SEO Research · Deep Dive

Traditional search optimization — ranking factors, changes since 2023, myths ruled out

Source: Google Search Central · Helpful Content updates 2023-2026 · Backlinko + Ahrefs studies

1 · RANKING FACTORS (2026, in order of effect size)

Content relevance + depth

▸ Query-intent match is #1 ranking signal
▸ Semantic depth > keyword density
▸ Comprehensive coverage of topic + sub-topics
▸ Original analysis beats summarized news
▸ Long-form (1,500-3,000w) for competitive terms
▸ E-E-A-T overlay on top of relevance

Verdict: write deep, write original, write what AI cannot

Cognilium edge: 50+ projects → first-hand engineering content

Highest-impact lever in 2026

Backlinks (still #1 off-page)

▸ Backlinks remain top-3 signal (per Google Spam team)
▸ Quality > quantity: 1 high-DA > 50 spam
▸ Topical relevance of linker matters
▸ Anchor text distribution (under 20% exact-match)
▸ Editorial > directory links
▸ Recency of backlink (new = stronger)
▸ Growth pattern matters (natural 15-20/mo)

Earn, never trade or buy (penalty)

Cognilium plan: research drops + guest posts

Technical SEO (Core Web Vitals)

▸ LCP < 2.5s (Largest Contentful Paint)
▸ INP < 200ms (Interaction to Next Paint)
▸ CLS < 0.1 (Cumulative Layout Shift)
▸ Mobile usability: passable on every page
▸ HTTPS: required (since 2018)
▸ Crawl budget: efficient internal linking
▸ JS render: progressive + SSR (we do)

Core Web Vitals are pass/fail, not ranked

Cognilium status: all green via Next.js 15 + Vercel

2 · WHAT CHANGED SINCE 2023 (the deltas that broke old playbooks)

AI Overviews everywhere

▸ ~40% of queries now show AI Overview
▸ AIO occupies prime real estate (above results)
▸ Clicks for top-3 organic drop 30-50% on AIO queries
▸ Snippet/answer = brand visibility play
▸ Strategy: optimize for being IN the AIO

Implication: track impressions + brand search, not just clicks

Biggest 2024-2025 SERP shift

Helpful Content System merged

▸ March 2024: HCU folded into core ranking
▸ Site-wide quality signal (was page-level)
▸ One bad section can drag the whole domain
▸ Forced cleanup of thin content
▸ Authorship + first-hand experience signals

Implication: prune low-quality pages aggressively

Why we archived 20 under-target blogs

Spam policies expanded

▸ Site Reputation Abuse (May 2024): kills "rented" subfolders
▸ Scaled Content Abuse: mass AI-generated content penalized
▸ Expired Domain Abuse: buying old domains for SEO
▸ INP replaced FID in Core Web Vitals
▸ Hreflang stricter for multi-regional

Implication: AI-mass = penalty, not shortcut

Why we do not auto-generate content

3 · MYTHS WE'VE RULED OUT

Myth: "Keyword density target"

→ DEBUNKED. Semantic relevance and topical coverage matter, not raw keyword counts. Stuffing penalized as keyword abuse.

Myth: "Exact-match anchor text wins"

→ DEBUNKED. Over-optimization penalty triggers above ~20% exact-match ratio. Diverse, descriptive anchors win.

Myth: "Word count = rank"

→ MISLEADING. Depth matters; length is a proxy. 800-word post that fully answers > 3,000-word fluff piece.

Myth: "Bounce rate / dwell time from GA4 affects ranking"

→ DEBUNKED publicly by Google. They do not use GA4 data for ranking. User signals come from SERP click behavior.

Myth: "Domain age strongly boosts rank"

→ OVERSTATED. Minor factor. Relevance + authority compound far faster than age.

Myth: "Daily publishing boosts rank"

→ ~~ONLY TRUE FOR NEWS. Evergreen content needs cadence for trust, not for re-ranking. Quality > velocity for blogs.~~

Myth: "Submit to 100 directories"

→ COUNTER-PRODUCTIVE. Quality directories (Clutch, Built In, Crunchbase) help. Spam directories trigger devaluation.

AEO Research · Answer Engine Optimization

Featured snippets, voice search, AI Overviews input — capturing "position 0"

Source: Google Search Central · Backlinko snippet studies · Stat counters 2026

1 · THE ANSWER-FIRST STRUCTURE (rules for snippet capture)

Question-format H2 hierarchy

- ▸ "What is X?" / "How does X work?" / "Why does X fail?"
- ▸ Each H2 maps to a search query
- ▸ Direct answer in the FIRST 40 words
- ▸ 50-150 character definition for snippet capture
- ▸ Then expand into the WHY and HOW
- ▸ Conclude with example or table
- ▸ Each H2 section ≈ 300-500 words (chunk-friendly)

Rule: lead with the answer, not the setup

Master pattern for AEO + GEO

Featured-snippet capture rate ~3× higher with this

Format-by-query-type

- ▸ "What is __" → paragraph snippet (50-150 chars)
- ▸ "How to __" → numbered list (5-8 steps)
- ▸ "Best __ for __" → bullet list with headers
- ▸ "__ vs __" → comparison table
- ▸ "Difference between" → side-by-side
- ▸ Date queries → year clearly visible
- ▸ Stats queries → big number + source

Match the format to the query intent

Google varies these in 2026

Google extracts these as-is into the snippet

FAQPage schema (under-used)

- ▸ Mark Q+A blocks with FAQPage JSON-LD
- ▸ Each Q is a featured-snippet candidate
- ▸ Voice assistants pull from FAQPage too
- ▸ "People Also Ask" (PAA) sources from FAQs
- ▸ Schema validates with Rich Results Test
- ▸ Don't hide answers behind JavaScript
- ▸ Don't use for promotional content (banned)

One FAQPage cluster = 5-10 snippet wins

High-impact, low-effort win

Cognilium uses on every long-form blog

2 · VOICE-SEARCH + ZERO-CLICK REALITY

Zero-click majority

- ▸ ~50% of all Google queries are zero-click
- ▸ User sees answer in SERP, doesn't need to click
- ▸ Brand visibility IS the win
- ▸ Track: impressions, brand search lift, share of voice
- ▸ Not just: clicks, CTR, sessions
- ▸ Featured snippet = position 0 = top brand surface

Optimize for the answer, not the click

Mental model shift since 2024

Voice + conversational queries

- ▸ Voice queries average 7+ words (vs 2-3 typed)
- ▸ Natural-language phrasing wins
- ▸ Active voice + declarative sentences extract better
- ▸ Alexa, Siri, Google Assistant all pull position 0
- ▸ "Near me" + entity queries favor NAP-consistent businesses
- ▸ Long-tail conversational queries less competitive

Write like you speak, not like a press release

Underserved channel, growing

AEO ↔ AI Overviews overlap

- ▸ A page in featured-snippet position is ~3× more likely to appear in AIO
- ▸ Position 0 content is the seed for AIO answers
- ▸ Tables, lists, structured definitions get reproduced verbatim
- ▸ DefinedTerm schema → AIO citation panels
- ▸ Code snippets reproduced if cleanly fenced
- ▸ Author + publish date shown in AIO citation chip

AEO investment is the bridge from SEO → GEO

Why AEO is the highest-ROI 2026 lever

GEO Research · Princeton Paper Deep Dive

Generative Engine Optimization — the only peer-reviewed study of LLM citation behavior

Source: Aggarwal et al. "GEO: Generative Engine Optimization" Princeton 2024-2026 · arXiv:2311.09735

1 · THE FOUR VALIDATED LIFT SIGNALS

Citations → +30%

▸ Add inline citations to authoritative sources
▸ "(Author, 2026)" or "[1]" format
▸ Primary sources (papers, vendor docs)
▸ Not blog-to-blog circles
▸ External links with rel=external for clarity
▸ Works for ALL major engines

Implementation: cite at least 3-5 primary sources per long-form blog

+30% citation rate baseline

Statistics → +32%

▸ Concrete numbers over hand-waves
▸ "5.2× faster" not "much faster"
▸ Cost in $, latency in ms, accuracy in %
▸ Source-attributed: "in our X-run benchmark"
▸ Tables with numeric data extract beautifully
▸ Round numbers feel suspicious; precision wins

Implementation: every claim gets a number or it gets cut

+32% — second-strongest signal

Quotations → +41% (STRONGEST)

▸ Direct quotes from authoritative sources
▸ Original quotes from Mudassir/Ali (E-E-A-T)
▸ Customer quotes in case studies
▸ <blockquote> semantic markup
▸ Attributed: "—Author, Title, Date"
▸ LLMs reproduce quotes verbatim with attribution

Implementation: every blog has 1+ original Mudassir/Ali quote

+41% — single highest-impact lever

Fluency → +28%

▸ Active voice · short clauses
▸ No filler ("in order to" → "to")
▸ Concrete nouns over abstractions
▸ Readability target: Flesch ≥ 50
▸ Avoid passive constructions
▸ One idea per sentence

Implementation: edit aggressively; first drafts always too wordy

+28% — easiest to improve, often skipped

2 · CHUNK-FRIENDLY STRUCTURE (what LLMs actually retrieve)

How LLM retrievers see content

▸ Content is chunked into 300-500 token windows
▸ Each chunk retrieved independently
▸ Chunks LOSE parent context — must self-contain
▸ Each H2 should answer ONE question completely
▸ "As discussed above" — useless in chunk-land
▸ Repeat key entities every ~300 words

Write as if every section travels alone

Architectural requirement, not style preference

Entity density

▸ Name versions, vendors, libraries, papers explicitly
▸ "GPT-4o" not "the OpenAI model"
▸ "Anthropic Claude 4.7" not "the AI"
▸ "Princeton GEO paper (Aggarwal et al)"
▸ Capitalized proper nouns extract cleanly
▸ DefinedTerm schema for glossary entities

Implementation: 5+ named entities per H2 section

Entity-rich text = AI-citable

Anti-patterns (HARM citation rate)

▸ Keyword stuffing — Princeton: NEGATIVE lift
▸ Clever metaphors — confuse extraction
▸ Over-clever transitions — break chunk independence
▸ Vague hand-wave claims (the banned-words list)
▸ Heavy JS (LLMs skip JS-rendered content)
▸ Hiding text behind interaction

SEO tricks BACKFIRE for GEO

GEO ≠ SEO; opposite of stuffing

3 · WHY THIS MATTERS — the math

The Princeton lifts COMPOUND, they don't average.

Apply all 4 signals to a single article (citations + stats + quotes + fluency):

▸ Baseline citation rate: 100 (index)
▸ Add citations: × 1.30 → 130
▸ Add statistics: × 1.32 → 171
▸ Add quotations: × 1.41 → 241
▸ Improve fluency: × 1.28 → 308

Result: 3.08× baseline citation rate per article — at zero distribution cost.

This is why content discipline matters more than distribution volume.

One well-written deep post citing 5 papers, with 3 quotes, with 10 specific numbers, and tight prose will out-cite 5 generic posts on the same topic in AI engines.

This validates Cognilium's "less, deeper" content strategy over "more, shallower".

It also validates why we archived the 20 under-target blogs (each fell short on multiple signals).

E-E-A-T Research · Trust Framework

Experience · Expertise · Authoritativeness · Trustworthiness — Google's quality-rater lens

Source: Google Search Quality Rater Guidelines (Dec 2022 update) + Helpful Content System docs

1 · THE FOUR LETTERS (each is a distinct signal cluster)

Experience (E — added Dec 2022)

"Has the creator demonstrably USED the thing they're writing about?"

- ▸ First-person prose: "When we built __"
- ▸ Specific failures + recovery: "Here's what broke"
- ▸ Original screenshots from real deployments
- ▸ Real numbers from real client work (NDA-safe)
- ▸ Anti-AI-generated content signal #1
- ▸ The 2022 addition was Google's response to AI mass-gen

Cognilium edge: 50+ shipped projects → first-hand evidence

Every blog has at least one "we did this" anecdote

The newest E, hardest to fake

Expertise

"Does the creator have the right credentials for this topic?"

- ▸ Author bio on every article (Mudassir / Ali)
- ▸ LinkedIn + sameAs in Person schema
- ▸ Specific tech mastery shown (not just claimed)
- ▸ Speaking, podcasts, original research as proof
- ▸ For YMYL topics: formal credentials required
- ▸ Authority/topical expertise > generalist

Status: bios live; conferences/podcasts Q3 2026

Person schema with jobTitle, alumniOf, sameAs

The credentials letter

Authoritativeness

"Is the creator recognized as a leading voice?"

- ▸ Quality backlinks from authority publications
- ▸ Citations in other writers' work
- ▸ Press mentions (TechCrunch, InfoQ)
- ▸ Wikipedia / Wikidata entity (long-term goal)
- ▸ Conference talks, podcast guest spots
- ▸ Industry awards (where applicable)

Status: building authority loops Q3-Q4 2026

Research drops → press → backlinks compounding

External validation letter

Trustworthiness — MOST IMPORTANT

"Can readers safely act on this information?"

- ▸ HTTPS · NAP consistency · privacy + terms pages
- ▸ No deceptive claims, no fake reviews
- ▸ Author + publisher schema (provenance)
- ▸ Editorial / correction policy disclosed
- ▸ Verified-stats discipline (banned-words list)
- ▸ Original research disclosed methodology

Cognilium: NO unverified claims · NO fake stats

Single source of truth in cta-config.ts

Per Google: "Trust is the most important of the four"

2 · COGNILIUM E-E-A-T APPLICATION (specific tactics, not generic advice)

Experience — first-hand evidence is the hardest moat to copy

- ▸ Every long-form blog includes "When we built ___ for a client" anecdote
- ▸ Real failure modes from 50+ projects (anonymized, NDA-safe)
- ▸ Screenshots of actual benchmarks, not stock images
- ▸ Specific tools + versions named, not "we used a vector DB"

Expertise — let the work prove it

- ▸ Author profiles with LinkedIn + Twitter + GitHub linked
- ▸ Tech mastery demonstrated through code, not buzzwords
- ▸ Speaking gigs and podcast guest spots tracked (Q3+)
- ▸ Person schema with explicit credentials

Authoritativeness — the slow compounding letter

- ▸ Research drops earn backlinks (target: 30-100 per major drop)
- ▸ Guest posts on InfoQ, The New Stack, Towards Data Science
- ▸ Podcast guest appearances (Latent Space, Pragmatic Engineer)
- ▸ Conference talks (AI Engineer Summit, KubeCon AI)

Trustworthiness — the don't-fuck-it-up letter

- ▸ Verified-stats discipline: 50+ projects, 96% satisfaction, etc.
- ▸ Banned-words list enforced via lint script (qa-tech-news.ts)
- ▸ Schema.org Organization + Person markup with sameAs links
- ▸ Editorial policy disclosed (corrections welcomed via contact)

Topical Authority Research · Hub-and-Spoke Cluster Strategy

Why depth-first publishing dominates scattered "30 random posts"

Source: Backlinko cluster studies · SEMrush topical authority research · Cognilium 5-pillar map

1 · HUB-AND-SPOKE THEORY

The architecture

▸ PILLAR PAGE (3,000+ words, broad survey of topic)
↑ linked from every spoke in the cluster
▸ SPOKE PAGES (1,500w each, specific angle)
↔ link to pillar + 3-5 sibling spokes
▸ CLUSTER HUB page (CollectionPage schema)
lists all spokes for human + crawler navigation
▸ GLOSSARY terms cross-link as entity anchors

Every spoke is independently rankable
Internal links pass topical PageRank
The shape that AI engines reward

Why it beats scattered publishing

30 random posts ≠ 30-post cluster

▸ Random posts: each ranks alone, no topical signal
▸ Cluster: posts reinforce each other's topical authority
▸ Backlinks naturally accrue to pillar, flow to spokes
▸ Long-tail keyword coverage compounds within cluster
▸ AI engines see "this site is the authority on X"
▸ Internal-link graph encodes topical relationships

Same effort: ~5× higher impact in a cluster
Why we picked 5 pillars, not 20 topics

Depth target (industry data)

▸ 30-50 posts/cluster → dominates AI Overviews
▸ 10-30 posts → competitive in long-tail
▸ < 10 posts → mostly invisible in AIO
▸ Pillar depth ranks 5-7× higher than scattered equivalents
▸ Cognilium current: 1-3 posts per cluster
▸ Target by Q3 2026: 12 posts per cluster minimum
▸ Long-term: 30+ posts in flagship cluster

Investment: depth-first, not breadth-first

Cognilium 90-day goal

2 · COGNILIUM 5-PILLAR MAP (selected from research)

1 · GraphRAG / Agentic KG (FLAGSHIP)

Brand-strategy decision: June 2026

▸ Pillar: "GraphRAG: The Complete 2026 Guide"
currently DRAFT (3,700w, 106 blocks)
▸ Spokes (live): entity-resolution, graph-rot
▸ Cluster hub: enterprise-graphrag-knowledge-systems
▸ Target depth: 12-15 posts by Q3

Why flagship: highest differentiation from competitors,
highest enterprise relevance, highest AI-engine demand
Cognilium's "we are THE GraphRAG company"

2 · Multi-Agent Systems

Cognilium core competency

▸ Live spokes: 8-stage docint pipeline,
agent-pipeline-failure recovery (DynamoDB+SQS)
▸ Cluster: aws-google-agent-frameworks
▸ Topics: supervisor patterns, tool registration,
failure recovery, multi-tenant routing
▸ Target: 12 posts by Q3

Why: AWS Bedrock + Google ADK + enterprise
Production agent expertise

3 · LLMOps + Evaluation

Production-readiness layer

▸ Cluster: production-llmops-evaluation
▸ Topics: cost optimization, latency, drift,
guardrails, bias detection, eval frameworks
▸ Archived spoke (rebuild target): bias-detection-multi-agent
▸ Target: 12 posts by Q3

Why: enterprises need this layer; few cover it well

The boring-but-critical pillar

4 · Voice AI / Enterprise

Specialty niche, low coverage

▸ Cluster: enterprise-voice-ai
▸ Topics: latency, barge-in, transcription accuracy,
domain-vocabulary grounding, multi-language
▸ Archived spoke (rebuild target): domain-vocab grounding
▸ Target: 8-10 posts by Q3

Why: high-value vertical, less competitive,
matches Cognilium's voice AI client work
Underserved cluster

5 · Enterprise Foundations

The "table stakes" pillar

▸ Clusters: enterprise-document-ai, graph-rot
▸ Topics: data quality, zero-trust, multi-tenant,
smart-category routing, document intelligence
▸ Archived: smart-category-routing, zero-trust-multi-tenant
▸ Target: 12 posts by Q3

Why: enterprise readiness — table stakes for any deal

Foundation layer

Cross-cluster glossary (the connective tissue)

GLOSSARY = entity anchors across all 5 pillars

▸ /glossary · 50 terms · 7 categories
▸ Every blog post links to glossary entries inline
▸ DefinedTermSet + DefinedTerm schema
▸ Each term is independently rank-able
▸ Glossary IS a citation magnet for AI engines

Cross-cluster cohesion via entity references
High-leverage SEO + GEO infrastructure
The site-wide entity layer

AI Crawlers · Per-Engine Behavior + Market Share

Which bots, what they want, where they cite — and how to optimize for each

Source: Cloudflare AI crawler reports · Andreessen Horowitz AI market data · vendor robots-policy pages

1 · MARKET SHARE 2026 (where the AI traffic actually comes from)

The Big 4 (95%+ combined)

ChatGPT (OpenAI) ..... 64.5% ← was 86.7% in 2025
Perplexity ..... 17% ← fastest-growing
Gemini (Google) ..... 9% ← AI Overviews-tied
Claude (Anthropic) ..... 5% ← growing Q4 2026

Copilot · Mistral · Others ... 4.5%

The shift: ChatGPT lost 22 points in 12 months
Diversification is mandatory, not optional

No single engine to optimize for

AI traffic conversion math

AI-referred conversion: 10.21%
vs organic: 3-4%
vs paid: 1-2%

▸ Pre-qualified by chatbot pre-filter
▸ User has high intent (specific question)
▸ AI-recommended = implicit endorsement
▸ Highest LTV-per-visitor channel in 2026

One AI referral ≈ 3× value of one organic visitor
Why GEO is the highest-ROI 2026 investment

Why diversification matters

Risk of monoculture optimization:

▸ ChatGPT was 86.7% (2025) → 64.5% (2026)
▸ Future could see further shift to Perplexity
▸ Each engine has different citation behavior
▸ Optimizing for one ≠ optimizing for all
▸ Princeton signals universal, but tone varies

Goal by Q3 2026: ≥35% AI mentions from
non-ChatGPT engines
Strategic resilience

2 · PER-ENGINE CITATION BEHAVIOR (how each one cites you)

ChatGPT (GPTBot + ChatGPT-User)

OpenAI · 64.5% share

▸ GPTBot = training crawl (corpus-building)
▸ ChatGPT-User = live retrieval (web mode)
▸ Cites 1-3 sources per answer (conservative)
▸ Favors authoritative + recent (within 6mo)
▸ "Search the web" mode pulls from live index
▸ Knowledge cutoff: training data + live retrieval

Optimize: deep, authoritative, recent
Maintain dateModified discipline

Perplexity (PerplexityBot) — HIGHEST VALUE

17% share, fastest-growing, generous citer

▸ Live-search-first (every query crawls fresh)
▸ Cites 5-10 sources per answer
▸ Pulls quotes verbatim with attribution
▸ Users click sources at 5× ChatGPT rate
▸ AI-referred conversion highest from Perplexity
▸ Citation appears as prominent UI chip

Optimize: quote-rich, recent, comprehensive
Highest ROI per optimization hour

Claude (ClaudeBot + Claude-User)

Anthropic · 5% share, growing fast

▸ Cites only when user explicitly asks
▸ Heavily values factual accuracy + nuance
▸ Strong on technical / engineering content
▸ Pulls long-form analytical posts more than short
▸ Cognilium audience overlap: highest of all engines
▸ Citation patterns: thorough, often longest

Optimize: long-form, technical, balanced
This is Cognilium's natural audience

Gemini (Google-Extended + Gemini)

Google · 9% share

▸ Google-Extended = separate from Googlebot (SEO bot)
▸ Opt-in for Gemini training corpus
▸ Gemini citations appear in AI Overviews
▸ Tight integration with Google Search
▸ Strong SEO content tends to also win Gemini
▸ AIO sources visible in citation chips

Optimize: SEO best-practices double as Gemini wins
Two-for-one optimization

The long tail (10+ smaller crawlers)

▸ CCBot (Common Crawl) — training seed for many LLMs
▸ Applebot-Extended — Apple Intelligence + Siri
▸ Bytespider — TikTok / Doubao (Chinese LLMs)
▸ Meta-ExternalAgent — Meta AI (Llama)
▸ Amazonbot — Alexa + Bedrock training
▸ cohere-ai — Cohere enterprise LLM
▸ YouBot · Phind · Andibot — smaller AI search
▸ Mistral · MistralAI-User — French AI lab

Cognilium policy: allow ALL · no upside to blocking
Every crawler is a citation chance

/llms.txt manifest standard

Proposed Sep 2024 by Jeremy Howard (Answer.AI)

▸ /llms.txt = curated TOC of site content for LLMs
▸ /llms-full.txt = full content corpus, single file
▸ Adopted by Anthropic docs, Stripe, Vercel
▸ Not yet a formal Google standard
▸ Hand-curated control over what LLMs prioritize
▸ Cognilium: live since 2026-06

Early-mover advantage
Generated via scripts/generate-llms*.ts

News SEO Research · Google News + Discover + Top Stories

A distinct discipline from evergreen SEO — different signals, different cadence, different funnel

Source: Google News Publisher Center docs · Discover guidelines · Wayfair 2018 case study

1 · DISCOVER ELIGIBILITY CHECKLIST

Content requirements

▸ Original reporting, NOT republished content
▸ Hi-res featured image (≥1200px wide, 16:9)
▸ NewsArticle JSON-LD schema
▸ Author byline with Person schema
▸ datePublished + dateModified (real, not faked)
▸ Substantive content (800w+ minimum)
▸ News-style: timely angle, not evergreen

Original is the keyword — aggregation fails

Discover wants exclusive content

Site-level signals

▸ Established cadence (5+ posts/week minimum)
▸ Google Publisher Center registration approved
▸ news-sitemap.xml with 48-hour rolling window
▸ Topical authority on the subject
▸ Mobile-optimized (Discover IS mobile)
▸ Fast LCP (Discover penalizes slow load)
▸ HTTPS + secure

Cadence is non-negotiable — skip a week → drop

Trust signals: site-wide, not per-page

What kills Discover eligibility

▸ AI-mass-generated content (Scaled Content Abuse)
▸ Republished press releases
▸ Aggressive paywalls before content
▸ Misleading clickbait headlines
▸ Drifting dateModified to fake freshness (Wayfair)
▸ Site Reputation Abuse (rented subfolders)
▸ Manual penalties take 6-18 months to clear

One mistake = months of recovery

Trust takes years to build, days to lose

2 · NEWS GENRES TAXONOMY (news:genres — pick the right one)

Genre tiers (highest → lowest trust)

▸ (untagged) — straight reporting (highest)
▸ Blog — opinion + commentary
▸ Opinion — analytical opinion
▸ OpEd — explicitly opinion
▸ Satire — humorous (rarely use)
▸ PressRelease — announcement (lower trust)
▸ UserGenerated — community content (lowest)

Cognilium tech-news = Blog + Opinion (engineering analysis)

Google's official news classification
Use the tag explicitly in news-sitemap.xml

Wayfair-style drift lesson

2018 case: Wayfair touched dateModified daily to fake freshness.
Discover de-indexed them for ~6 months.

▸ Updating < 5% of content but bumping date = penalty
▸ "Updated" labels must reflect REAL changes (≥15%)
▸ Quarterly evergreen refresh = OK
▸ Daily date-touch = NOT OK
▸ Google verifies content delta against last crawl

Implementation: check-last-updated.ts script

The discipline that prevents disaster
compares file mtime vs git history

Top Stories vs Discover vs News tab

▸ Top Stories: in main SERP, query-driven
needs: news-sitemap, NewsArticle schema, GPC approval
▸ Discover: mobile home feed, no query
needs: image quality, topical authority, engagement
▸ News tab: dedicated news SERP
needs: Publisher Center registration approved

▸ All three flow from the same content + schema
▸ Multi-surface presence compounds

Google's algorithm picks which surface(s) per query + user — we don't control the selection algorithm

Goal: be in all 3 surfaces

Distribution Strategy · 4-Tier Channel Hierarchy

Which channels matter, in what order, why — investment matched to ROI window

Source: docs/PLATFORMS.md (130+ platform inventory) · distribution research 2026

1 · THE 4-TIER MODEL (do Tier 1 first, always)

TIER 1 · Owned (do FIRST, ALWAYS)

No platform risk · full control

▸ Site (cognilium.ai) — the only canonical asset
▸ Newsletter (Resend) — only owned audience
▸ RSS feeds (3) — discovery surface
▸ Sitemaps (13) — search input
▸ /llms.txt manifest — LLM input

Investment: 100% before any other tier
Cost: build infrastructure once, runs forever
Risk: zero (we own it)

Status: LIVE since 2026-06

Skipping Tier 1 = building on rented land

The foundation, non-negotiable

TIER 2 · Auto-mirror (after Tier 1)

Push automatically · low ongoing cost

▸ Dev.to · canonical-safe · LIVE
▸ Resend Broadcast · email to subs · LIVE
▸ Bluesky · social discovery · LIVE
▸ LinkedIn (org) · PAUSED (API gating)
▸ Hashnode · DORMANT (paywall)
▸ Tumblr · WIP (scaffolding)
▸ X / Twitter · OPTED OUT

Investment: build once, run forever

Each mirror module = ~1 day setup, ∞ runtime
Add new = drop one file in src/lib/mirrors/

Force multiplier on Tier 1 content

TIER 3 · Pull-based aggregators

Zero ongoing cost · they come to you

▸ Substack — pulls /all-content/rss · LIVE
▸ Flipboard — auto-mag · LIVE
▸ LinkedIn Newsletter — org page · LIVE
▸ Google Publisher Center → News surface · LIVE
▸ Feedly · Inoreader · passive
▸ Mastodon-RSS bots · passive

Investment: feed quality (RSS hygiene)

They show up because RSS is good
Medium SKIPPED (cannibalizes canonical)

Multiplier on Tier 1 RSS infrastructure

TIER 4 · Manual community (top 5% only)

High cost per post · only for best content

▸ Hacker News — genuinely interesting longform
▸ Reddit — value-first only (anti-promo culture)
▸ Lobsters — invite-only · highest signal
▸ Product Hunt — tool/asset launches
▸ Hugging Face / Discord — community presence
▸ Quora · Stack Overflow · GitHub — ongoing presence

Investment: ~1 hour per top-5% post

Don't spray content here — gets you banned
Submit thoughtfully, engage in comments

Hand-curated, never automated

2 · INVESTMENT ALLOCATION (where time + money go)

Time allocation per week (typical 2026 cadence):

▸ Tier 1 — Content creation (writing news + blogs): 60% of content time
▸ Tier 2 — Auto-mirror maintenance: 0% (runs itself; only fix when broken)
▸ Tier 3 — Aggregator quality (RSS, Publisher Center): 5% (audits + tweaks)
▸ Tier 4 — Community engagement (HN, Reddit, comments): 10-15% (top posts only)
▸ Newsletter operations: 5%
▸ Audience research + strategy: 10%

Why this allocation works:

▸ Most leverage is in Tier 1 content (what feeds everything else)
▸ Tier 2 is build-once-run-forever (already built)
▸ Tier 3 doesn't need work if Tier 1 is good
▸ Tier 4 is the diminishing-returns layer

Common mistakes:

▸ Inverting the hierarchy: 50% time on Twitter, 10% on site — backwards
▸ Skipping Tier 1 to chase Tier 4 trends — content has no home
▸ Auto-cross-posting to Medium — cannibalizes canonical
▸ Buying directory listings (Tier 4 spam) — penalty risk