LangInMotion®

WHITEPAPER · HEALTHCARE PAYERS

Denial Letters: The Unchecked Paragraph

Denial letters misstate the criteria they cite. Courts have documented it for a decade. AI is likely to scale it.

LANGINMOTION® · VERSION 1.0 · 21 AUGUST 2026 · LIM-WP-2026-01

Executive summary

A denial notice is one of the most heavily regulated documents a health plan produces. The template is prescribed down to the font size — for CMS model notices, *"it is not permissible to omit any standardized language, nor alter the template, including font size, without CMS approval."*[7] Who may sign it is controlled by accreditation and, increasingly, by state law. Whether it went out on time is audited.

One paragraph escapes all of it. The clinical rationale — the passage that says *why*, and cites the criteria applied — is written free-hand. The governing standard is that it be *"sufficiently specific to enable the enrollee to determine the basis for appeal."*[8] Specific. Not *accurate to the criteria it cites*.

> **The boilerplate is regulated down to the font size. The paragraph that actually says why is written free-hand, and nothing checks it against the criteria it cites.**

That gap is not hypothetical. Federal courts have repeatedly found denial letters that restate a plan's own criteria incorrectly: a disjunction reduced to one branch, an **and** rewritten as an **or**, a qualifying criterion used as grounds for refusal. In one case the plan's own second-level reviewer conceded the error in writing.[11] These letters were written by named physicians between 2015 and 2022 — **before** LLM-generated correspondence was in play.[9][10][11]

So this is not a new risk introduced by AI. It is an old, documented, litigated one that **AI is likely to scale**: raising the rate, making the same error again and again rather than one-off ones, and removing any possibility of full human review at volume.

Meanwhile the pressure is rising from both sides. CMS-0057-F, in force since 1 January 2026, requires impacted payers to give a **specific reason** for each denial rather than a generic one.[6] And a mature vendor ecosystem now helps *providers* generate appeals at near-zero marginal cost — appeals that attack precisely this paragraph.[12]

This paper sets out what the record shows, why every existing control misses it, and what the missing control looks like.

PART I

The problem is documented, not theoretical

Denials are wrong at a rate that is difficult to explain

FINDING	FIGURE	SOURCE
MA skilled-nursing admission denials **overturned on appeal**	**95%**	[1]
…the same, for one large processor	97%	[1]
SNF denials that were **appealed at all**	18%	[1]
Long-term acute care / inpatient rehab denials overturned	36% / 43%	[2]
Denied MA prior-auth requests that **met Medicare coverage rules**	**13%**	[3]
HealthCare.gov in-network claims denied, 2024	19%	[4]
Marketplace denials **ever appealed**	**<1%**	[4]

The 95% overturn rate is the strongest single figure available. A first-pass decision reversed nineteen times out of twenty is not a close call being adjudicated. And it must be read against the 18% appeal rate: if 95% of the appealed 18% were wrong, the unappealed 82% is uncosted exposure. [1]

These figures describe *decisions*. What follows describes the *letters* — and the two are connected, because the letter is where the decision's reasoning becomes reviewable.

Four errors, each from a court case

Each of the following comes from a federal court's own reading of a plan's criteria alongside the letter that applied them. None is our characterisation.

Dropped disjunct. The governing guideline authorised residential treatment where other levels were inappropriate **and** the patient either posed a danger to self or others **or** suffered a moderately severe psychiatric disorder causing serious dysfunction. The letter rested on risk of harm alone, ignoring the second branch — and three of the four continuation criteria. *T.E. v. Anthem*, Sixth Circuit, precedential. [9]

Conjunction rewritten as disjunction. The plan's guideline required symptoms representing a deterioration from usual status **and** self-injurious or risk-taking behaviours **and** inability to be managed outside a 24-hour setting.

The letter's own restatement rendered this as *"1) their behaviours have worsened **or** their actions risk serious harm."* An **and** became an **or** in the plan's description of its own criteria. [10]

Determination on an unstated ground. In the same letter, the operative rationale was a single sentence — *"[t]he information we have does not show 24 hour structured care is needed"* — resting on one conjunct while the criteria required several. The court treated the gap as fatal. [10]

Exclusion criterion inverted into a denial ground. A criterion phrased in the negative — *"the member is **not** in imminent or current risk of harm"* — functions as an admission gate distinguishing residential care from *more* restrictive inpatient care. The reviewer used it as a reason to deny. The plan's own second-level reviewer conceded in writing: *"You are correct that by itself it is not grounds for a non-coverage determination."*[11]

Two properties of this evidence worth noting

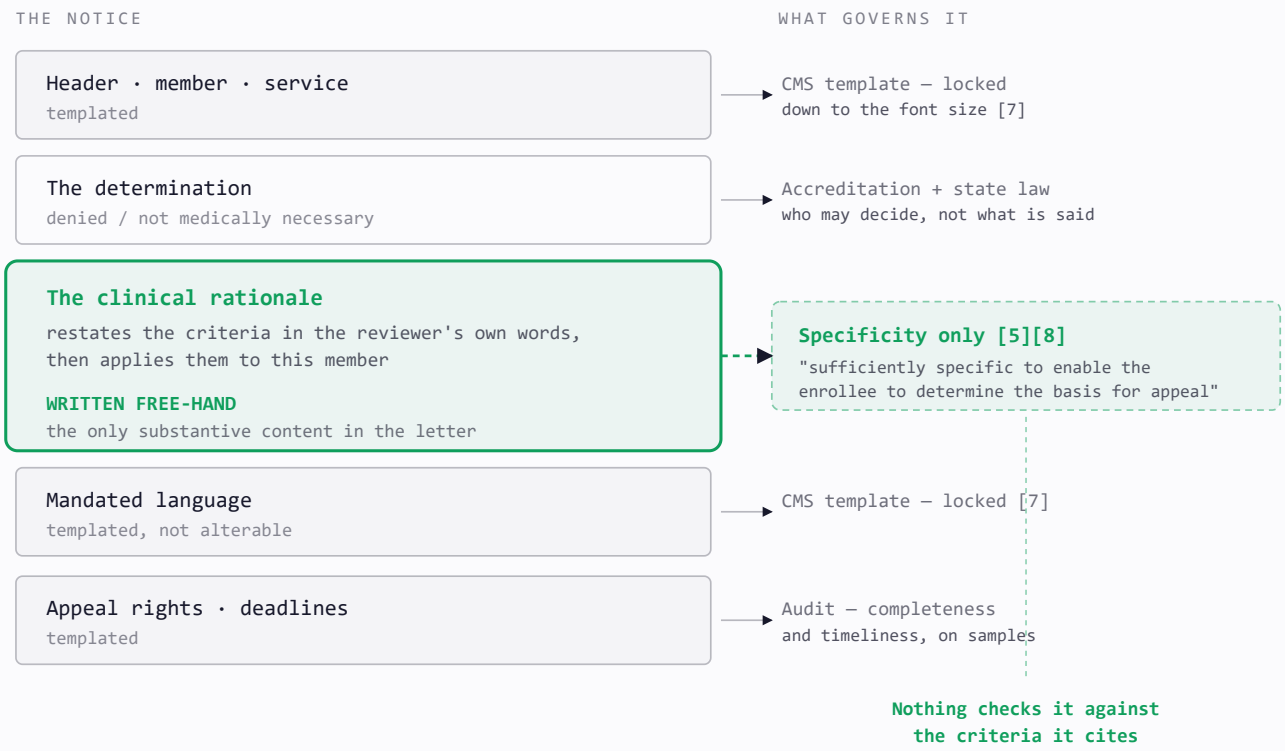
It is human-authored. Each of these letters was written and signed by a named, licensed physician — in 2015 in one matter, [11] and in the *Robert D.* matter a few years later[10] — years before LLM correspondence. **The failure mode does not require AI and is not caused by it.**

The rationale was always composed, never copied. In these letters the plan restates its criteria *in its own words* before applying them. This matters for any claim about what automation changed: there was no verbatim-copy safeguard that generation removed. Paraphrase was always the mechanism, and paraphrase was already where the error lived.

PART II

Where the paragraph comes from

Payer-published provider manuals describe a consistent path. [13][14]



The header, mandated language and appeal-rights sections are governed by a locked template. Who signs is governed by accreditation and state law. Timeliness is audited. **The clinical rationale — the only substantive content in the letter — is written free-hand.**

Steps 1 through 3 are the controlled part: who screens, who may deny, what triggers escalation. Step 5's template is prescribed. [7]

Step 4 is the whole exposure. It is the only substantive content in the letter that no control examines for accuracy against the source it cites.

PART III

Four controls, one shared blind spot

The template is locked. CMS model notices carry mandated language and may not be altered without regional-office approval.[7]

The clinical rationale is free text. The standard is specificity sufficient to let the enrollee determine a basis for appeal[8] — not fidelity to the criteria cited.

The federal requirement works the same way. ERISA's claims-procedure regulation obliges a notice to state *"[t]he specific reason or reasons for the adverse determination"* and give *"[r]eference to the specific plan provisions on which the determination is based."*[5] Both halves are satisfied by a letter that names the right provision and then **describes it inaccurately**. The obligation is to cite and to be specific; nothing in it reaches whether the description matches the provision named.

Authority is controlled, not content. A nurse may not issue a medical-necessity denial; the case escalates to a medical director.[13][14] State AI statutes likewise regulate *who* decides, not *what the letter says*.

Audits test completeness and timeliness, on samples. CMS program audits and state market-conduct exams check that required elements are present and deadlines met.

Every one of these answers *did the letter contain the required parts*, *was it decided by the right person*, and *was it on time*. **None answers whether the rationale accurately represents the policy it quotes.**

This gap comes from the controls themselves. We are not asserting it.

PART IV

What AI changes, and what it does not

It does not introduce the failure. Part I establishes that.

WHAT CHANGES	WHY IT MATTERS
Rate	A medical director composing a rationale under volume pressure errs occasionally. A model composes every rationale.
It repeats — the one that matters	A person's misreading of a provision is a one-off, confined to one letter. A model makes the **same** misreading: the same provision, misread the same way, across every letter that touches it. Scattered individual errors are individual appeals; the same error repeated is a pattern.
You can no longer review it all	The volume that makes automation attractive makes 100% human review uneconomic. Sampling was adequate when errors were random; it is weak against an error that recurs by construction.
Individualisation defeats the existing check	For years the criticism was that letters were generic. LLM-written rationales are individualised and case-specific — they pass every check built to detect boilerplate, while the accuracy question goes unasked.

And the pressure is now two-sided. CMS-0057-F requires a specific reason for each denial. [6] Simultaneously, providers have AI tools that generate appeals at near-zero marginal cost, targeting the clinical rationale directly. [12] Payers are being asked to write that paragraph more precisely at exactly the moment it is being attacked more efficiently.

PART V

Why existing detection tools miss it

Physician sign-off is accountability, not a fidelity check. A reviewer reading a fluent, individualised, clinically plausible letter has no signal that a *should* became a *must*. Nothing reads wrong.

Template and reason-code governance constrains the skeleton. The error is in the free text, which is precisely what templates do not constrain.

General-purpose hallucination detection looks for unsupported factual claims. In the inversion cases there is no unsupported claim — the criterion is quoted or closely paraphrased and the facts are accurate. Tools tuned for factual grounding cannot see this kind of error: a correct fact used with the wrong authority, or turned around to mean the opposite.

The gap is not negligence. **No existing control takes the source provision as its referent.**

PART VI

The missing control

What actually gets checked

Not the sentence. Always the pair:

> **(output sentence, governing provision)**

Authorization fidelity cannot be assessed on a sentence in isolation, because the property is a *relation*. A substantial part of the engineering problem is determining which provision governs a given passage before any judgement is meaningful.

Certification: turning policy into evidence

Before anything can be measured against a policy, the policy is reduced to **certified provisions**: each provision's exact text plus **how strongly it is worded** — whether it requires something, forbids it, permits it, recommends it, or simply describes it. This is a human step, performed once per policy document.

It is a familiar control. This is **validation**, not verification — *did we build the right thing*, not *did we build the thing right* — and the discipline is established: independent review, separation of author from verifier, documented inter-rater agreement.

What we check for

ERROR	WHAT IT CATCHES
Modality escalation	force upgraded — *"should be considered"* → *"is required"*
Scope drift	a rule scoped to one population restated as universal
Conditionality loss	a triggering condition dropped; the rule asserted unconditionally
Unsupported detail invention	a threshold, document or deadline the policy never contains
Source-claim overreach	*"the policy explicitly requires…"* where the source is silent or permissive
Decision-stage leakage	an outcome announced from a stage not authorised to decide it
Criterion inversion	a criterion cited accurately, used with its logical role reversed

Each is defined as a specification rather than as code, so the catalog extends as new kinds of error turn up in real letters. Four of the seven correspond directly to defects in Part I.

What this control is not

It does not assess clinical appropriateness, second-guess the medical director's judgement, or determine whether a denial was correct.

It determines whether the letter claims only what the policy authorizes — a narrower question than "is this decision right," and far more tractable, which is precisely why it can be measured.

PART VII

What an engagement looks like

This paper describes the first of three engagements: an **AI Readiness & Risk Scan**. It is an assessment, not a platform adoption — no integration, no runtime dependency, no change to the generation path.

The scan establishes what your current correspondence actually does against your own policies. What follows from it — a **Runtime Enforcement Pilot** that moves the check into the generation path, and **Continuous Drift Monitoring** as policies version and models change — is only worth discussing once the scan has shown whether there is anything to enforce.

Stage 1 · Policy certification. Candidate provisions are extracted from the governing policy documents and routed through blind independent review. **Your medical directors are the right adjudicators** — they own the policy. Deliverable: a certified provision set with documented reviewer agreement, versioned.

Stage 2 · Retrospective audit. Run against **test or pre-production correspondence first** — no PHI, no BAA, no security review, and findings arrive before member-facing deployment. Production audit follows once value is demonstrated.

Stage 3 · The report. Findings grouped by type of error, by provision, and by workflow stage. Every finding traces to a specific provision and a specific span of text, answerable to *"which passage did you compare this to, and who says that passage governs here?"*

Note that the scan does not require the payer to have deployed AI. The evidence in Part I is drawn entirely from human-authored letters. A plan whose rationales are still written by physicians has the same exposure, and the same absence of any control that checks it. The scan is diagnostic either way; what differs is what it recommends next.

On the business case

We will not hand you a return-on-investment model, because we would have to invent the inputs. There is no reliable public figure for what an appeal costs a health plan to work. What is public is the scale: roughly $35 billion a year in US prior-authorization administration, and a 95% overturn rate on appealed Medicare Advantage skilled-nursing denials. [1]

The numbers that matter are yours — your appeal volume, your overturn rate, what an appeal costs you to work, and how often the rationale is the thing being challenged. You already have all four.

The scan supplies the missing piece: how often your letters misstate your own policies, and in what way. Set that against your own cost per appeal and the business case is arithmetic rather than assertion. We do that calculation with you, using your figures, as part of the report.

One trend to price in either way. Providers now generate appeals with AI at close to zero marginal cost, and they aim at the clinical rationale. [12] Every appeal costs you real money and costs them almost nothing. That gap widens on its own, and the paragraph they aim at is the one paragraph nothing in your control stack checks.

Appendix · What we do not claim

Included deliberately. Where the product is measurement discipline, the boundary of what has been measured is part of the deliverable.

- **We publish no detector precision or recall figures.** Our reference corpus is not independently adjudicated, and a performance number scored against an unadjudicated corpus measures agreement with ourselves. Engagement findings are adjudicated with the client.
- **We do not claim automated policy interpretation.** Certification is a human judgement. Machine-suggested labels exist to draw the reviewer's eye, are withheld from the blind review pass, and are never presented as certified.
- **We do not claim complete recall.** We have not catalogued every kind of error. The catalog is explicitly extensible.
- **We do not assess clinical correctness.** Authorization fidelity is a language property, not a medical one.
- **We make no claim about the internal mechanics of language models.** What is documented is *that* these errors occur, in human and machine-written text alike.
- **Where two qualified reviewers genuinely disagree** on a provision's force, the honest artifact is a flagged provision, not a coin-flip label presented as certified.

References

[1] HHS Office of Inspector General, *Medicare Advantage Organizations Overturned Nearly All Appealed Prior Authorization Denials for Skilled Nursing Facility Admission*, OEI-09-24-00331, 8 June 2026. oig.hhs.gov

[2] HHS Office of Inspector General, *The Three Largest Medicare Advantage Organizations Denied Requests for Long-Term Acute Care and Inpatient Rehabilitation at Some of the Highest Rates*, 2026. oig.hhs.gov

[3] HHS Office of Inspector General, *Some Medicare Advantage Organization Denials of Prior Authorization Requests Raise Concerns About Beneficiary Access to Medically Necessary Care*, April 2022. oig.hhs.gov

[4] KFF, *Claims Denials and Appeals in ACA Marketplace Plans in 2024*. kff.org

[5] 29 C.F.R. § 2560.503-1(g)(1)(i)–(ii), requiring an adverse benefit determination notice to set out *"[t]he specific reason or reasons for the adverse determination"* and *"[r]eference to the specific plan provisions on which the determination is based."* Heightened obligations at (g)(1)(vii) and (j)(6) are directed at disability benefit determinations. Text verified at govinfo.gov

[6] CMS-0057-F, in force 1 January 2026. Impacted payers must give a specific reason for each denied prior-authorization decision, which *"could include explaining how the submitted documentation failed to support the request, or citing the specific plan provisions on which the denial is based."* Note the disjunction: citing the provisions is one of two permitted ways to satisfy the requirement, not a mandate to cite criteria.

[7] Centers for Medicare & Medicaid Services, managed-care notices and forms guidance: *"it is not permissible to omit any standardized language, nor alter the template, including font size, without CMS approval."* cms.gov

[8] New York State Department of Health, final adverse determination notice guidance — rationale must be *"sufficiently specific to enable the enrollee to determine the basis for appeal."* health.ny.gov

[9] *T. E. v. Anthem Blue Cross and Blue Shield*, No. 25-5407 (6th Cir. 22 Jan. 2026) (appeal from W.D. Ky. No. 3:22-cv-00202); recommended for publication. Governing policy: MCG Guideline for Residential Behavioral Health Level of Care, Child or Adolescent. govinfo.gov

[10] *Robert D. and K.D. v. Blue Cross of California d/b/a Anthem Blue Cross*, No. 2:20-cv-138-HCN-DAO (D. Utah 30 Jan. 2024) (Nielson, J.), reviewed de novo. Governing policy: Anthem Clinical UM Guideline, *Psychiatric Disorder Treatment — Residential Treatment Center*. govinfo.gov

[11] *Mark M. and Nina M. v. United Behavioral Health*, No. 2:18-cv-00018-BSJ (D. Utah 3 Sept. 2020) (Jenkins, J.). Governing policies: UBH Level of Care Guidelines (Residential, Mental Health Conditions); AACAP CASII v1.5; AACP LOCUS Adult Version. Note: the court **upheld** the denial under arbitrary-and-capricious review notwithstanding the conceded mismatch — a documented criteria/letter mismatch is not automatically dispositive. govinfo.gov

[12] On the provider-side appeal-automation ecosystem: HFMA, *Predict, Prevent, Perform: The AI Evolution of Denials Management*. hfma.org

[13] Independence Blue Cross, *Provider Manual — Clinical Services / Utilization Management*, April 2026. provcomm.ibx.com

[14] Health Net California, *Provider Library — Non-Delegated Medical Management*, describing referral of cases not meeting criteria to the medical director. providerlibrary.healthnetcalifornia.com ---

About this paper

Suggested citation. LangInMotion.ai, Inc., *Denial Letters: The Unchecked Paragraph*, Version 1.0, 21 August 2026 (LIM-WP-2026-01).

Revisions. Version 1.0 is the first public release. Material changes to the argument or to the sources will increment the version and be noted here.

Sources. All quoted plan criteria and denial-letter text in this paper are reproduced from published opinions of United States federal courts, which are public records. We quote the courts' reproductions, not the underlying licensed criteria products.

Trademarks. LangInMotion® is a registered trademark of LangInMotion.ai, Inc. Reg. No. 8,348,708.

InterQual is a trademark of Optum, Inc. MCG is a trademark of MCG Health, LLC. Payer and plan names appear only in case citations. Those marks are used for identification only; no affiliation, sponsorship or endorsement is implied.

Rights. © 2026 LangInMotion.ai, Inc. May be shared and quoted with attribution.

Contact. langinmotion.ai

Find out what your letters actually say

An AI Readiness & Risk Scan runs against your test correspondence — no PHI, no integration, no change to how you write today. You get findings traced to your own policy provisions, adjudicated with your medical directors.

Book a call

Or reply to the address this reached you from.