Le Gradient de Contrôle : présence de l'IA, degré de reprise humaine et détectabilité de la production numérique

Claes G.[1], Vermeersch A.-L.[2], De Backer T.[3]

[1] *Universiteit Hasselt-Continentaal, Vakgroep digitale cultuur en receptie*
[2] *Universiteit Mechelen-Maritiem, Laboratorium voor authenticiteitsperceptie*
[3] *Universiteit Turnhout-Australis, Eenheid voor mensmachine-onderscheid*

DOI : 10.9999/icf.2026.gradient.de.controle · icf.news/article-gradient-de-controle

Résumé

Contexte. La production numérique assistée par intelligence artificielle se répand. On observe, dans son sillage, une homogénéisation des formes — mises en page, tournures, structures qui se ressemblent. L'étude ne mesure pas la qualité de ces productions, question de goût qui n'est pas de son ressort, mais leur *détectabilité* : la mesure dans laquelle un observateur repère qu'une IA est intervenue, et comment cette mesure varie selon le degré de reprise humaine.

Méthode. Six cent quatre-vingt-quatre participants ont évalué un fil de trente publications de type réseau social, mêlant trois niveaux de contrôle : production entièrement déléguée à l'IA sans reprise (contrôle nul), production IA reprise et corrigée par un humain (contrôle fort), et production humaine sans IA (sans IA). Trois mesures : détectabilité, saillance mnésique, et confiance déclarée dans le service annoncé. Une variable individuelle a été relevée : l'exposition de chaque participant aux outils d'IA.

Résultats. La détectabilité décroît fortement avec le contrôle humain : 84 % pour la production non reprise, 41 % pour la production reprise, 33 % pour le fait-main. La différence entre production reprise et fait-main n'atteint pas le seuil de signification statistique. La production non reprise est aussi la moins mémorisée (22 % de rappel spontané contre 58 % pour le fait-main) et la moins créditée de confiance (4,1 contre 7,0 sur 10). Surtout, la capacité individuelle à détecter l'IA décroît avec l'exposition du participant à l'IA (r = −0,41) : les usagers les plus intensifs sont les moins aptes à la repérer.

Conclusion. L'intelligence artificielle ne produit pas l'uniformité ; elle produit une norme, et c'est l'écart à cette norme — la reprise humaine — qui devient saillant. Quand le fond devient synthétique, la main qui corrige ressort. L'Institut relève que cette conclusion vaut aussi pour les instruments qui l'ont produite, et renvoie à la déclaration correspondante.

Mots-clés : intelligence artificielle — détectabilité — gradient de contrôle — reprise humaine — saillance — perception d'authenticité

1. Un incident préalable

L'étude est née d'une anecdote qu'un des auteurs tient d'un enseignant de néerlandais. La classe était composée d'élèves francophones en immersion, à qui une dissertation avait été confiée à faire à domicile. À la remise des copies, l'enseignant a félicité le groupe, puis s'est interrompu, le paquet de feuilles à la main :

Goed werk, kinderen. Alleen — zo goed spreek ik zelf geen Nederlands. « *Beau travail. Seulement — moi-même, je ne parle pas si bien le néerlandais.* »

Le détail décisif n'est pas que les copies étaient excellentes. C'est qu'elles l'étaient d'une manière qui ne pouvait pas être celle de ces élèves — ni, l'enseignant le reconnaissait, tout à fait la sienne. Une maîtrise sans les hésitations, les calques du français, les tournures apprises qui trahissent d'ordinaire un apprenant. Une perfection sans origine.

L'enseignant n'a pas eu besoin de comparer les copies à un corpus, ni de mesurer quoi que ce soit. Il a détecté un écart entre le niveau produit et le niveau attendu de cette source précise. C'est exactement l'opération que la présente étude cherche à mettre en nombres : non pas « est-ce bon ? », mais « cela vient-il d'où cela prétend venir ? »

2. Problème et cadrage

L'usage d'outils génératifs pour produire des contenus numériques — visuels promotionnels, publications de réseaux sociaux, textes courts — s'est généralisé. Une observation informelle accompagne cette diffusion : ces productions tendent à se ressembler. Mêmes structures de vidéos, mêmes mises en page de visuels, mêmes tournures d'accroche.

L'Institut a choisi de ne pas mesurer ce qui relèverait d'un jugement de valeur. La question « ces productions sont-elles moins bonnes ? » appelle un critère esthétique que l'Institut n'a pas à fixer. Une question voisine est en revanche mesurable : ces productions sont-elles *reconnaissables* comme issues d'une IA, et cette reconnaissance dépend-elle de la manière dont l'IA a été employée ?

© 2025 Institut du Constat Fondamental — icf.news · ISSN 2025-0001 (fictif) · Peer-reviewed par nos soins

Le déplacement est important, car il transforme un débat d'opinion en observation. On ne demande pas aux participants s'ils aiment. On leur demande de dire ce qu'ils repèrent, ce qu'ils retiennent, et ce qu'ils croient.

2.1 Le gradient de contrôle

L'hypothèse centrale porte sur une variable rarement isolée : le degré de reprise humaine. Entre « l'humain a tout fait » et « l'IA a tout fait » s'étend un continuum que l'étude segmente en trois niveaux.

Niveau	Procédure de production	Ce que le niveau isole
Contrôle nul	Requête générique adressée à un outil génératif, sortie publiée sans retouche	L'IA livrée à elle-même
Contrôle fort	Sortie IA reprise par un humain : correction, personnalisation, choix conservés ou écartés	L'IA comme instrument tenu en main
Sans IA	Production faite main, sans outil génératif (mise en page manuelle, rédaction directe)	Le référent humain

Tableau 1. Les trois niveaux du gradient de contrôle. Les contenus portaient tous sur des services ordinaires — dépannage, cours particuliers, activités locales — pour neutraliser l'effet du sujet.

2.2 Les trois mesures

Chaque participant a fait défiler un fil de trente publications, dix par niveau, dans un ordre aléatoire, sur une interface reproduisant celle d'un réseau social. Trois mesures ont été recueillies :

La détectabilité. Après chaque publication, le participant indiquait s'il pensait qu'une IA était intervenue dans sa production. On en tire le pourcentage d'attribution à l'IA par niveau.

La saillance. En fin de parcours, sans consigne préalable de mémorisation, le participant citait les publications dont il se souvenait. On en tire un taux de rappel spontané.

La confiance. Pour chaque service annoncé, le participant évaluait sur dix la probabilité qu'il ferait appel à ce service. Cette mesure ne dit rien de la qualité réelle du service — inconnue et hors sujet — mais de l'inférence que le participant en tire à partir de l'annonce.

3. Résultats

3.1 La détectabilité décroît avec le contrôle

Niveau	Attribué à l'IA	Rappel spontané	Confiance (/10)
Contrôle nul	84 %	22 %	4,1
Contrôle fort	41 %	51 %	6,8
Sans IA	33 %	58 %	7,0

Tableau 2. Les trois mesures par niveau de contrôle. La production reprise (contrôle fort) se rapproche du fait-main sur les trois dimensions et s'éloigne de la production non reprise.

La production non reprise est identifiée comme telle par 84 % des participants. La production reprise ne l'est que par 41 %, et le fait-main par 33 %. L'écart entre contrôle nul et contrôle fort atteint quarante-trois points ; l'écart entre contrôle fort et fait-main n'est pas significatif ($\chi^2 = 2{,}9$; $p = 0{,}09$). Autrement dit : **dès qu'un humain reprend la sortie, elle se rapproche du fait-main jusqu'à ne plus pouvoir en être distinguée avec assurance dans le cadre de ce protocole.**

Ce n'est pas la présence de l'IA que l'œil détecte. C'est son abandon.

3.2 Ce qu'on oublie et ce à quoi on se fie

Les deux autres colonnes du Tableau 2 suivent la même pente. La production non reprise est la moins mémorisée : 22 % de rappel spontané, contre 58 % pour le fait-main. Elle est aussi la moins créditée de confiance : 4,1 sur 10, contre 7,0.

Sur les seules publications de niveau contrôle nul, la détectabilité et la confiance sont corrélées négativement ($r = -0{,}52$; $p < 0{,}001$) : plus une annonce est perçue comme non reprise, moins le participant se dit prêt à recourir au service.

L'Institut insiste sur ce que cette corrélation n'établit pas. Elle n'établit pas que les services promus sans reprise sont de moindre qualité — le protocole ne mesure aucun service. Elle établit que les participants *infèrent* une moindre qualité à partir d'une moindre reprise. L'absence de soin visible dans l'annonce est traitée comme un indice sur le soin réel. Cette inférence peut être juste ou fausse ; l'étude constate seulement qu'elle est faite.

L'annonce sans main visible n'est pas jugée laide. Elle est jugée suspecte. Étude n°48 — Institut du Constat Fondamental

3.3 La variable qui retourne l'étude

Avant le début du protocole, chaque participant avait renseigné son exposition hebdomadaire aux outils d'IA générative. Cette variable, prévue comme simple covariable de contrôle, a produit le résultat le plus net de l'étude.

La capacité individuelle à détecter correctement l'IA — mesurée par un indice de discrimination entre les niveaux (d′) — décroît à mesure que l'exposition augmente ($r = -0{,}41$; $p < 0{,}001$). L'indice d′ distingue la capacité réelle à différencier les conditions de la simple tendance individuelle à attribuer plus ou moins volontiers les contenus à une IA : une personne qui crierait « IA » à tout propos obtiendrait un fort taux de détection sans discrimination réelle, et le d′ ne la crédite pas de cette suspicion indistincte.

Exposition à l'IA (quartile)	Indice de discrimination (d′)
Q1 — faible exposition	3,09
Q2	2,83
Q3	2,70
Q4 — forte exposition	2,25

Tableau 3. Capacité de discrimination selon l'exposition du participant aux outils d'IA. Plus l'exposition est forte, moins la production IA est détectée.

Le sens de l'effet mérite d'être posé clairement, car il est le cœur de l'article. Ce ne sont pas les personnes les plus étrangères à l'IA qui la détectent le moins. C'est l'inverse. Les usagers les plus intensifs sont les moins aptes à la repérer.

L'interprétation que les auteurs avancent — sans pouvoir l'établir par ce seul protocole — est une hypothèse d'accoutumance perceptive. À force de côtoyer une forme, on cesse de la percevoir comme une forme. Elle devient le fond sur lequel on lit tout le reste. Ce qui détonne pour un œil neuf passe pour normal à un œil saturé.

Encart — le dessin d'arbre

À l'école maternelle, les enfants dessinent les arbres à peu près de la même façon : un trait vertical brun, une masse verte au sommet. Ce schéma n'est pas ce qu'ils voient par la fenêtre — les arbres réels n'ont ni ce tronc ni cette boule. C'est un pattern acquis, transmis, reproduit sans qu'aucun enfant ne le perçoive comme un pattern. Pour lui, c'est simplement « comme on dessine un arbre ».

On ne voit le schéma qu'après l'avoir désappris — en réapprenant à regarder l'arbre plutôt qu'à réciter sa forme convenue. La perception du cliché suppose une sortie du cliché.

Le Tableau 3 décrit la même opération, à front renversé. L'usager intensif d'IA a intériorisé la forme générative au point qu'elle ne lui apparaît plus comme une forme. Il ne la détecte pas, non par défaut d'attention, mais parce qu'elle est devenue sa manière par défaut de voir un contenu. Il dessine tous ses arbres pareils et ne sait pas qu'il le fait.

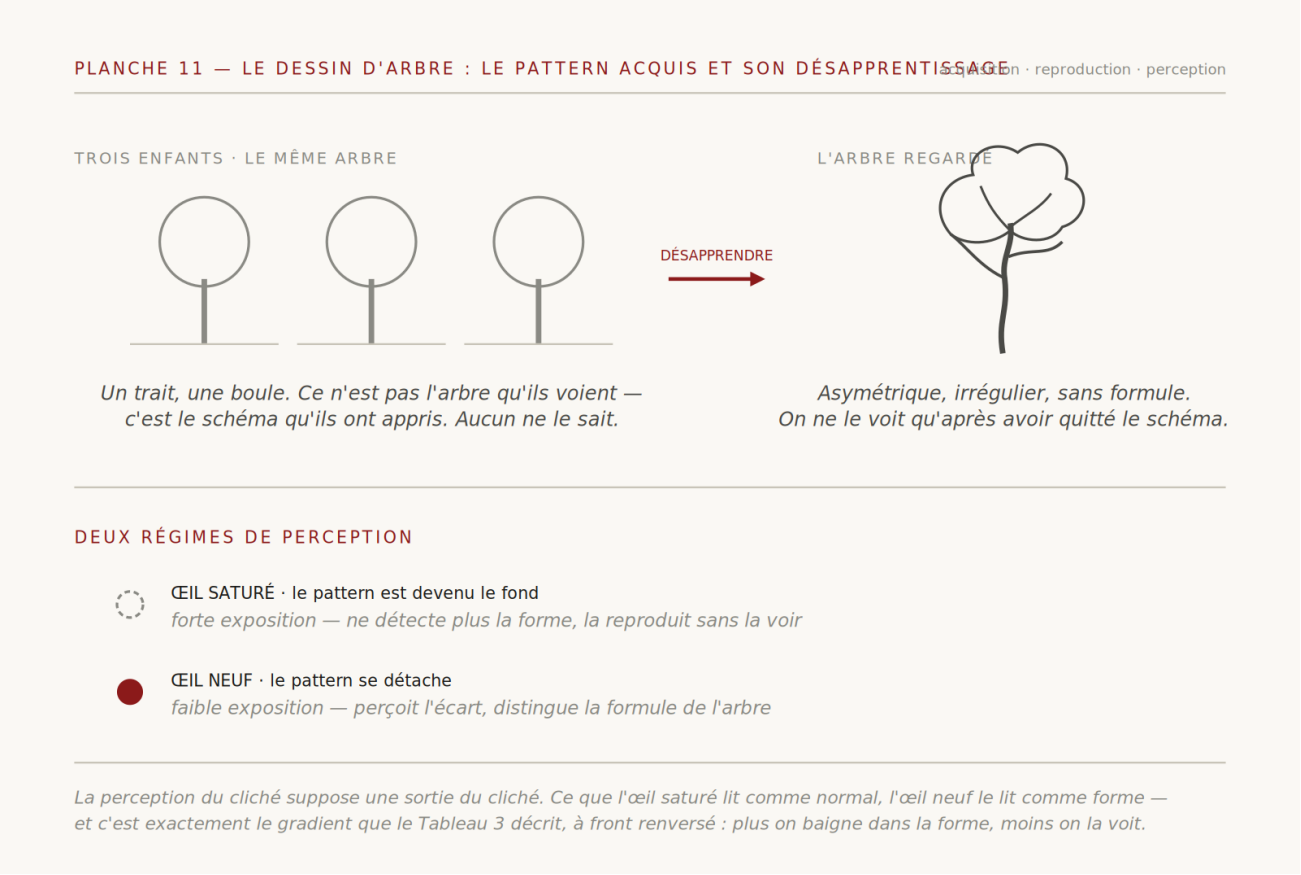

Planche 11. Le dessin d'arbre et son désapprentissage. À gauche, le schéma convenu — un trait, une boule — reproduit à l'identique sans qu'aucun enfant ne le perçoive comme un schéma. À droite, l'arbre effectivement regardé, irrégulier et sans formule, que l'on ne voit qu'après avoir quitté le cliché. En bas, les deux régimes de perception que le Tableau 3 met en nombres : l'œil saturé, pour qui la forme est devenue le fond, et l'œil neuf, qui en détecte l'écart.

Encart méthodologique — Dr. K. Osei-Mensah, Responsable Statistiques & Analyses

Une corrélation de –0,41 entre exposition et détection est modérée. Je tiens à ce qu'elle ne soit pas lue comme davantage.

Elle ne dit pas que l'usage de l'IA rend aveugle. Elle dit qu'à travers cet échantillon, les deux varient ensemble, et en sens opposé. Trois lectures restent ouvertes et le protocole ne les départage pas : l'exposition émousse la détection ; ou les personnes qui détectent mal l'IA sont celles qui l'adoptent le plus volontiers ; ou une troisième variable — l'âge, le métier, le temps passé en ligne — porte les deux.

Je signale néanmoins que le gradient par quartile est monotone. Chaque palier d'exposition détecte moins bien que le précédent, sans exception. Cette régularité ne prouve pas la cause, mais elle est trop propre pour être écartée d'un revers de main.

4. Discussion

4.1 L'uniformité n'est pas là où on l'attend

Le sens commun voudrait que l'IA uniformise et que l'humain diversifie. Les résultats invitent à une formulation plus exacte. L'IA ne produit pas l'uniformité en elle-même : elle produit une *norme*, statistiquement centrale, dépourvue d'écart. C'est cette absence d'écart qui la rend détectable, et c'est le retour de l'écart — la reprise humaine — qui la rend indétectable à nouveau.

La production humaine ne ressort donc pas parce qu'elle serait meilleure. Elle ressort parce qu'elle dévie. L'erreur conservée, l'asymétrie, le choix un peu étrange qu'aucune requête générique ne produirait : voilà les marqueurs que l'œil enregistre. Dans un environnement où la norme devient synthétique, la déviation devient un signal, et le signal se lit comme une présence.

4.2 Le biais de sur-attribution

Un résultat secondaire éclaire l'envers du phénomène. Trente-trois pour cent des productions faites main, sans aucune IA, ont été attribuées à une IA. Le fait-main est donc soupçonné à tort dans un cas sur trois.

Cet effet a une conséquence que les auteurs jugent importante. Quand la forme générative devient la norme perceptive, elle ne jette pas seulement le soupçon sur les productions IA : elle le jette sur toute production régulière, soignée, sans aspérité — y compris humaine. La bouillie, pour reprendre un terme qui a circulé pendant la conception de l'étude, ne noie pas que l'IA. Elle installe un doute général, dont l'humain soigné fait aussi les frais.

4.3 La bouillie de la bouillie

Les auteurs relèvent un dernier étage, qu'ils n'ont pas mesuré mais que la logique de l'étude appelle. Un marché de formations à l'usage de l'IA s'est développé, dont une part promet d'apprendre à « sortir du lot ». Ces formations diffusent des méthodes, et notamment des requêtes types, à un grand nombre de personnes simultanément.

Si la thèse de l'article est exacte, l'effet est prévisible : une requête enseignée à tous produit une norme au second degré. Ce qui devait distinguer uniformise, un cran plus haut. Le pattern se reproduit à l'étage censé y échapper. Les auteurs proposent d'appeler ce phénomène, faute de terme consacré, la reproduction du pattern par sa propre remédiation.

5. Limites

La première limite est constitutive, et les auteurs la formulent sans détour. Les contenus de niveau « sans IA » ont été réalisés manuellement et vérifiés comme tels : la variable indépendante est garantie sur ce point. En revanche, l'assemblage du fil, la mise en forme de l'interface et la préparation du matériel périphérique ont eux-mêmes mobilisé des outils génératifs. L'instrument de mesure partage donc, pour son enveloppe sinon pour ses stimuli, la propriété qu'il mesure — et les auteurs ne peuvent situer avec certitude leur propre production quelque part sur le gradient qu'ils décrivent. La classification expérimentale n'en est pas affectée ; la posture des expérimentateurs, si.

La deuxième tient à la détectabilité déclarée. Les participants disaient s'ils *pensaient* qu'une IA était intervenue ; on ne dispose d'aucun étalon objectif de ce qu'ils auraient dû répondre pour les stimuli ambigus. La mesure est une perception, non une vérité de terrain.

La troisième porte sur la confiance. Elle a été recueillie comme intention déclarée, dans un contexte sans enjeu réel. Rien n'assure qu'elle prédit un comportement effectif de recours à un service.

La quatrième concerne l'exposition à l'IA, mesurée par auto-déclaration hebdomadaire, instrument grossier et sujet aux biais de mémoire et de désirabilité.

6. Conclusion

L'intelligence artificielle est souvent présentée comme une menace d'uniformisation, qui effacerait les singularités sous une production standardisée. Les résultats de cette étude suggèrent une lecture inverse, et plus favorable qu'il n'y paraît.

En imposant une norme, l'IA rend visible ce qui s'en écarte. Elle ne dissout pas la main humaine : elle la met en relief. Un contenu repris, corrigé, tenu en main par quelqu'un redevient indétectable comme IA, non parce qu'il dissimule l'outil, mais parce que la reprise réintroduit exactement ce qui manquait — la déviation, le choix, l'origine.

L'outil n'est donc pas l'ennemi de la singularité. Il en est, pour qui le tient, le révélateur. Un dictionnaire ne remplace pas celui qui écrit ; il lui permet de se relire. L'IA employée avec un fort gradient de contrôle occupe la même place : un instrument qui, loin d'effacer la personne, lui permet de mieux paraître.

Reste la position, moins confortable, de ceux qui lâchent la main. Leur production est détectée, oubliée, et créditée d'une moindre confiance — pour un service qui, peut-être, ne la méritait pas. L'Institut ne formule pas de recommandation. Il observe que, dans ses données, la différence entre paraître authentique et paraître synthétique ne tient pas à l'usage de l'outil, mais au fait de le tenir.

La situation est stable. Elle l'est un peu moins pour ceux qui ont cessé de regarder.

Déclarations

Financement. Institut du Constat Fondamental, sur fonds propres.

Conflits d'intérêts. Les auteurs déclarent avoir recouru à des outils d'intelligence artificielle pour la préparation du matériel expérimental et pour d'autres travaux de l'Institut. L'ampleur exacte de ce recours fait l'objet de la note ci-dessous.

Déclaration d'auto-implication de la rédaction. Le Prof. Parmentier, informé des conclusions de la présente étude par un membre des services de documentation, a demandé qu'il soit consigné une question qu'il juge « non rhétorique ». Il souhaite savoir dans quelle mesure les instruments d'IA ont contribué à la production des études de l'Institut, au-delà de ce qui figure dans le rapport d'opportunité interne consacré à ces outils — lequel, rédigé par une doctorante de l'Institut avec sa propre contribution qualifiée d'involontaire, actait le recours à NotebookLM (Google DeepMind) pour produire les synthèses audio des études à partir de leur texte. Il observe que si le gradient de contrôle décrit dans cet article est exact, l'Institut ne peut pas préjuger de sa propre position sur ce gradient sans un examen qu'il n'a pas conduit. Le Comité de lecture a jugé la question fondée et a renoncé à y répondre, au motif qu'il n'était pas certain de pouvoir le faire sans recourir à l'un des outils concernés.

Approbation éthique. Le protocole a été soumis au Comité d'Éthique de l'Institut. Le Comité n'a pas répondu, ce qui valait approbation selon les statuts. Les auteurs relèvent que cette non-réponse, cette fois, ne peut être attribuée à une automatisation, le Comité n'ayant jamais démenti être composé d'êtres humains.

Remerciements. À l'enseignant de néerlandais dont la remarque a ouvert cette recherche, et qui a demandé à ne pas être nommé — précisant qu'il n'était pas certain, à la réflexion, d'avoir vraiment prononcé sa phrase aussi bien.

Références

– Claes, G. et De Backer, T. (2025). Vormherkenning en machinale productie : een verkennend kader. *Tijdschrift voor Digitale Cultuur en Authenticiteitsperceptie*, 3(2), 88–109.

– Institut du Constat Fondamental (2025). Indifférenciation acoustique et plasticité perceptive intergénérationnelle. *Journal of Intergenerational Acoustic Perception*, Étude n°21.

– Institut du Constat Fondamental (2026). L'Anamnèse Sans Destinataire : entretien clinique automatisé et ressenti post-passation. *Nordic Journal of Clinical Automation and Assessment*, Étude n°47.

– Vermeersch, A.-L. (2024). Blootstelling en perceptuele gewenning : een overzicht. *Vlaams Tijdschrift voor Cognitieve Psychologie*, 19(1), 22–41.