

NTRL

Article Categorization & Feed Construction

Canonical Product Specification

Version 1.0 · Production System

1. Purpose & Core Principles

This document defines the production-grade system for categorizing all ingested articles, enabling user-selected feed topics, and assembling a calm, complete NTRL feed.

Locked principles: Categorization is descriptive, not interpretive. Every article has exactly one feed category. Categories are user-selectable. Empty categories are omitted. All decisions are explainable and auditable.

2. Conceptual Model

Feed Categories are user-facing organizational buckets.

Internal Domains & Tags represent system knowledge used for accuracy, filtering, and quality assurance.

This separation prevents UI simplification from corrupting classification quality.

3. Canonical Feed Categories

World · United States · Local · Business · Technology · Science · Health · Environment · Sports · Culture

Categories are stable nouns, not editorial constructs. Order and visibility are user-controlled.

4. Internal Domain Taxonomy

Each article is classified into exactly one internal domain, never shown directly to users.

Domains include: Global Affairs; Governance; Law & Justice; Security & Defense; Crime & Public Safety; Economy; Finance & Markets; Business & Industry; Labor & Demographics; Infrastructure; Energy; Environment; Science; Health; Technology; Media; Sports; Culture; Lifestyle; Incidents.

5. Domain → Feed Category Mapping

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

This mapping allows accuracy without inflating the UI.

Internal Domain	Feed Category
Global Affairs	World
Governance (US)	United States
Crime (City)	Local
Markets / Finance	Business
Technology	Technology
Health & Medicine	Health
Science & Research	Science

6. Canonical Tagging Axes

All articles receive structured tags across fixed, non-evaluative axes: Internal Domain; Geography; Actors; Action Type; Temporal State; Evidence State; Impact Scope; Quantitative Signals.

Tags may never imply urgency, importance, moral judgment, or political leaning.

7. Categorization Pipeline

Inputs include raw article text, entities, geographic signals, and NTRL-generated neutral outputs.

Decision order: Geographic rules → Domain classification → Domain mapping → Conflict resolution → Confidence scoring → Review flagging.

8. Feed Assembly Rules

The feed is assembled based on user-selected categories.

Within each category, articles are sorted by time.

If a selected category has no qualifying articles, it is silently omitted.

9. User Settings: Topic Selection

Users may toggle any category on or off and reorder categories.

Categories without updates are hidden automatically.

10. Admin Configuration

Admins may manage category definitions, domain mappings, quotas, and review low-confidence classifications.

All overrides require an audit trail.

11. Backend Data Contract

Required fields include feed_category, internal_domain, tags, categorization_confidence, needs_review, override_history, and taxonomy_version.

No silent changes are permitted.

12. Quality & Monitoring

Metrics track misclassification rate, override frequency, domain coverage gaps, and local accuracy.

Engagement metrics are explicitly excluded.

13. Design Outcome

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

It is designed for long-term production use.