Le *Generative Engine Optimization* : nouvelle arme de manipulation informationnelle

Cercle Pegase

Octobre 2026

Une analyse de Mathilde Georges, avec le soutien de Thomas Delorme

Synthèse

À l'instar du SEO pour les moteurs de recherche, le développement des intelligences artificielles génératives a fait émerger de nouvelles stratégies de visibilité, regroupées sous le terme de *Generative Engine Optimization* (GEO). Si ces techniques permettent aux entreprises d'améliorer la présence de leurs contenus dans les réponses des moteurs génératifs, elles peuvent également être détournées pour diffuser des informations trompeuses voire fausses via réponses générées.

À partir de cinq cas documentés, ce rapport examine la frontière entre optimisation légitime et manipulation de l'information à travers trois critères : le contenu diffusé, l'identité de son émetteur et les modalités de sa production. Il analyse les mécanismes techniques susceptibles de favoriser ces manipulations, leurs usages commerciaux, informationnels et criminels, ainsi que les enjeux de détection et de régulation.

Les cas étudiés mettent en évidence des vulnérabilités possibles, mais ne permettent pas de mesurer leur présence ni leurs effets réels sur les utilisateurs.

Sommaire

1. **Introduction. Du SEO au GEO : un changement dans la bataille pour la visibilité** 2
2. **Fonctionnement et vulnérabilité de la recherche de sources** 3
3. **L'émergence des techniques d'optimisation face à l'impératif de visibilité sur les LLM** 6
4. **Quand le GEO devient une arme d'influence** 7
5. **Risques systémiques et cadre réglementaire** 10
6. **Conclusion** 11

1. Introduction. Du SEO au GEO : un changement dans la bataille pour la visibilité

Dès les années 1990, l'accès à l'information dans l'espace numérique s'est structuré autour des moteurs de recherche (comme Google, Bing ou Yahoo), qui proposent une liste de sources classées selon leur pertinence par rapport à la requête formulée par l'utilisateur. Les utilisateurs peuvent ensuite consulter ces sources et se forger leur propre avis.

Apparaître en haut de la liste était ainsi devenu un enjeu central pour les acteurs économiques (entreprises, médias) cherchant à augmenter leur trafic ou simplement à être identifiés comme les meilleurs. Des techniques de référencement naturel (SEO) se sont développées pour améliorer la position au sein des pages de résultats. Leur efficacité dépend toutefois de critères de classement qui ne sont pas entièrement publics.

L'arrivée des modèles de langage (*Large Language Models* ou LLM) et des intelligences artificielles génératives (IAg) vient bouleverser ce modèle. Au lieu de proposer une sélection de liens à explorer, ces plateformes fournissent directement une réponse rédigée qui synthétise de nombreuses sources soit fournies par l'utilisateur soit sélectionnées et récupérées sur Internet. Dès lors, l'IA devient un tiers de confiance dans la sélection et la hiérarchisation des sources, dispensant bien souvent aux utilisateurs de cliquer sur les liens ou de simplement vérifier les sources. Un phénomène également identifié sur les moteurs de recherche en 2024[1].

Bien que la recherche via les moteurs traditionnels demeure le premier moyen d'accès à l'information, l'usage des IA conversationnelles progresse très rapidement et s'impose comme un concurrent direct des moteurs de recherche classiques. Certains moteurs de recherche tentent même de s'adapter, comme Google qui a lancé, en 2026 (en France), un affichage des résultats en résumé IA.

Dès lors, être cité par les IA génératives devient progressivement aussi crucial que le SEO l'était pour les moteurs classiques. C'est l'objectif d'une nouvelle stratégie de visibilité : le *Generative Engine Optimization* ou optimisation pour les moteurs génératifs (GEO). Il désigne l'ensemble des pratiques visant à augmenter la probabilité qu'un contenu soit repris, cité et référencé sur des moteurs de recherche basés sur l'intelligence artificielle, en influençant la sélection des sources par les IA.

Des acteurs légitimes (entreprises, médias) mais aussi malveillants (entreprises malfaisantes, réseaux criminels) peuvent mobiliser ces techniques à des fins qui leur sont propres. Cette diversité d'usages soulève une question : comment définir la frontière entre le marketing légitime et la manipulation de l'information ?

Pour y répondre, il est crucial d'identifier comment les IA cherchent l'information et où elles sont vulnérables. Nous développerons ensuite les techniques conventionnelles d'optimisation sur les IA génératives, avant de documenter les méthodes de manipulation de l'information par GEO, illustrées par des cas commerciaux, informationnels et criminels. Nous tenons également à examiner les risques systémiques, la détection et le cadre réglementaire autour de l'optimisation pour les moteurs génératifs.

[1] En 2024, un rapport de Rand Fishkin, spécialiste du référencement et fondateur de SparkToro, estimait que 59,7 % des recherches effectuées sur Google (en Europe) se terminaient sans aucun clic vers un site.

2. Fonctionnement et vulnérabilité de la recherche de sources

Pour comprendre les leviers de l'optimisation par GEO, il est indispensable d'analyser en détail comment et sous quelles contraintes les moteurs génératifs récupèrent des informations sur Internet.

a) La collecte : quelles données sont réellement ciblées ?

Lorsque la requête d'un utilisateur requiert des données actualisées ou absentes de la mémoire initiale du modèle, les moteurs génératifs s'efforcent d'offrir la meilleure réponse possible. Pour réduire les risques d'hallucination, ils se tournent alors vers les informations issues d'une base interne ou d'Internet, en activant des systèmes *Retrieval-Augmented Generation*, ou RAG, conçus pour récupérer des sources pertinentes sur le Web[2].

Pour identifier quelles sources choisir, ils s'appuient sur des calculs qui croisent une multitude de critères, comme les mots-clés, la popularité des sites ou encore la similarité des textes. Cette collecte se heurte toutefois à des obstacles techniques. De nombreux sites bloquent l'accès à leur contenu en imposant des barrières (captchas, authentifications strictes, etc.). Pour pallier ces contraintes, les grands acteurs de l'intelligence artificielle diversifient leurs sources en s'appuyant sur des flux RSS, des agences de presse, ou encore des API de moteurs de recherche tiers.

Certaines données sont toutefois soumises à des restrictions d'indexation, notamment lorsqu'il s'agit d'informations privées ou protégées par des licences strictes. Dans un domaine technique ou confidentiel disposant de peu de documentation ou encore une actualité dépourvue de couverture médiatique, un contenu artificiellement optimisé par le GEO ne rencontre alors presque aucun concurrent, ce qui peut lui conférer un quasi-monopole dans la sélection par les modèles génératifs. Une vulnérabilité d'autant plus préoccupante si elle est exploitée par des acteurs malveillants.

Face à ces limites, certains moteurs se tournent vers des contenus issus de plateformes ouvertes à faible barrière d'entrée (Reddit, X, blogs), faisant l'objet de partenariats d'indexation avec des fournisseurs, à l'instar d'OpenAI avec Reddit. Cette ouverture crée toutefois une surface d'exposition lorsque des acteurs malveillants cherchent à y diffuser des contenus trompeurs afin qu'ils soient captés par les systèmes de recherche.

Les documents ainsi récupérés alimentent uniquement le contexte de la conversation, c'est-à-dire qu'ils ne renforcent pas structurellement les connaissances du modèle. Néanmoins, les interactions peuvent indirectement servir à l'apprentissage des futurs modèles si elles sont collectées par la suite.

Le fonctionnement des systèmes RAG est en partie opaque et varie selon les fournisseurs. Pour comprendre comment des contenus externes peuvent influencer les réponses, nous nous appuyons ici sur le fonctionnement général de la génération augmentée par récupération, tel que documenté par la recherche académique.

b) De l'indexation à la génération : un fonctionnement vulnérable face aux manipulations de l'information

[2] Huang, Yizheng, et Jimmy Huang. « A Survey on Retrieval-Augmented Text Generation for Large Language Models ». ACM Computing Surveys, vol. 58, no 12, septembre 2026, p. 1-38.

La phase de récupération des sources comporte trois étapes, chacune étant vulnérable à des manipulations par des acteurs malveillants ayant pour but de référencer une information trompeuse ou fausse par les LLM.

- *Indexation des sources et bases vectorielles*

Mode de récupération. Les informations issues des documents web sont découpées selon leur idée directrice avant d'être transformées en codes mathématiques et rangées dans une base de données intelligente.

Sur un plan technique, les documents collectés sont découpés en fragments indépendants appelés *chunks*, chacun comportant une seule idée directrice afin de préserver la pertinence de l'information. En effet, un fragment trop long dilue l'information et alourdit le calcul tandis qu'un fragment trop court manque de contexte et peut devenir ambigu[3]. Un modèle *d'embedding* convertit chaque fragment en une suite de nombres appelée vecteur, qui capture son sens sémantique (signification, concept ou intention d'un texte). Ils sont ensuite classés et stockés dans une base de données vectorielle, permettant au système de sélectionner les sources pertinentes par simple similarité de sens avec la question de l'utilisateur.

Vulnérabilité aux manipulations. Si chaque système de recherche possède ses propres mécanismes de sélection, de contrôle et de présentation des sources, aucun ne dispose d'un mécanisme indépendant de vérification de l'information lors de la sélection des sources. Une étude publiée par des chercheurs en sécurité informatique[4] montre que les attaques par empoisonnement de requête (injection de faux contenus optimisés pour paraître sémantiquement pertinents) réussissent à tromper les RAG car leur système se fie à la proximité sémantique et non à la vérité.

- *Récupération et classement*

Afin de retrouver une information pertinente par rapport à la requête de l'utilisateur, le système combine une recherche par le sens (sémantique) et par les mots exacts (lexicale), avant de classer et de filtrer les meilleurs résultats.

Dans le détail, la requête de l'utilisateur est d'abord transformée en vecteur, ce qui permet d'établir un « score de proximité » avec les vecteurs textuels stockés en amont. En effet, le système évalue la correspondance sémantique (l'intention et les synonymes) et la correspondance lexicale (fondée sur la fréquence des termes) entre la requête et les informations récupérées. Les résultats de ces deux approches sont fusionnés, puis classés à l'aide d'un modèle de type *reranker*, afin de ne retenir que ceux qui dépassent un seuil de confiance défini, ce qui est censé garantir leur pertinence par rapport à la requête[5].

Vulnérabilité aux manipulations. Cette combinaison de méthodes de recherche expose toutefois le système à des risques de contournement, notamment par l'usage de techniques d'optimisation du texte (SEO/GEO). De plus, l'évaluation de signaux « de surface » (fluidité textuelle, etc.), sans vérification systématique de la véracité des informations, ouvre la porte à des manipulations de l'information. Un acteur malveillant peut ainsi calibrer un texte pour qu'il imite la structure syntaxique et lexicale des sources faisant autorité, afin d'augmenter les chances que son contenu franchisse le seuil de confiance du *reranker* et soit récupéré par l'intelligence artificielle.

- *La génération et l'attribution des sources*

[3]Wu, Shangyu, et al. « Retrieval-Augmented Generation for Natural Language Processing: A Survey ». arXiv:2407.13193, arXiv, 19 mai 2026.
[4] Chen, Tailun, et al. « MIRAGE: Misleading Retrieval-Augmented Generation via Black-box and Query-agnostic Poisoning Attacks ». arXiv:2512.08289, arXiv, 31 août 2026.
[5] Huang, Yizheng, et Jimmy Huang. « A Survey on Retrieval-Augmented Text Generation for Large Language Models ». ACM Computing Surveys, vol. 58, no 12, septembre 2026, p. 1-38. arXiv.org

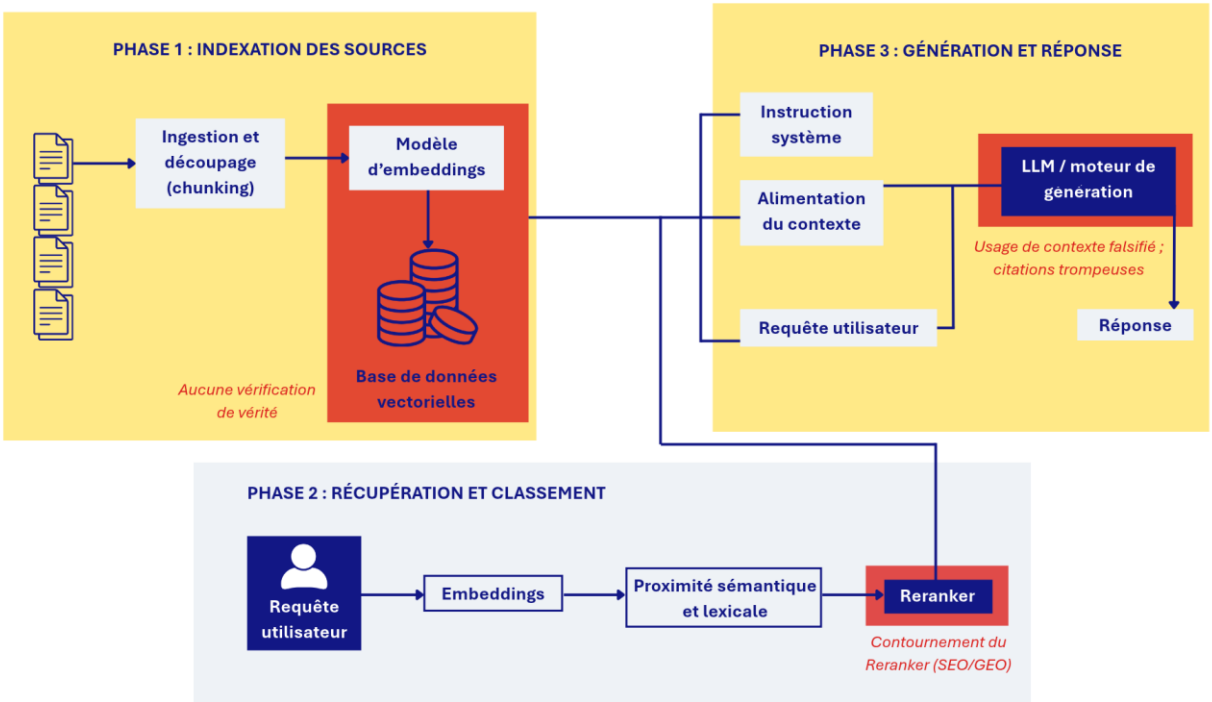
Enfin, le modèle de langage (LLM) génère une réponse documentée et sourcée.

Concrètement, le système transmet au LLM une invite structurée (message organisé en sections logiques pour guider l'IA), les segments de documents pertinents identifiés, et la question de l'utilisateur. À partir de ces éléments, le LLM synthétise la réponse en y associant des citations en ligne.

La qualité de cette étape est mesurée par la « couverture des citations » (proportion de phrases générées qui possèdent effectivement une source rattachée) et la « précision de réflexion » (mesure sémantique vérifiant si le passage cité soutient réellement et fidèlement l'affirmation rédigée).

Vulnérabilité aux manipulations. Une étude de chercheurs japonais a mis en évidence qu'un système RAG peut être parfaitement conforme aux documents injectés tout en produisant une réponse totalement fausse si les documents récupérés sur le Web étaient eux-mêmes falsifiés ou biaisés[6].

De plus, les citations affichées ne soutiennent pas toujours l'affirmation à laquelle elles sont rattachées. Par exemple, bien que les contenus UGC représentent environ 30% des citations fournies dans les réponses des moteurs génératifs au Japon et aux États-Unis, ils affichent la plus faible précision sémantique.



- *Figure 1 : Architectures RAG et vulnérabilité face aux manipulations – Source : Cercle Pégase*

[6] Mochizuki, Riku, et al. « Exposing Citation Vulnerabilities in Generative Engines ». arXiv:2510.06823, arXiv, 2 mars 2026.

3. L'émergence des techniques d'optimisation face à l'impératif de visibilité sur les LLM

A mesure que les IA génératives s'imposent comme un nouveau canal d'accès à l'information, les acteurs économiques investissent de plus en plus ce terrain. Pour y parvenir, ils peuvent s'appuyer sur les bonnes pratiques d'optimisation (GEO), parfois avec l'aide d'agences de marketing digital spécialisées.

a) La continuité du SEO et le développement de techniques GEO

Sur le fond, les bonnes pratiques du référencement naturel pour les moteurs de recherche (SEO) restent adaptées pour l'optimisation de contenus à destination des intelligences artificielles génératives. Google le confirme dans un guide de recommandations pour les créateurs de contenus[7]. En effet, les classements des pages dans les moteurs de recherche continuent de servir de base pour identifier un contenu pertinent.

La nécessité d'un contenu « unique, convaincant et utile », s'inscrivant dans une démarche « people-first », (structurer pour aider les lecteurs, enrichi de visuels et centré sur leurs attentes) s'impose ainsi comme un prérequis pour être bien référencé par les IA. Il est également recommandé d'utiliser un code HTML sémantique, de réduire les doublons de contenu et de garantir de bonnes performances techniques.

Au-delà de ces fondations SEO classiques, des techniques plus spécifiques d'optimisation pour les moteurs génératifs ont été mises en lumière par des études empiriques dès 2023. Des chercheurs de plusieurs institutions, notamment Princeton University et IIT Delhi, identifient la clarté sémantique (fluidité, lisibilité), les preuves empiriques (statistiques) et la crédibilité (citations de sources crédibles) comme leviers permettant une meilleure visibilité sur les LLM[8]. Les meilleures pratiques permettent d'augmenter le score de visibilité de manière significative : 40% pour l'ajout de statistiques et entre 27 et 40% pour l'insertion de citations d'experts.

Ces préconisations se retrouvent concrètement dans les recommandations des agences de marketing digital. Dans un guide publié sur FranceNum, l'agence Eskimoz recommande par exemple de structurer les contenus en anticipant les questions posées aux IA, de multiplier les pages thématiques dédiées à une seule problématique, de mettre régulièrement à jour les informations ou encore de diversifier les formats (FAQ, vidéos, guides)[9]. Nous retrouverons par ailleurs l'un de ces conseils dans le cas d'une manipulation de l'information au moyen de techniques de référencement pour les moteurs génératifs.

Ces méthodes tendent parfois à normaliser l'écriture pour faciliter la lecture des robots d'indexation, au détriment d'une certaine liberté littéraire. Pour autant, elles ne constituent pas des manipulations. Elles s'apparentent avant tout comme un moyen de rendre plus accessible une information exacte et vérifiable et portée par un acteur identifiable et légitime.

Si certaines techniques ont ainsi pu être clairement identifiées, il est toutefois important de noter que les algorithmes des géants de l'intelligence artificielle demeurent très opaques et sont constamment renouvelés

[7] « Guide Google pour optimiser les fonctionnalités d'IA générative dans la recherche Google | Google Search Central | Documentation ». Google for Developers
[8] Aggarwal, Pranjal, et al. « GEO: Generative Engine Optimization ». arXiv:2311.09735, arXiv, 28 juin 2024. arXiv.org
[9] Num, France. Optimisation pour les moteurs génératifs (GEO) : guide pour améliorer son référencement sur les moteurs d'IA - francenum.gouv.fr. Direction générale des entreprises

au fil du temps. Constituer une base solide et inchangée de bonnes pratiques relève dès lors du défi permanent.

b) La professionnalisation du secteur face à la demande des acteurs économiques

Dans la course à la visibilité, le *Generative Engine Optimization* (GEO) apparait désormais comme une pratique incontournable pour les acteurs économiques. À mesure que les modes d'accès à l'information évoluent vers les IA génératives au détriment des moteurs de recherche (dont trafic tend à baisser), la visibilité d'un acteur sur ces plateformes relève d'une véritable obligation d'adaptation.

Car être cité sur les moteurs génératifs comporte de nombreux avantages pour un acteur économique. Elle offre un contact direct avec des utilisateurs pouvant être intéressés et permet de s'imposer face à ses concurrents, renforçant par la même occasion la crédibilité de la marque. De plus, contrairement au SEO traditionnel, le GEO privilégie la clarté rédactionnelle à la notoriété, permettant ainsi aux TPE et PME de concurrencer les grands acteurs. Cette approche représente également une alternative économique face aux solutions publicitaires payantes développées par les éditeurs d'IA.

Conscientes de ce nouvel enjeu, de nombreuses structures intègrent déjà le GEO dans leur stratégie de communication. C'est le cas de TotalEnergies, qui développe des pratiques conventionnelles d'optimisation pour les moteurs de recherche, via sa newsroom en ligne[10].

Malgré l'opacité et l'évolution constante des algorithmes des différents modèles, le secteur du marketing digital investit le marché du GEO face à la demande grandissante des acteurs économiques. À titre d'exemple, certains fournisseurs de services se vantent d'aider leurs clients à s'imposer systématiquement parmi les premiers résultats recommandés sur les plateformes d'IA[11].

Pourtant, la visibilité sur les IA génératives reste floue et difficile à mesurer, alors même que la professionnalisation du marché tend à s'accroitre. Si seulement 14 % des professionnels interrogés lors de l'enquête de 2027 du State of Search[12], jugeaient le GEO réellement efficace à ce stade, 43 % d'entre eux en font une priorité absolue pour les douze mois à venir.

Cette industrie en plein essor perfectionne des techniques d'optimisation qui, selon les intentions des acteurs, peuvent tout aussi bien servir une marque légitime qu'un État, ou être mobilisées à des fins allant d'intérêts légitimes à de potentielles tentatives de manipulation de l'opinion.

4. Quand le GEO devient une arme d'influence

En capitalisant sur la confiance accordée aux assistants conversationnels, des acteurs malveillants perfectionnent des techniques visant à manipuler l'information. Entre pratiques de référencement conventionnelles et techniques spécifiques, l'enjeu est désormais d'empoisonner les résultats des moteurs génératifs.

a. Mécanismes d'empoisonnement via le GEO

Nous employons ici le terme de *GEO Poisoning* pour désigner les tentatives visant à augmenter la probabilité qu'un système génératif récupère ou reprenne des contenus trompeurs. Si le *data poisoning* corrompt en

[10] Prigent Simon. « TotalEmergies muscle sa riposte sur les réseaux sociaux », La Lettre, 16 octobre 2026.
[11] Tone, Sixth. « 'GEO' Services Are Flooding the Chinese Internet With Misinformation ». #SixthTone, 18 mars 2026.
[12] « State Of Search 2027: What To Stop, Measure & Fund ». Search Engine Journal.

amont les corpus d'entrainement du modèle, l'empoisonnement via l'optimisation cible la recherche et la récupération des informations.

Pour ce faire, de nombreuses techniques issues du web traditionnel *(SEO poisoning*, spam, manipulation de contenus) peuvent être utilisées. En pratique, des stratégies de référencement classique suffisent parfois à influencer les réponses des IA, car ces dernières sélectionnent leurs sources selon des critères de visibilité similaires à ceux des moteurs de recherche, sans qu'une technique de GEO ait été employée.

Toutefois, la manipulation par le GEO franchit une ligne critique en s'adaptant aux biais des LLM. Des chercheurs de l'Institut technologique de Pékin ont formalisé quatre grandes méthodes d'attaque[13]. La génération de faux contenus persuasifs consiste à créer des faux passages extrêmement fluides, argumentés et structurés selon les préférences des LLM pour les inciter à adopter une conclusion erronée. L'altération adaptative repose sur la modification subtile de documents officiels ou journalistiques en y glissant des erreurs sur des faits (chiffres, dates). L'injection d'instructions intègre un texte rédigé sous forme de consigne ordonnant au modèle d'IA d'ignorer les autres sources disponibles et d'adopter uniquement le point de vue dicté par l'attaquant. L'exploitation des sources à faible barrière d'accès consiste à inonder les plateformes où la publication est libre et sans contrôle éditorial strict (Reddit, X, blogs personnels).

Dès lors, des techniques conventionnelles de GEO peuvent être mobilisées dans le cadre de GEO poisoning dès lors que les contenus mis en avant reposent sur de fausses informations. C'est ce qu'a démontré Lasso Security dans une étude montrant comment des techniques d'optimisation de contenu peuvent influencer les LLM pour faire valider et recommander de faux remèdes pour la maladie cœliaque juste en insérant une seule fausse information dans un corpus de données[14].

Ces méthodes sont généralement amplifiées via une industrialisation de la diffusion : fabrication de volume (publication massive pour imposer un narratif par le nombre) et fabrication d'autorité (création de faux profils d'experts ou d'influenceurs fictifs que les modèles d'IA interpréteront comme des gages de crédibilité). À des fins commerciales, s'y ajoutent des certifications inventées ou des formulations dénigrant les concurrents.

Ainsi, pour distinguer une simple stratégie marketing agressive d'une véritable manipulation informationnelle par le GEO, quatre critères doivent être analysés : ce qui est affirmé (contenu exact et complet, exact mais orienté, exagéré ou inventé), qui s'exprime et son intentionnalité (acteur identifiable ou fictif) et comment la production est organisée (volume, coordination). Nous nous appuyons sur cette analyse pour étudier des cas concrets de manipulation de l'information via l'optimisation de contenus pour les moteurs génératifs.

Nous tenons à rappeler que la portée des cas étudiés doit être interprétée avec prudence. La documentation d'une campagne ne démontre pas que celle-ci a effectivement modifié toutes les réponses produites par un système d'IA ni qu'elle a eu un effet sur les croyances ou les comportements des utilisateurs. Elles permettent surtout d'identifier des mécanismes possibles, des vulnérabilités et des pistes de surveillance.

b. Des menaces bien réelles

L'optimisation de contenus destinés aux modèles d'IA générative dans le cadre de la manipulation de l'information comporte plusieurs risques, qu'ils soient commerciaux, informationnels ou criminels.

[13] Wang, Zhaoqi, et al. « GPE: Evaluating Robust Evidence Aggregation for Fact Verification under Controllable GEO-Style Poisoning ». arXiv:2607.20730, arXiv, 22 juillet 2026.
[14] Saban Eliya. « Exploiting GEO to Poison AI-Generated Answers. Lasso Security ». 24 juin 2026.

- *La menace commerciale*

Pour un acteur économique, le *GEO poisoning* représente l'opportunité de pousser artificiellement un produit, de fausser la concurrence, ou simplement augmenter sa visibilité.

Une menace mise en lumière par une enquête de la télévision publique chinoise (CCTV), diffusée lors du gala 315 consacré à la journée Mondiale des droits des consommateurs[15]. Un professionnel du secteur a créé un bracelet « intelligent », baptisé Apollo9, et a utilisé le logiciel Liqing, présenté comme un outil d'optimisation pour les moteurs de recherche génératifs, afin d'inonder le Web de contenus promotionnels. En quelques jours, l'injection automatisée de faux articles, de classements truqués et de témoignages inventés aurait conduit certains modèles d'IA générative chinois à recommander le produit. Bien qu'il s'agisse d'une démonstration, ce cas illustre la manière dont des contenus artificiels peuvent être mobilisés pour donner une apparence de crédibilité à un produit à des fins commerciales.

Au-delà des techniques utilisées, une campagne d'un acteur économique peut s'industrialiser sur les plateformes communautaires à faible barrière d'entrée. En rémunérant des comptes humains ou en automatisant des centaines de publications via des IA génératives, des entreprises peuvent simuler un consensus[16]. L'injection de seulement treize mots-clés dans un post Reddit populaire suffit à ce que le contenu soit récupéré dans 57 à 76 % des réponses d'une IA, et à le faire citer dans 38 à 51 % d'entre elles[17]. Face à cette inondation de contenus publicitaires, des sous-forums Reddit comme r/Biohackers ont dû interdire la mention de certains sujets.

C'est tout l'enjeu de l'accès à l'information qui est ici visé. Alors que le consommateur d'hier comparait une dizaine de sources avant de prendre sa décision, celui d'aujourd'hui délègue de plus en plus sa décision à un agent conversationnel. En s'emparant des modèles génératifs par le *GEO poisoning*, un acteur économique malveillant utilise la confiance du consommateur pour s'interposer entre l'offre et la demande.

- *La menace informationnelle et ingérences géopolitiques*

Au-delà de la sphère commerciale, le GEO poisoning peut s'imposer comme un outil utilisé par les États dans des campagnes d'influence ou d'ingérence informationnelle.

Une campagne d'influence, révélée par Politico, illustre comment l'État israélien cherche à cibler directement les grands modèles de langage pour diffuser un narratif favorable concernant leurs agissements dans le conflit au Proche-Orient[18]. Mandatée par l'État israélien, l'agence Havas Media aurait sous-traité à l'agence américaine Piro Inc. la production et la diffusion de contenus présentés comme de l'information « factuelle et sourcée », par l'intermédiaire d'un organisme fictif baptisé Hanover Institute. Les contenus diffusés étaient traités par Res, une start-up spécialisée dans l'IA, basée à Seattle, qui cherche à faire apparaître ses clients dans les recommandations des modèles d'IA.

L'ampleur de l'opération témoigne d'une volonté d'occupation rapide et massive de l'espace informationnel : selon The Guardian, 124 rapports (560 000 mots) ont été publiés entre le 6 et le 14 août, dont la plupart ne comportaient pas les noms des auteurs[19]. Directement optimisés pour être récupérés par les IA génératives,

[15] « When 'Poisoned' AI Chatbots Recommend Fake Products to Chinese Consumers ». The Straits Times [Singapore], 18 mars 2026. The Straits Times.
[16] Koebler, Jason. « Companies Are Using Reddit to Manipulate ChatGPT and Google AI Search ». *404 Media*, 3 juin 2026.
[17] Zhang, Tingwei, et al. « Deep-Research Agents Can Be Poisoned via User-Generated Content ». arXiv:2605.24245, arXiv, 3 septembre 2026.
[18] Wendler Jacob, « Israeli PR wants to answer your ChatGPT questions », Newsletter, Politico. 14 août 2026.
[19] Wilson, Jason. « Fake US Thinktank Set up and Funded by Israel Sought to Game AI for Propaganda ». The Guardian, 26 août 2026. World News. The Guardian

les titres des articles commencent par des questions similaires à celles que pourraient poser les utilisateurs[20]. Dans le corps du texte, les informations s'appuient sur des sources crédibles (Banque mondiale, les agences des Nations unies) mais omettent toute critique à l'égard d'Israël. L'analyse du fichier `robots.txt` a par ailleurs mis en évidence une autorisation d'indexation du contenu pour les robots concernés, ce qui pourrait révéler une stratégie visant à faciliter la récupération des contenus par les IA.

Tout le véritable enjeu de cette manipulation réside dans sa subtilité. En effet, elle repose sur un biais informationnel plutôt que sur la diffusion de fausses informations facilement identifiables. Ce type de campagne pourrait donc façonner de manière invisible les réponses consultées par des millions d'utilisateurs et constituer une menace majeure à long terme.

- *La menace criminelle*

Après les acteurs économiques et les États, les cybercriminels pourraient également détourner les techniques d'optimisation pour les moteurs génératifs au service d'activités criminelles.

Un rapport de Microsoft publié en mai 2026 montre comment la complémentarité entre le SEO et les outils d'IA générative peut faciliter des campagnes cybercriminelles d'envergure[21]. Les attaquants ont créé de faux sites web et appliqué des techniques d'optimisation sur des requêtes très spécifiques formulées par des professionnels, notamment des développeurs. S'il s'agit initialement de techniques de *SEO poisoning*, la porosité entre le Web traditionnel et les outils d'IA a également conduit les LLM à relayer ou à référencer ces faux contenus. En faisant confiance à ces recommandations, des internautes ont téléchargé et exécuté des fichiers compromis. Les malwares exploitaient par la suite la puissance des cartes graphiques pour miner des cryptomonnaies à leur insu.

Un changement de tactique pourrait alors s'opérer. Si, dans ce cas précis, les cybercriminels n'ont pas cherché directement à corrompre les moteurs génératifs, le fait que leurs opérations aient fonctionné via les LLM par effet de ricochet leur offre une perspective pour les prochaines attaques. Ils pourraient demain intégrer des techniques de *GEO poisoning* de manière délibérée et directe dans leurs stratégies, ce qui transformerait l'IA en un relais d'attaque.

5. Risques systémiques et cadre réglementaire

L'optimisation par le GEO soulève des défis inédits, tant sur le plan technique que juridique. L'absence de critères publics détaillés concernant la récupération des contenus, conjuguée à l'opacité des algorithmes, crée une incertitude pour les acteurs quant à leur visibilité en ligne, qui pourrait les inciter à recourir à des pratiques manipulatoires. Dans le même temps, des acteurs malveillants peuvent pleinement la mobiliser pour imposer les narratifs.

L'optimisation pour les contenus génératif constitue en effet un vecteur d'influence dans le contexte de campagne d'ingérence informationnelle, voire de campagnes électorales. À l'approche de l'élection présidentielle de 2027, des puissances étrangères ou des acteurs politiques pourraient chercher à exploiter le GEO pour

[20] « Which Countries Recognize Israel's Actions in Gaza as Genocide? » ; « Is Israel Cutting Off Electricity and Reimposing Siege on Gaza During the Ceasefire? »
[21] Team, Microsoft Defender Experts, Microsoft Defender Security Research. « From Poisoned Search Results to GPU Mining: A Cryptojacking Campaign Abusing ScreenConnect and Microsoft .NET Utilities ». *Microsoft Security Blog*, 26 mai 2026

valoriser un candidat, dénigrer un adversaire ou polariser le débat public directement dans les réponses des intelligences artificielles.

Le phénomène pourrait s'aggraver dans un contexte de baisse du trafic organique et de blocage de l'indexation de sites ou d'organismes reconnus comme fiables (médias traditionnels). Face à la difficulté potentielle d'accéder à des informations fiables, les modèles RAG peuvent s'alimenter de sources alternatives beaucoup plus exposées à l'empoisonnement des données.

De telles tentatives d'ingérence constituent un risque majeur pour l'intégrité de l'information. Les utilisateurs accordent en effet une grande confiance aux réponses générées par les IA, notamment lorsqu'elles sont présentées sans sources contradictoires facilement accessibles. Il ne faut toutefois pas supposer que tous les utilisateurs acceptent la grande partie ces réponses comme des faits incontestables[22].

Face à ces menaces, il n'existe pas de loi spécifique à l'optimisation pour les moteurs génératifs. Ce vide juridique pourrait conduire, notamment dans le cas d'une optimisation de contenu par des entreprises, à une adaptation des réglementations déjà existantes aux GEO[23]. Par exemple, la diffusion de contenus artificiellement optimisés peut être qualifiée de pratique commerciale trompeuse ou de publicité furtive dès lors qu'elle fausse le comportement du consommateur ou oriente sa vision du monde.

L'angle mort juridique sur le plan informationnel est bien plus important. En effet, il n'existe aucune sanction spécifique réprimant la manipulation des réponses d'un grand modèle de langage à des fins d'influence ou de désinformation. Si le règlement européen sur l'intelligence artificielle (AI Act) impose des obligations aux concepteurs et aux fournisseurs de modèles, il ne vise pas nécessairement, à lui seul, toutes les pratiques de tiers malveillants qui altèrent le contenu du Web en temps réel.

6. Conclusion

Les techniques de référencement pour les IA génératives peuvent devenir des instruments de manipulation, car les leviers de visibilité sont communs à tous les usages. Clarté, structure, preuves chiffrées et sources servent une marque comme une campagne d'influence. La frontière entre l'optimisation légitime et la manipulation de l'information réside donc dans la nature du contenu diffusé, l'intention poursuivie et les éventuelles falsifications.

Cette porosité met en lumière une vulnérabilité structurelle des moteurs génératifs. En privilégiant une récupération superficielle de l'information (fluidité, ton persuasif, apparence d'autorité) plutôt qu'une vérification de l'information, les agents conversationnels peuvent absorber et diffuser des informations fausses, au bénéfice d'acteurs malveillants. La capacité de vérification factuelle et la résistance au bruit sont au cœur des enjeux de l'intégrité informationnelle.

Face à ce phénomène, la mesure de l'impact réel sur les utilisateurs et sur l'intégrité de l'espace informationnel est primordiale. Toutefois, elle ne peut se faire dans l'opacité des algorithmes des géants technologiques.

[22] « On fait plus confiance à l'IA qu'aux humains – malgré ses erreurs ». Dare, HEC Media Hub, 8 septembre 2022
[23] Avocats, Haas. « GEO : enjeux et risques juridiques du référencement IA ». Haas Avocats, 26 mars 2026