LE TERRAIN TECHNIQUE

Produire ne coûte plus rien, vérifier coûte toujours autant

Tyler Kohler . 18 août 2026 . note 03

Une page de texte plausible se produit aujourd'hui pour un coût proche de zéro. La vérifier coûte ce qu'elle a toujours coûté : le temps d'une personne compétente. Ces deux coûts ont longtemps varié ensemble ; ils ne varient plus.

Le prix d'une phrase plausible

Produire un texte crédible supposait quelqu'un qui écrit et un support qui publie. Chacun avait un prix, qui limitait le volume. Ce plancher a cédé avec la génération automatique de texte. Un modèle statistique entraîné sur de grands corpus restitue à la demande des paragraphes corrects et adaptés au sujet. La quantité ne dépend plus du temps de quelqu'un. Le coût unitaire ne tombe pas à zéro : il tombe au prix du calcul.

La vérification n'a pas suivi la même pente. Contrôler une affirmation suppose d'identifier ce qu'elle avance, de remonter à une source primaire, de la lire, parfois d'appeler quelqu'un. Ces opérations restent humaines et se comptent en dizaines de minutes. Le coût de production dépend désormais d'une machine, celui de la vérification toujours d'une personne.

La réfutation ne connaît pas d'économies d'échelle

Une formule circule depuis 2013 pour décrire ce déséquilibre : réfuter demande beaucoup plus d'énergie que produire. Elle oppose deux régimes de coût.

La production bénéficie d'économies d'échelle : le millième texte ne coûte presque rien. La vérification n'en bénéficie pas. Chaque affirmation a ses propres faits et son propre interlocuteur, et la millième coûte ce qu'a coûté la première.

Devant un volume, on ne réfute donc pas tout : on choisit. Le reste demeure, non parce qu'il est vrai, mais parce que personne n'a eu le temps.

Ce que le volume fait aux signaux de classement

Les systèmes qui trient l'information publique se sont éprouvés sur des volumes humains. Leurs signaux gardaient leur valeur tant que produire coûtait quelque chose.

La reprise en est l'exemple le plus net. Une affirmation relayée par des sources séparées valait indice, parce que chaque relais supposait une décision et un coût. L'exception était connue : une dépêche d'agence produisait des dizaines de reprises sans autant de vérifications. Elle est devenue la règle. Le signal perd sa valeur informative sans perdre son poids dans le calcul. Pour qui lit, la reprise a cessé d'être un indice et doit être traitée comme telle.

La modération obéit à une autre logique. Les capacités d'examen sont finies, le traitement se fait par priorité. Pour qui signale un contenu, la conséquence est concrète : des délais, et une absence de réponse qui ne vaut pas décision. Un contenu resté en ligne n'a pas été jugé conforme : il n'a pas été examiné.

Faux n’est pas creux

Le débat public se concentre sur le faux contenu : l’affirmation inexacte, la citation inventée. Ce contenu est nuisible, mais il est attaquable : une affirmation fausse se compare à un fait, et cette comparaison ouvre le démenti comme la voie judiciaire.

Le contenu creux n’a pas cette propriété. Il n’avance rien de vérifiable. Il évoque un contexte sans conclure et rapproche des éléments sans les relier. Il ne ment pas, faute d’affirmer.

Sa fonction n’est pas nulle pour autant. L’analyste qui l’observe constate un vocabulaire qui s’installe autour d’un nom, et des associations que les moteurs mesurent. Pour la personne visée, l’impasse tient en une phrase : sans affirmation à confronter, ni le démenti ni la voie judiciaire n’ont de prise.

Le coût de ce contenu a toujours été bas. Il est devenu négligeable comme celui de tout le reste, et ce qui restait marginal devient une masse.

Pourquoi la détection automatique est fragile

L’idée d’un détecteur de texte généré se heurte à trois obstacles, indépendants de la qualité des outils.

Le premier tient à ce qui est mesuré. Un détecteur n’observe pas l’origine d’un texte, il observe des régularités statistiques. Ces régularités caractérisent aussi une prose administrative, ou celle d’une personne qui n’écrit pas dans sa langue maternelle. Une étude publiée en 2023 a soumis sept détecteurs à deux corpus : des rédactions de candidats à un test d’anglais, des rédactions d’élèves américains. Plus de la moitié des premières ont été classées comme générées, contre moins d’une sur dix des secondes.

Le deuxième tient aux taux de base. Sur un grand nombre de textes majoritairement humains, un taux d’erreur même faible produit beaucoup de mises en cause injustifiées. Et l’erreur n’est pas symétrique : classer à tort un texte humain accuse quelqu’un.

Le troisième tient au mouvement. Un détecteur est entraîné sur une distribution datée, alors que les modèles évoluent plus vite que les corpus de référence.

Un score de détection ne dit rien sur l’origine d’un texte. Il ne peut donc pas fonder à lui seul une mise en cause.

Ce que l’on regarde à la place

La question de la fabrication ne se laisse pas trancher. Celle du soutien factuel, oui. Trois familles d’indices y suffisent, indépendantes de la façon dont le texte a été écrit.

La cohérence des faits vérifiables, d’abord. Un texte généré est plausible dans sa forme et faible dans son détail. Les dates, les références de décisions et les citations attribuées sont des points de contact avec le monde. Une référence qui n’existe pas règle la fiabilité sans rien dire de l’origine.

La traçabilité des sources, ensuite. On remonte les reprises jusqu’à leur premier état. Soit la chaîne aboutit à un document antérieur et identifiable, soit elle tourne en rond entre des pages qui se citent l’une l’autre. Le second cas se lit sans compétence technique.

Le comportement de diffusion, enfin. On en tire des dates de publication, des écarts entre ces dates, et des recoupements entre des comptes présentés comme indépendants.

Aucun de ces axes ne dit si une machine a écrit le texte. C’est ce qui les rend opposables : ils

portent sur des éléments qu'un tiers contrôle lui-même.

Le temps de vérification est rationné

La vérification coûte du temps humain, donc elle est rationnée. Celui qui s'interroge, journaliste ou juriste, dispose d'un budget qu'il n'a pas choisi. Il s'arrête quand ce budget est épuisé, et retient ce qui y était contrôlable. La question qui décide n'est donc pas ce qui est vrai, mais ce qui se vérifie dans le temps disponible.

Le texte est devenu abondant, donc bon marché, donc faible comme preuve. Ce qui devient rare, et cher, c'est la source vérifiable. Elle est datée et déposée chez un tiers qui en répond. Elle est signée par quelqu'un que l'on peut joindre. Aucune de ces propriétés ne s'obtient après coup.

En Suisse, ce socle existe : inscription au registre du commerce, publications à la Feuille officielle suisse du commerce, attestations de fonction par des tiers, prises de parole datées sous son nom. Les comptes n'en font pas partie : l'article 958e du code des obligations réserve leur publication aux sociétés cotées ou débitrices d'un emprunt par obligations, et n'accorde ailleurs qu'un droit de consultation au créancier justifiant d'un intérêt digne de protection. En France, le dépôt au greffe est la règle, avec option de confidentialité pour les plus petites sociétés.

La préparation s'en trouve inversée. Elle portait sur la visibilité, avantage annulé dès lors que publier ne coûte plus rien à personne. Elle porte désormais sur la vérifiabilité : disposer, avant l'incident, d'un socle documentaire qu'un tiers contrôle dans le temps qu'il s'accorde.

Celui qui doit établir ses propres faits au moment où la question se pose consomme le budget de celui qui la pose. Il n'obtient pas un jugement défavorable, mais un dossier laissé ouvert, ce qui produit le même effet et ne se conteste nulle part.

La formule est attribuée à Alberto Brandolini, qui l'a énoncée publiquement en 2013. Aucun protocole expérimental ne la sous-tend : elle décrit un rapport d'effort sans le mesurer.

Weixin Liang, Mert Yuksekgonul, Yining Mao, Eric Wu et James Zou, dans la revue Patterns (juillet 2023), ont soumis sept détecteurs à 91 rédactions de candidats au TOEFL et à 88 rédactions d'élèves américains. Plus de la moitié des premières ont été classées comme générées, contre moins d'une sur dix des secondes. Les rédactions TOEFL provenaient d'un même forum en ligne, donc de locuteurs partageant une langue première : le résultat ne se généralise pas à l'ensemble des non-anglophones.

Kohler, T. (2026). Produire ne coûte plus rien, vérifier coûte toujours autant. Notes, tylerkohler.ch.
https://tylerkohler.ch/notes/produire-ne-coute-plus-rien/