NTRL
Article Categorization & Feed Construction
Canonical Product Specification (v1.1)

Production System · Expanded Diagrams & Admin Flows
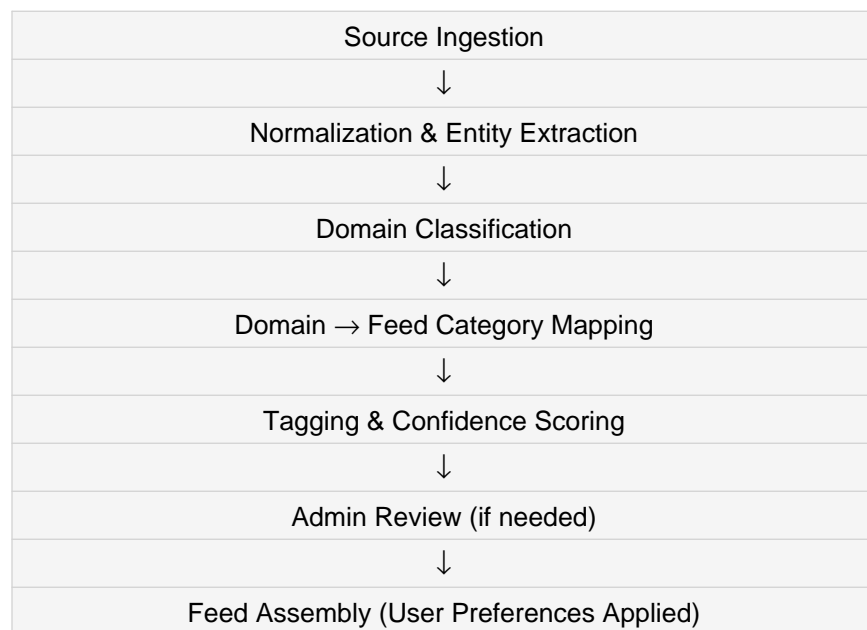
1. Purpose & Scope

This document defines the production-grade system for categorizing articles, mapping them to user-selected feed categories, and assembling the NTRL feed.

Version 1.1 expands on v1.0 with explicit system diagrams, admin workflows, and operational detail.

2. End-to-End System Overview

The diagram below shows the full lifecycle of an article from ingestion to user feed rendering.

Source Ingestion
↓
Normalization & Entity Extraction
↓
Domain Classification
↓
Domain → Feed Category Mapping
↓
Tagging & Confidence Scoring
↓
Admin Review (if needed)
↓
Feed Assembly (User Preferences Applied)

3. Canonical Feed Categories (User-Facing)

Feed categories are stable, descriptive nouns that users explicitly opt into.

Default available categories include: World, United States, Local, Business, Technology, Science, Health, Environment, Sports, Culture.

4. Internal Domain Taxonomy (System-Only)

Each article is assigned exactly one internal domain. Domains are exhaustive and non-overlapping.

- Global Affairs
- Governance & Politics
- Law & Justice
- Security & Defense
- Crime & Public Safety
- Economy & Macroeconomics
- Finance & Markets
- Business & Industry
- Labor & Demographics
- Infrastructure & Systems
- Energy
- Environment & Climate
- Science & Research
- Health & Medicine
- Technology
- Media & Information Systems
- Sports & Competition
- Society & Culture
- Lifestyle & Personal Affairs
- Incidents & Disasters

5. Domain → Feed Category Mapping

Internal domains map deterministically to feed categories using admin-configured rules.

This indirection preserves accuracy without inflating the UI.

Internal Domain	Mapped Feed Category
Global Affairs	World
Governance (US)	United States
Crime & Public Safety (City)	Local
Finance & Markets	Business
Technology	Technology
Science & Research	Science

Health & Medicine	Health
Sports & Competition	Sports
Society & Culture	Culture

6. Categorization Decision Flow

Categorization follows a strict, explainable order of operations.

1. Geographic Rules
2. Internal Domain Classification
3. Domain → Category Mapping
4. Conflict Resolution Rules
5. Confidence Scoring
6. Review Flagging (if needed)

7. Feed Assembly with User Preferences

User-selected categories directly control which sections appear in the feed.

Selected categories with no qualifying articles are silently omitted.

User Topic Selection
↓
Enabled Categories (Ordered)
↓
Articles Filtered by Category
↓
Empty Categories Removed
↓
Rendered Feed

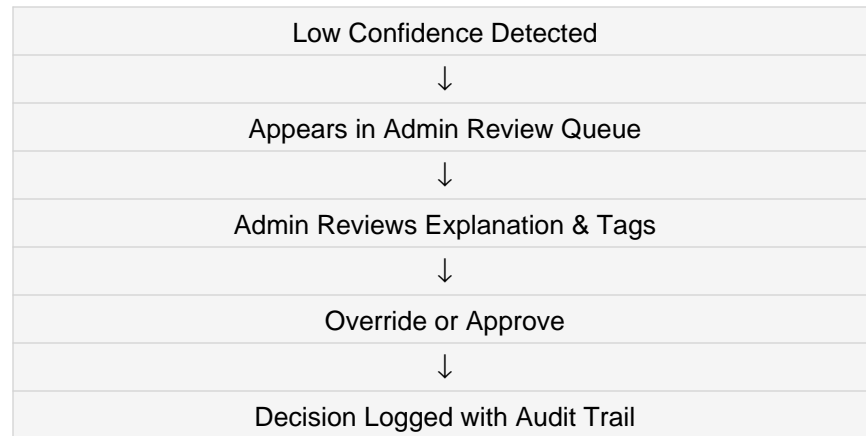## 8. Admin Configuration Flows (Expanded)

Admin tooling exists to ensure correctness, transparency, and long-term taxonomy stability.

- Category Management: Add, remove, rename feed categories.

- Domain Mapping: Map internal domains to feed categories.

- Quota Configuration: Set per-category and per-source limits.

- Review Queue: Inspect low-confidence categorizations.

- Override Actions: Apply manual overrides with required reason.

- Audit & Versioning: Track taxonomy and override history.

9. Admin Review Queue Flow

Low-confidence or conflicting categorizations are routed to a structured review queue.

Low Confidence Detected
↓
Appears in Admin Review Queue
↓
Admin Reviews Explanation & Tags
↓
Override or Approve
↓
Decision Logged with Audit Trail

10. Backend Data Contract

The following fields are mandatory for each article record:

feed_category, internal_domain, tags, categorization_confidence, needs_review, override_history, taxonomy_version.

11. Design Outcomes

This system enables user control without editorial bias, scales indefinitely, and structurally prevents manipulation.

The separation of domains, categories, and tags ensures long-term stability.