certificate validation enforced.

Audit and governance

Every content decision is traceable back to a named reviewer through the ClickUp approval history. Workflow execution logs persist for 90 days for debugging and compliance review. API cost logs persist indefinitely because they feed the cost-accounting layer that justified the build in the first place. If a piece of content produces a problem post-publication, the full decision trail, from topic approval through to publish, is recoverable in minutes.

Migration Readiness

The system was designed to outgrow itself. n8n was the right orchestration choice for the initial build because visual workflows meant non-engineering team members could read and modify the pipelines. At portfolio scale, some components earn a migration to a dedicated backend service.

The migration path is a NextJS backend for the custom-logic layers (scoring algorithms, brand-configuration management, dashboard surfaces) while retaining n8n for the connector-heavy integration flows (API calls, webhook handling, scheduled triggers). The split isn’t opportunistic; it maps to where each tool is strongest. n8n excels at orchestration; a backend framework excels at custom algorithms and typed APIs. Running both deliberately is cheaper than picking one and forcing it.

The database layer stays on PostgreSQL through NocoDB. The LLM layer stays on OpenRouter. The publishing layer stays on the WordPress REST API. The migration is additive rather than replacement-style, which is the only way a production content operation with 315K+ monthly visits can tolerate an infrastructure change.

Strategic Outcomes

The financial numbers are the headline. They are not the most important outcome.

- **80% reduction in manual process overhead.** The approval stages kept humans in the decisions that matter and removed them from the work that didn’t. This is what team leverage looks like in a multi-site operation.
- **Quality-at-scale held.** Per-brand voice consistency, factual accuracy, and on-page SEO discipline all improved versus the Tely.ai baseline. The five-stage architecture is the reason.
- **Durable in-house advantage over vendor dependency.** The system cannot be acquired, re-priced, or discontinued by a third party. Every component is either open-source, consumption-priced, or self-hosted.
- **Scaling economics inverted.** Adding the fifth, sixth, and seventh sites cost $50 to $100 per brand per month in incremental infrastructure. Under the prior vendor model, the same sites would have cost $375 each.

The underlying thesis: content-generation SaaS optimises the wrong constraint for a multi-brand operation. Once the constraint is correctly identified as QC bandwidth rather than generation capacity, the build decision becomes straightforward. The system documented here is the execution of that thesis.