Refuser le soin, détecter la fraude : le même calcul

De la décision contestable à la trajectoire par défaut : neutralité directionnelle et légitimation terminale du soin algorithmique

Ce qui a basculé en mai, et ce que « déshérence » ne veut pas dire

Trois faits se sont rapprochés au mois de mai 2026.

1. Dans l'affaire *Estate of Lokken v. UnitedHealth Group*, une juridiction du Minnesota a ordonné en mars la production de documents, accédant pour l'essentiel à la demande des plaignants (Becker's Payer) ; le litige porte sur le modèle nH Predict, auquel est imputé un taux d'erreur de l'ordre de 90 % dans le refus de soins post-aigus (CBS).
2. La presse santé décrit, le 19 mai, l'examen humain du contexte clinique quasi entièrement délégué à des systèmes automatisés (WUSF).
3. Et le 21 mai, le HHS annonce intensifier son recours à l'IA pour traquer la fraude dans la dépense fédérale de santé (Washington Post).

Disons-le tout de suite, pour fermer la lecture facile : le rationnement du soin n'a pas attendu l'intelligence artificielle. L'autorisation préalable, le tri des remboursements, la priorisation des ressources rares lui préexistent. La thèse n'est donc pas que l'IA refuse des soins. C'est une formulation commode, fausse, immédiatement classée comme énième scandale assurantiel. Elle est plus froide : le calcul ne crée pas le périmètre économique du refus ; il en change la structure causale, échelle, vitesse, opacité, diffusion de la causalité, jusqu'à rendre le soin statistiquement improbable sans qu'aucun refus n'ait jamais été prononcé.

Un mot enfin sur le terme. En droit français, la déshérence évoque la succession sans héritier, l'abandon institutionnel, une absence passive. Ce n'est pas ce dont il s'agit ici, et ce n'est pas une métaphore. Je décris une privation active, produite par calcul, et distribuée dans une architecture. La distinction est tout l'objet du texte.

Ce que le calcul change, et le périmètre qu'il optimise

Un modèle de gestion de couverture ne prédit pas l'état du patient. Il prédit une trajectoire de coût et calibre une prise en charge sur cette trajectoire. La différence est de nature, et elle a un nom : le périmètre d'optimisation.

Trois périmètres se disputent la même décision. Le périmètre *clinique* maximise le bénéfice du patient et minimise la perte de chance. Le périmètre *budgétaire* minimise la dépense évitable. Le périmètre *capacitaire* absorbe le flux et évite la saturation. Aucun n'est illégitime en soi. Le tort naît d'une seule chose : qu'un système prétende opérer dans le périmètre clinique alors qu'il optimise en réalité un périmètre budgétaire ou capacitaire. Ce n'est pas un défaut de performance ; c'est une erreur de catégorie sur la finalité, une fausse ontologie de l'objectif. Le système ne se trompe pas dans son périmètre ; il opère dans un périmètre différent de celui qu'on lui prête.

D'où la première distinction qui tranche, et qui survivra à tout l'article : *human-in-the-loop* contre *human-as-alibi*. La défense de UnitedHealth, selon laquelle la couverture est tranchée par des directeurs médicaux et non par l'IA, est exacte dans sa lettre et trompeuse dans sa portée. La question pertinente n'est pas qui signe, mais qui compose le périmètre sur lequel la signature porte. Si le médecin-conseil valide en bloc une recommandation produite sur le périmètre du coût, sa signature ne réintroduit pas le périmètre clinique : elle l'authentifie absent. L'humain est alors dans la boucle sans pouvoir sur la boucle. Présent, mais sans périmètre.

Trois générations : G2 optimise une décision, G3 optimise un espace de décision

On peut ordonner le phénomène en trois générations, et la distinction décisive n'est pas celle qu'on attend.

1. La première génération est le refus humain explicite : un agent décide, et sa décision est contestable parce qu'elle est datée, signée, située.
2. La deuxième est la recommandation algorithmique validée par un humain : le modèle propose, l'humain dispose. En principe. Elle conserve l'objet de la première : il y a toujours une décision, sur un dossier, qu'on peut isoler et attaquer.
3. La troisième, celle que décrit ce texte, change d'objet. Elle n'optimise plus une décision : elle optimise un espace de décision.

G2 optimise une décision. G3 optimise un espace de décision.

Une génération 3 apparaît lorsqu'un système n'agit plus principalement par décision explicite sur dossier individuel, mais par modification distribuée des conditions d'accès, de priorité, de friction ou de capacité, en amont de la décision clinique locale. Il ne refuse

pas le dossier : il déforme le terrain sur lequel le dossier sera traité. Le refus, quand il survient, n'est plus une cause, mais l'effet de bord d'une trajectoire déjà inclinée.

Il faut tenir cette troisième génération avec prudence. Affirmer que la validation humaine converge mécaniquement avec la sortie du modèle serait une thèse forte, et je n'en ai pas la preuve chiffrée : aucun taux d'infirmation public ne l'établit. La bonne formulation n'est donc pas « l'humain valide machinalement ». Elle est : la question n'est pas de savoir si l'humain est formellement présent, mais à quelle fréquence il infirme réellement le modèle, dans quelles classes de cas, avec quelle traçabilité, et sous quelle responsabilité. La génération 3 est une hypothèse structurelle ; son falsifieur est connu et mesurable : le taux d'infirmation humaine, par classe de cas. Cette donnée n'est, à ce jour, publique nulle part ; cette absence de publication est déjà un signal de faible gouvernabilité.

Ce déplacement a une conséquence que les formulations centrées sur la « décision automatisée » manquent. Les outils de tri, de scoring, de capacité ne décident pas toujours d'un patient ; ils transforment le contexte dans lequel sa privation devient probable. Le tort ne se loge pas dans un acte, mais dans une trajectoire par défaut.

On peut résumer la progression entière dans une matrice qui en donne la colonne vertébrale.

Objet optimisé	**Effet visible**	**Mode de contestation**
Décision individuelle	Refus explicite	Recours classique
Décision individuelle	Refus explicite	Recours classique
Trajectoire populationnelle	Attrition probabiliste	Quasi invisible

La lecture verticale est tout l'argument : à mesure que l'objet optimisé s'élève, de la décision à la trajectoire, l'effet se dilue et le mode de contestation s'effondre. On sait attaquer un refus. On ne sait pas attaquer une probabilité.

La neutralité directionnelle

Le rapprochement des faits de mai prend ici tout son sens. Le même type d'outil, le modèle prédictif appliqué à une dépense de santé, sert, chez l'assureur, à refuser le soin, et chez le régulateur public, à traquer la fraude. En France, l'Assurance Maladie a détecté et stoppé 723 millions d'euros de fraude en 2025, en hausse de 15 %, au moyen d'un datamining mobilisé depuis une décennie (ameli.fr) ; la CPAM de Paris attribue depuis août 2025 un niveau d'alerte à chaque dossier pour prioriser les enjeux financiers (Acuité). Le débat dominant lit ces usages sur un axe moral : l'un serait abusif, l'autre vertueux. Cette lecture manque l'essentiel.

Le calcul est neutre quant à la direction du tort, et c'est l'axe doctrinal de ce texte : il ne distingue pas moralement la dépense indue de la dépense légitime ; il distingue

seulement les trajectoires coûteuses des trajectoires acceptables dans le périmètre qui lui a été assigné. Le modèle ne sait pas si la dépense qu'il écarte est une fraude, un soin nécessaire, une erreur de codage ou un coût politiquement indésirable. Il optimise un périmètre. La même indifférence calculatoire produit, selon le périmètre qu'on lui confie, le scandale ou la bonne gestion.

Le terme de *neutralité directionnelle* demande alors une définition stricte, faute de quoi il invite la lecture exactement inverse de celle qu'il porte. Il ne désigne pas une neutralité axiologique du modèle, l'idée, fausse, que l'IA serait « sans valeurs ». Il désigne une propriété d'architecture : le même artefact se laisse réorienter vers des finalités opposées sans changer de grammaire technique. On pourrait le nommer plus sèchement *invariance instrumentale* ou *indifférence téléologique du moteur d'optimisation*. Je retiens « neutralité directionnelle » à une condition de lecture : neutre quant à la direction du tort, jamais quant à la priorité encodée par le périmètre. C'est pourquoi la neutralité directionnelle du calcul ne signifie pas la neutralité politique du système : l'architecture est commutable, la finalité ne l'est pas. Le danger n'est pas que l'IA soit neutre, c'est qu'elle soit réutilisable, et que cette réutilisabilité passe inaperçue derrière la moralité apparente du cas d'usage.

Une observation, à énoncer froidement. Le même instrument sert, dans une administration, à protéger la dépense publique contre la fraude, et chez un assureur, si les allégations de Lokken prospèrent, à comprimer la dépense au détriment du patient. Le calcul ne sait pas laquelle des deux finalités il sert ; il distingue des trajectoires, pas des intentions. La morale du cas d'usage réside dans l'institution qui assigne le périmètre, jamais dans le modèle. Un dispositif de gouvernance qui surveille le modèle sans surveiller l'assignation du périmètre surveille la mauvaise variable.

L'Europe : moduler plutôt que refuser

Un dirigeant français pourrait, arrivé ici, classer le dossier : affaire américaine, assureurs privés, rien à voir avec un système solidaire. Ce serait une erreur de cadrage, qu'il faut fermer dès maintenant.

L'Europe n'a pas besoin d'importer la *prior authorization* américaine pour rencontrer le problème. Il lui suffit de moduler l'accès, le délai, l'intensité, la priorité et la charge administrative du parcours. C'est la deuxième distinction qui tranche : le refus n'est qu'une modalité de la privation ; il y a aussi le retard, et il y a la friction.

La privation moderne passe rarement par un non explicite. Elle passe par le délai, la demande de pièces, la réorientation, la priorisation basse, le contrôle renforcé, la suspension temporaire, le parcours rendu plus coûteux cognitivement. Le refus moderne ne dit pas non : il ralentit, complique, dé-priorise. C'est une privation sans événement :

aucune décision datée, signée, opposable, seulement une probabilité d'accès qui décroît.

Cette friction n'a rien d'une abstraction sur « l'IA et le soin » : elle s'écrit dans des couches d'architecture identifiables. Une priorisation de file d'attente. Une orchestration de processus (BPM) qui ordonne les étapes d'un dossier. Une règle de routage qui dirige une demande vers un circuit lent ou rapide. Un moteur de règles (*policy engine*) qui code, sous forme de seuils, une politique de risque. Un *admission control* qui régule l'entrée dans une file. Un seuil dynamique de contrôle documentaire qui décide quels dossiers exigeront des pièces supplémentaires. Un score antifraude injecté dans le workflow de traitement des demandes (*claims*). Une pondération de SLA qui hiérarchise les délais garantis. Un *ranking* de dossiers. Une optimisation du *scheduling* hospitalier. Autant de points où une condition d'accès se calibre, sans qu'aucune décision clinique ne soit formellement prise, donc sans qu'aucune ne soit formellement contestable. La modulation est dans le système d'information, pas dans le prétoire.

Le terrain français est déjà là, sans rien de spectaculaire. Le scoring antifraude trie les dossiers par niveau d'alerte ; le pilotage capacitaire optimise les flux : l'algorithme Calyps prédit l'activité au centre hospitalier de Valenciennes depuis 2021, avec une fiabilité de l'ordre de 95 % à 48 heures (esanum). Aucun de ces outils ne « refuse » un patient ; chacun déplace une condition initiale : qui sera vu, quand, dans quel ordre, après quel contrôle. La France n'industrialise pas (encore) le refus algorithmique explicite ; elle industrialise déjà la modulation algorithmique du parcours.

C'est ici qu'apparaît la dissociation centrale de ce texte, et il faut l'énoncer frontalement : un système peut rester localement conforme, chaque règle de routage, chaque seuil, chaque SLA respectant sa spécification, tout en produisant globalement une attrition d'accès au soin qu'aucune décision n'a ordonnée. La conformité se vérifie au niveau du composant ; l'attrition se produit au niveau du système. Aucune ne contredit l'autre, et c'est précisément ce qui les rend redoutables ensemble.

D'autres mécanismes complètent ce tableau, comme le tri automatisé des remboursements, l'optimisation médico-économique des complémentaires ou la régulation populationnelle, mais leur documentation précise reste à établir [SOURCE À DOCUMENTER : déploiements datés de modulation algorithmique du parcours en France au-delà de l'antifraude et du capacitaire], et je préfère le signaler que de l'inventer.

Pourquoi la revue humaine ne suffit pas

Vient l'objection sérieuse, celle des partisans d'une régulation par la qualité : auditez le modèle, imposez une revue humaine effective, et le dispositif redevient gouvernable. Elle contient une part de vrai : si la revue réattribuait la décision à son périmètre clinique, ma thèse tomberait. Mais elle présuppose que la convergence entre l'humain et le modèle

serait une défaillance, une paresse, un alibi qu'il suffirait de corriger par la discipline. Or cette convergence est une propriété attendue du système de travail, pas une faute des *reviewers*.

L'humain converge avec le modèle par design : temps limité par dossier, asymétrie informationnelle face à un système qu'il n'a pas construit, pression de productivité, peur de déroger à une recommandation tracée, absence d'accès au raisonnement causal du modèle, responsabilité diluée sur la chaîne, et, le plus déterminant, un coût d'infirmation supérieur au coût de validation. Infirmer demande du temps, une justification, une exposition personnelle ; valider n'en demande aucun. Un système qui rend l'infirmation plus coûteuse que la validation produit mécaniquement de la validation. La signature humaine cesse alors d'être un contrôle pour devenir une *légitimation terminale*, non par vice, par structure. C'est pourquoi « plus de revue » ne suffit pas : on ajoute de la signature à un système qui en produit déjà.

Ce qui manque n'est pas la présence humaine, c'est la mesure de son effectivité et la réattribution du périmètre. Cinq exigences rendraient le dispositif opposable, et elles n'ont rien d'une doctrine, ce sont des conditions terrain. Déclarer le périmètre réellement optimisé : clinique, économique, capacitaire ou antifraude. Mesurer le taux d'infirmation humaine réel, par classe de cas. Tracer les transitions causales, du score à l'alerte, de l'alerte à la revue, de la revue à la décision, de la décision à l'effet sur le parcours. Identifier une responsabilité nominative sur chaque transition critique. Publier ou auditer les classes de faux positifs et de faux négatifs ayant un effet sur l'accès au soin.

Le cadre réglementaire confirme l'enjeu par sa propre prudence. L'article 14 de l'AI Act impose une supervision humaine pour les systèmes à haut risque, mais son effectivité reste à instrumenter. Et un compromis politique entre le Conseil et le Parlement européen, annoncé le 7 mai 2026 dans le cadre du Digital Omnibus, doit reporter, sous réserve d'adoption formelle, l'application des obligations *high-risk* au 2 décembre 2027 pour les systèmes autonomes et au 2 août 2028 pour les systèmes embarqués dans des produits (Conseil de l'UE). Cet interrègne, où l'exigence se dessine avant que son effectivité ne devienne pleinement opposable, est exactement l'espace où la revue de façade prospère, et où la dissociation devient une stratégie viable.

Le système moderne ne nie pas le droit au soin. Il dégrade la probabilité effective d'y accéder sans produire d'événement juridiquement spectaculaire. Rien à attaquer, parce que rien n'a, au sens juridique, eu lieu.

Gouverner les conditions initiales, pas les décisions finales

Reste l'exigence opératoire, en se tenant à distance de l'avis juridique, qui relève d'un conseil spécialisé que ce texte ne prétend pas formuler. L'exigence n'est pas « auditez vos modèles ». Elle est : déclarez le périmètre sur lequel la trajectoire de soin est calculée, et mesurez à quelle fréquence un humain s'en écarte. C'est une exigence de périmètre avant d'être une exigence de performance.

Un contraste le montre. Un dispositif de médecine prédictive orienté vers la décompensation du patient porte sur une trajectoire physiologique : la fenêtre d'anticipation, le seuil d'alerte, le moment d'intervention. PREDICARE, dans le cadre du programme territorial de médecine prédictive, est construit dans ce périmètre. Sa difficulté de gouvernance est entière, mais elle est de bon périmètre : l'erreur qu'on y craint est une erreur sur l'état du patient, donc une erreur que le clinicien peut contester sur son propre terrain. Un système de couverture ou de régulation, à structure prédictive comparable, porte sur la trajectoire de coût ou de capacité. Même forme, périmètre inverse. La prédiction hérite de son périmètre, et gouverner un système prédictif commence par gouverner ce dont il prédit la trajectoire.

Au terme du parcours, le dispositif se laisse énoncer en une phrase : le soin n'a plus besoin d'être explicitement refusé ; il peut être rendu statistiquement improbable par un environnement de décision optimisé pour un autre périmètre que le bénéfice clinique. La conséquence pour le décideur n'est pas un surcroît de vigilance sur les décisions finales : il n'y en a, justement, plus à surveiller. Quand le tort s'est dissous dans une trajectoire, le dernier point de prise n'est ni la décision, qui n'a pas eu lieu, ni la signature, qui n'était qu'une légitimation : c'est la condition initiale. Le périmètre assigné, le seuil calibré, la règle de routage, la priorité encodée, tout ce qui incline le terrain avant que le premier dossier n'y entre.

Gouverner le soin algorithmique, ce n'est plus arbitrer des refus. C'est reprendre la main sur les conditions initiales, avant qu'elles ne deviennent une trajectoire que plus personne ne saura contester : ni le patient, qui n'a rien à attaquer, ni le clinicien, qui n'a rien signé, ni l'institution, qui n'a fait qu'assigner un périmètre.