Refusing Care, Detecting Fraud: The Same Calculation

From the Contestable Decision to the Default Trajectory: Directional Neutrality and Terminal Legitimation of Algorithmic Care

What Shifted in May, and What "Abandonment" Does Not Mean

Three facts converged in May 2026.

1. First: in *Estate of Lokken v. UnitedHealth Group*, a Minnesota court ordered in March the production of documents, granting the plaintiffs' request in substance (Becker's Payer); the litigation concerns the nH Predict model, to which is attributed an error rate of approximately 90% in post-acute care denials (CBS).
2. Second: the health press reported on 19 May that human review of clinical context has been almost entirely delegated to automated systems (WUSF).
3. Third: on 21 May, the HHS announced an intensification of its use of AI to track fraud in federal health expenditure (Washington Post).

Let us say it immediately, to close the easy reading: the rationing of care did not wait for artificial intelligence. Prior authorization, reimbursement triage, and prioritization of scarce resources predate it. The thesis is therefore not that AI denies care. That formulation is convenient, false, and immediately filed away as yet another insurance scandal. The real thesis is colder: the calculation does not create the economic perimeter of denial; it changes its causal structure, scale, speed, opacity, and diffusion of causality, to the point of rendering care statistically improbable without any denial ever having been issued.

A word on the term. In French law, *déshérence* evokes the heirless succession, institutional abandonment, a passive absence. That is not what is at stake here, and this is not a metaphor. What is described is an active deprivation, produced by calculation, and distributed across an architecture. That distinction is the entire object of this text.

What the Calculation Changes, and the Perimeter It Optimizes

A coverage management model does not predict the patient's condition. It predicts a cost trajectory and calibrates care against that trajectory. The difference is categorical, and it has a name: the optimization perimeter.

Three perimeters compete over the same decision. The clinical perimeter maximizes patient benefit and minimizes loss of chance. The budgetary perimeter minimizes avoidable expenditure. The capacity perimeter absorbs flow and prevents saturation. None of these is inherently illegitimate. The wrong arises from one thing only: that a system claims to operate within the clinical perimeter while in fact optimizing a budgetary or capacity perimeter. This is not a performance failure; it is a categorical error concerning the objective, a false ontology of purpose. The system is not mistaken within its perimeter; it operates within a different perimeter than the one attributed to it.

Hence the first distinction that cuts, and that survives everything that follows: *human-in-the-loop* versus *human-as-alibi*. UnitedHealth's defense, to the effect that coverage decisions are made by medical directors and not by the AI, is accurate in its letter and misleading in its scope. The relevant question is not who signs, but who composes the perimeter on which the signature bears. If the medical reviewer validates in bulk a recommendation produced on the cost perimeter, the signature does not reintroduce the clinical perimeter: it authenticates its absence. The human is then in the loop without power over the loop. Present, but without perimeter.

Three Generations: G2 Optimizes a Decision, G3 Optimizes a Decision Space

The phenomenon can be ordered across three generations, and the decisive distinction is not the one expected.

1. The first generation is the explicit human refusal: an agent decides, and the decision is contestable because it is dated, signed, situated.
2. The second is the algorithmically recommended decision validated by a human: the model proposes, the human disposes, in principle. It retains the object of the first: there is still a decision, on a case file, that can be isolated and challenged.
3. The third generation, the one this text describes, changes the object. It no longer optimizes a decision; it optimizes a decision space.

G2 optimizes a decision. G3 optimizes a decision space.

A generation 3 system appears when a system no longer acts primarily through explicit decisions on individual case files, but through the distributed modification of conditions of access, priority, friction, or capacity, upstream of the local clinical decision. It does not reject the case file: it deforms the terrain on which the case file will be processed. When a denial occurs, it is no longer a cause, but the side effect of an already inclined trajectory.

This third generation must be held with care. Claiming that human validation converges mechanically with model output would be a strong thesis, and I have no quantitative proof: no public override rate establishes it. The correct formulation is not "humans validate mechanically." It is: the question is not whether the human is formally present, but how frequently the human actually overrides the model, in which classes of cases, with what traceability, and under what responsibility. Generation 3 is a structural hypothesis; its falsifier is known and measurable: the human override rate, by class of case. This data is publicly available nowhere; that absence of publication is itself a signal of poor governability.

This displacement has a consequence that formulations centered on "automated decisions" miss. Triage tools, scoring systems, and capacity instruments do not always decide on a patient; they transform the context in which that patient's deprivation becomes probable. The wrong does not reside in an act, but in a default trajectory.

The entire progression can be summarized in a matrix that forms its backbone.

Object Optimized	**Visible Effect**	**Mode of Contestation**
Individual decision	Explicit denial	Classical recourse
Individual decision	Explicit denial	Classical recourse
Population trajectory	Probabilistic attrition	Near-invisible

The vertical reading is the entire argument: as the optimized object rises from decision to trajectory, the effect dilutes and the mode of contestation collapses. We know how to challenge a denial. We do not know how to challenge a probability.

Directional Neutrality

The convergence of May's facts takes on its full meaning here. The same type of tool, the predictive model applied to a healthcare expenditure, serves, at the insurer, to deny care, and at the public regulator, to detect fraud. In France, the Assurance Maladie detected and halted 723 million euros of fraud in 2025, up 15%, through datamining deployed over a decade (ameli.fr); the CPAM de Paris has, since August 2025, assigned an alert level to each case file to prioritize financial stakes (Acuité). The dominant debate reads these uses along a moral axis: one would be abusive, the other virtuous. That reading misses the essential point.

The calculation is neutral as to the direction of the wrong, and that is the doctrinal axis of this text: it does not morally distinguish between undue and legitimate expenditure; it distinguishes only costly trajectories from acceptable ones within the perimeter assigned to it. The model does not know whether the expenditure it screens out is fraud, necessary care, a coding error, or a politically inconvenient cost. It optimizes a perimeter. The same calculatory indifference produces, according to the perimeter entrusted to it, scandal or sound management.

The term directional neutrality then demands a strict definition, lest it invite precisely the opposite reading from the one it carries. It does not designate an axiological neutrality of the model, the false idea that AI is "without values." It designates an architectural property: the same artifact can be reoriented toward opposite ends without changing its technical grammar. It could be named more dryly as *instrumental invariance* or *teleological indifference of the optimization engine*. I retain "directional neutrality" on one condition of reading: neutral as to the direction of the wrong, never as to the priority encoded in the perimeter. This is why the directional neutrality of the calculation does not mean the political neutrality of the system: the architecture is commutable; the purpose is not. The danger is not that AI is neutral, but that it is reusable, and that this reusability goes unnoticed behind the apparent morality of the use case.

One observation, to be stated coldly. The same instrument serves, in a government agency, to protect public expenditure against fraud, and at an insurer, if the Lokken allegations prevail, to compress expenditure at the patient's expense. The calculation does not know which of the two ends it serves; it distinguishes trajectories, not intentions. The morality of the use case resides in the institution that assigns the perimeter, never in the model. A governance device that monitors the model without monitoring the assignment of the perimeter is watching the wrong variable.

Europe: Modulating Rather Than Refusing

A French executive might, at this point, close the file: American matter, private insurers, nothing to do with a solidarity-based system. That would be a framing error, and it must be closed immediately.

Europe does not need to import the American prior authorization model to encounter the problem. It need only modulate access, delay, intensity, priority, and the administrative burden of the care pathway. That is the second distinction that cuts: denial is merely one modality of deprivation; there is also delay, and there is friction.

Modern deprivation rarely passes through an explicit no. It passes through delay, document requests, redirection, low prioritization, enhanced scrutiny, temporary suspension, pathways rendered more costly in cognitive terms. The modern denial does

not say no: it slows, complicates, de-prioritizes. It is a deprivation without event: no dated, signed, actionable decision, only an access probability that declines.

This friction is nothing abstract about "AI and care": it is written in identifiable architectural layers. A queue prioritization. A business process management orchestration that orders the steps of a case file. A routing rule that directs a request toward a fast or slow circuit. A policy engine that encodes, in the form of thresholds, a risk policy. An admission control that regulates entry into a queue. A dynamic documentary control threshold deciding which case files will require additional documentation. A fraud score injected into the claims processing workflow. An SLA weighting that hierarchizes guaranteed timelines. A case file ranking. A hospital scheduling optimization. Each of these is a point where a condition of access is calibrated, without any clinical decision being formally made, and therefore without any formally being contestable. The modulation is in the information system, not in the courtroom.

The French terrain is already there, without anything spectacular. Fraud scoring sorts case files by alert level; capacity management optimizes flows: the Calyps algorithm has predicted activity at the Valenciennes hospital center since 2021, with a reliability of approximately 95% at 48 hours (esanum). None of these tools "refuses" a patient; each displaces an initial condition: who will be seen, when, in what order, after what scrutiny. France is not (yet) industrializing explicit algorithmic refusal; it is already industrializing algorithmic modulation of the care pathway.

This is where the central dissociation of this text appears, and it must be stated frontally: a system can remain locally compliant, with each routing rule, each threshold, each SLA meeting its specification, while globally producing an attrition of access to care that no decision has ordered. Compliance is verified at the component level; attrition occurs at the system level. Neither contradicts the other, and it is precisely this that makes them formidable in combination.

Other mechanisms complete this picture, such as automated reimbursement sorting, the medico-economic optimization of supplementary insurers, or population-level regulation, but their precise documentation remains to be established [SOURCE TO DOCUMENT: dated deployments of algorithmic pathway modulation in France beyond fraud detection and capacity management], and I prefer to signal that gap than to fill it with invention.

Why Human Review Does Not Suffice

Here comes the serious objection, from advocates of quality-based regulation: audit the model, impose effective human review, and the device becomes governable again. It contains a share of truth: if review genuinely reassigned the decision to its clinical

perimeter, my thesis would fall. But it presupposes that convergence between human and model would be a failure, a laziness, an alibi that it suffices to correct through discipline. That convergence is an expected property of the work system, not a fault of the reviewers.

The human converges with the model by design: limited time per case file, informational asymmetry facing a system the reviewer did not build, productivity pressure, reluctance to deviate from a traced recommendation, absence of access to the model's causal reasoning, responsibility diffused across the chain, and, most determinatively, an override cost that exceeds the validation cost. Overriding requires time, a justification, personal exposure; validating requires none of these. A system that makes overriding more costly than validating mechanically produces validation. The human signature then ceases to be a control and becomes a terminal legitimation, not through vice, but through structure. That is why "more review" does not suffice: it adds signature to a system that already produces it.

What is missing is not human presence; it is the measurement of its effectivity and the reattribution of the perimeter. Five requirements would make the device actionable, and they are not doctrine, they are field conditions. Declare the perimeter that is actually optimized: clinical, economic, capacity-based, or fraud-detection. Measure the actual human override rate, by class of case. Trace the causal transitions, from score to alert, from alert to review, from review to decision, from decision to pathway effect. Identify nominative responsibility for each critical transition. Publish or audit the classes of false positives and false negatives with an effect on access to care.

The regulatory framework confirms the stakes through its own prudence. Article 14 of the AI Act imposes human oversight for high-risk systems, but its effectivity remains to be operationalized. And a political compromise between the Council and the European Parliament, announced on 7 May 2026 within the Digital Omnibus, is set to defer, pending formal adoption, the application of high-risk obligations to 2 December 2027 for autonomous systems and to 2 August 2028 for systems embedded in products (EU Council). This interregnum, where the requirement takes shape before its effectivity becomes fully actionable, is precisely the space where facade review flourishes, and where dissociation becomes a viable strategy.

The modern system does not deny the right to care. It degrades the effective probability of accessing it without producing any legally spectacular event. Nothing to challenge, because nothing has, in the legal sense, occurred.

Governing Initial Conditions, Not Final Decisions

There remains the operational requirement, keeping a distance from legal opinion, which falls to specialized counsel that this text does not claim to formulate. The requirement is

not "audit your models." It is: declare the perimeter on which the care trajectory is calculated, and measure how frequently a human departs from it. This is a perimeter requirement before it is a performance requirement.

A contrast illustrates this. A predictive medicine device oriented toward patient decompensation bears on a physiological trajectory: the anticipation window, the alert threshold, the moment of intervention. PREDICARE, within the territorial predictive medicine program, is built within this perimeter. Its governance difficulty is intact, but it is the right perimeter: the error one fears there is an error concerning the patient's condition, therefore an error the clinician can contest on their own terrain. A coverage or regulatory system, with comparable predictive structure, bears on the cost or capacity trajectory. Same form, inverse perimeter. The prediction inherits from its perimeter, and governing a predictive system begins with governing what it predicts the trajectory of.

At the close of this inquiry, the device can be stated in one sentence: care no longer needs to be explicitly refused; it can be rendered statistically improbable by a decision environment optimized for a perimeter other than clinical benefit. The consequence for decision-makers is not heightened vigilance over final decisions: there are, precisely, none left to monitor. When the wrong has dissolved into a trajectory, the last point of leverage is neither the decision, which did not occur, nor the signature, which was only a legitimation: it is the initial condition. The assigned perimeter, the calibrated threshold, the routing rule, the encoded priority, everything that inclines the terrain before the first case file enters it.

Governing algorithmic care is no longer a matter of arbitrating denials. It is a matter of reclaiming the initial conditions, before they become a trajectory that no one will know how to contest: not the patient, who has nothing to challenge; not the clinician, who signed nothing; not the institution, which only assigned a perimeter.