La fuite de données n'aura pas lieu un jour, elle a déjà eu lieu lundi matin

***Par** Myriam BELAID - septembre 2026*

Une simple demande de devis et soudain, une fuite de données.

Lundi matin.

Vous êtes dans votre bureau. La période de rush fiscal bat son plein et, comme souvent à cette période de l'année, tout le monde manque de temps.

Dans votre cabinet d'expertise comptable, une vingtaine de collaborateurs travaillent sur plusieurs dossiers en même temps. Un associé en retard sur plusieurs dossiers de conseil, cherche une solution pour aller plus vite.

Pour gagner du temps, il décide d'utiliser l'intelligence artificielle pour résumer rapidement des documents fiscaux clients.

Il ouvre ChatGPT avec son compte personnel, puis copie-colle les documents dont il a besoin.

En quelques secondes, un résumé clair et exploitable apparaît. Le résultat est bon. Il passe au dossier suivant.

Le travail avance. Personne ne se pose vraiment de question.

Puis quelques jours plus tard, une cliente de longue date contacte le dirigeant du cabinet.

Elle lui pose une question qui semble, au départ, assez simple :

Sa déclaration fiscale a-t-elle, d'une manière ou d'une autre, pu servir à entraîner une intelligence artificielle ?

Le cabinet commence alors à vérifier ce qui s'est passé.

Il faudra plusieurs heures pour comprendre que le compte utilisé n'était pas un compte professionnel encadré, mais un compte grand public. Les données avaient donc été transmises à un service dont les conditions d'utilisation ne garantissaient pas la même protection des informations.

Parmi les documents transmis figuraient des informations fiscales sensibles appartenant aux clients du cabinet.

Personne n'avait cherché à mal faire.

Il s'agissait simplement d'un raccourci pris pour gagner quelques minutes sur un dossier.

Mais ce raccourci a suffi à faire sortir des données confidentielles du périmètre du cabinet.

Un gain de temps de quelques secondes peut parfois entraîner des conséquences qui elles, durent beaucoup plus longtemps.

Prenons le temps d'étudier tout cela.

Quand on parle de fuite de données, les premiers éléments qui nous viennent à l'esprit sont souvent une intrusion sur un ordinateur portable par un attaquant, un mail frauduleux ou encore un vol de mot de passe.

Ces situations sont bien réelles, mais elles occultent une autre réalité plus courante souvent méconnue.

Elle est d'autant plus difficile à identifier qu'elle peut survenir de manière tout à fait banale dans notre quotidien.

Elle peut se produire un lundi matin, autour d'un café, lorsqu'une assistante souhaite supprimer les doublons de son fichier clients et copie celui-ci dans une intelligence artificielle pour effectuer l'opération plus rapidement.

C'est ce que l'on nomme le « **Shadow AI** », c'est-à-dire l'usage d'outils d'intelligence artificielle par des collaborateurs en dehors d'un cadre défini par l'entreprise.

En effet, contrairement aux idées reçues, une fuite de données n'est pas toujours le fruit d'une cyberattaque. Elle peut naître d'un simple réflexe de productivité : utiliser une intelligence artificielle pour gagner du temps. C'est ce qui la rend aussi fréquente que difficile à identifier.

Seulement, derrière cette simplicité se cachent de véritables enjeux de confidentialité et de protection des données.

Car une fois l'information transmise, la question n'est plus seulement de savoir si une fuite a eu lieu. Le véritable sujet devient celui de ses conséquences : **qui pourra exploiter ces données et dans quel objectif ?**

Quels chemins empruntent les pirates ?

Plus inquiétant encore, un cybercriminel n'a pas nécessairement besoin de pirater un système ou de contourner des mesures de sécurité pour obtenir des informations confidentielles.

Ces données confidentielles lui parviennent par divers chemins identifiés [2] et notamment :

- La fuite d'informations confidentielles via l'IA (OWASP LLM02) ;
- La réutilisation des données transmises aux IA pour l'entraînement des modèles ;

- La compromission d'un fournisseur ou d'un composant utilisé par l'IA (OWASP LLM04) ;
- La manipulation des réponses de l'IA par des instructions malveillantes (OWASP LLM01).

La question qui en dit trop ?

La divulgation d'informations sensibles

Il s'agit du cas le plus courant : un collaborateur copie-colle des données sensibles dans un outil d'intelligence artificielle afin de gagner du temps ou d'obtenir une réponse plus rapidement.

Dans ce contexte, ces données sensibles peuvent alors ressortir dans les réponses faîtes à d'autres utilisateurs par fragments par exemple, lorsque ce dernier formule une question en lien direct avec celle de l'utilisateur lésé ou aborde des problématiques proches ou convergentes.

« Par exemple, dans notre cabinet, le collaborateur fait un copié-collé des documents fiscaux clients dans une IA pour gagner du temps. Ces données quittent alors le périmètre de l'entreprise et peuvent être exposées à des usages qu'elle ne maîtrise plus ».

C'est notamment ce qu'il s'est produit le 20 mars 2023. *Un bug dans une bibliothèque open source a permis à certains utilisateurs de ChatGPT de voir les titres de conversations d'autres comptes et parfois le premier message d'une conversation avec des informations sensibles, dont des informations de paiement [3].*

L'enseignement est simple : dès lors que la donnée quitte le périmètre de l'entreprise, celle-ci cesse d'en avoir la maîtrise complète et s'expose à des risques qu'elle ne contrôle plus.

Partagé une fois, exploité pour toujours ?

L'utilisation des données des utilisateurs pour l'entraînement des modèles IA

Un autre risque découle directement des informations saisies dans les outils d'intelligence artificielle.

Afin de fournir des réponses toujours plus pertinentes, certains services utilisent les données transmises par leurs utilisateurs pour entraîner ou améliorer leurs modèles.

Le risque réside dans le fait que les informations partagées ne servent plus uniquement à répondre à une demande ponctuelle. Elles peuvent contribuer à enrichir les connaissances du modèle et influencer les réponses fournies ultérieurement à d'autres utilisateurs.

« Pour retourner à notre cabinet, si un collaborateur transmet à une IA des informations détaillées sur les tarifs les marges ou une méthode de travail, ces données peuvent être utilisées pour améliorer le modèle. Elles pourraient alors indirectement influencer les réponses données à d'autres utilisateurs, voire permettre à un concurrent de retrouver une partie de ces informations ».

Pour une TPE ou une PME dont la valeur repose souvent sur une expertise ou un procédé spécifique, cette perte de maîtrise représente un risque important. La diffusion involontaire

d'informations stratégiques peut réduire un avantage concurrentiel, fragiliser la position de l'entreprise sur son marché et, dans certains cas, bénéficier directement à des concurrents.

Ce risque est d'autant plus difficile à maîtriser que les données sont fréquemment traitées ou hébergées sur des infrastructures situées hors de l'Union européenne. L'entreprise dispose alors d'une visibilité limitée sur leur localisation, leur durée de conservation et les conditions de leur réutilisation, ce qui renforce les enjeux de confidentialité et de souveraineté des données.

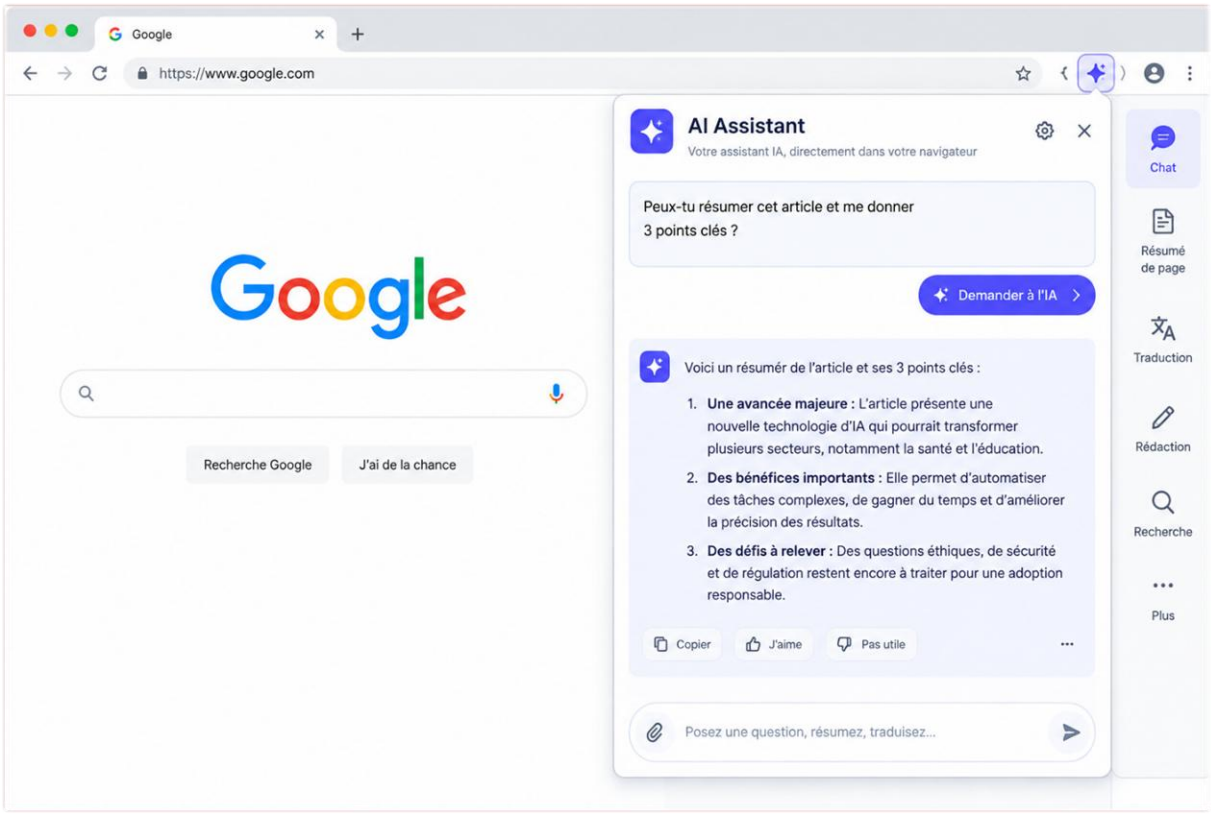
Piraté par ricochet ?

La compromission de la chaîne d'approvisionnement

Parmi les vecteurs de compromission les plus discrets on trouve la compromission d'une extension de navigateur utilisée pour interagir avec une intelligence artificielle.

En effet, une extension dispose souvent d'autorisations lui permettant d'accéder au contenu affiché dans le navigateur et se met régulièrement à jour. Si son éditeur est compromis ou diffuse une mise à jour malveillante, celle-ci peut être déployée automatiquement chez tous les utilisateurs. L'extension peut alors collecter discrètement les données affichées ou saisies, comme des devis, des informations clients ou des documents internes et les transmettre à un tiers.

L'utilisateur continue d'utiliser son outil en toute confiance, tandis que les données de l'entreprise quittent silencieusement son périmètre de contrôle.



L’attaque était là mais personne ne l’a vue ?

Les attaques par injection de prompt

Au-delà des outils qui entourent l'IA, le contenu qu'elle traite peut lui-même devenir un vecteur d'attaque. C'est le principe des attaques par injection de prompt.

Cette attaque consiste pour un attaquant à cacher dans un document soumis à une IA (PDF, e-mail, page web), une instruction que vous ne voyez pas mais qu’elle lit et prend pour consigne légitime.

Si l'assistant IA dispose d’un accès aux fichiers de l'entreprise (par l’intermédiaire de SharePoint, OneDrive, Google Drive, Microsoft 365 Copilot etc.) et que les mécanismes de protection sont insuffisants ou inexistants, il peut alors exécuter cette instruction et divulguer des informations sensibles.

« Dans notre cas, aucune instruction cachée n’était présente dans les documents transmis à l’IA. Mais si un document avait contenu une instruction malveillante dissimulée, l’IA aurait pu être manipulée pour accéder à d’autres fichiers du cabinet et divulguer des informations confidentielles ».

Le danger est que l'attaque se dissimule dans un document apparemment légitime, mais qui s’avère en fin de compte devenir un vecteur de fuite de données.

Un perdant, plusieurs gagnants

- **Le concurrent**

Il constitue la réponse la plus évidente à la question : « À qui profite le crime ? ». Si vos informations commerciales tombent entre ses mains, celui-ci pourra analyser vos marges ou vos conditions d'achat afin d'ajuster ses propres offres.

Il pourra alors proposer des prix plus compétitifs que les votre, de cibler vos clients les plus rentables et de remettre en cause l'avantage commercial que vous aviez construit au fil des années.

- **Le fournisseur d’IA**

Dans certains cas, le fournisseur d'IA peut utiliser les données saisies par ses utilisateurs pour améliorer ou entraîner ses modèles, lorsque cette pratique n'est pas explicitement exclue par des dispositions contractuelles ou des paramètres techniques.

Dans ce cas de figure, ce que vous saisissez dans l’outil ne sert plus seulement à générer une réponse pour vous : ces données peuvent aussi alimenter l’entraînement des algorithmes du fournisseur. Cela lui permet ainsi d’améliorer ses services existants ou d’en créer de nouveaux.

Même sans intention malveillante, cette pratique peut faire perdre à l’entreprise le contrôle sur ses informations sensibles. Des données relatives à vos projets, à votre stratégie, vos méthodes de travail ou à vos clients risquent alors d’être conservées au-delà de leur usage initial.

- **L'attaquant**

Dans une démarche plus discrète, l'attaquant n'a pas besoin d'exploiter immédiatement les données récupérées. Il peut les utiliser pour mieux vous connaître et construire des attaques sur mesure.

En reprenant le jargon de l'entreprise, le nom de vos fournisseurs ou vos habitudes de paiement, il est en mesure de produire des e-mails quasiment impossibles à distinguer d'un échange légitime. Son objectif n'est alors plus seulement de détenir l'information, mais de s'en servir pour gagner votre confiance et obtenir un accès à votre système d'information.

Les données pourront également être revendues à d'autres acteurs malveillants ou conservées pour être exploitées dans le cadre d'attaques futures, visant votre entreprise, vos clients ou vos partenaires.

Comme le veut l'adage, la discrétion est mère de toutes les vertus. En matière de cybersécurité, elle est parfois aussi celle de tous les vices.

- **Le client**

Si les informations concernent vos clients ou révèlent un manque de maîtrise de vos données, la conséquence serait une perte de confiance de vos clients.

Un client qui apprend que des informations le concernant ont été exposées pourra s'interroger sur votre capacité à protéger ses données et celles de son activité. Dans certains cas, cette perte de confiance peut conduire à la remise en cause d'une relation commerciale construite depuis plusieurs années, voire à son départ vers un concurrent.

Au fond, lorsqu'une donnée s'échappe, c'est souvent la confiance qui sort en premier.

Cela n'arrive pas qu'aux autres ?

Le cas Samsung

En mars 2023, *The Economist Korea*, confirmé par la suite par *Bloomberg*, révèle que trois incidents distincts ont eu lieu chez Samsung en l'espace de vingt jours [4].

En effet, des ingénieurs de l'entreprise ont copié dans ChatGPT du code source confidentiel, des séquences de programmes destinées à corriger des anomalies ainsi que le contenu de réunions internes, dans le but de résoudre plus rapidement des problèmes techniques ou de gagner du temps.

Ces collaborateurs n'avaient aucune intention malveillante. Pourtant, en transmettant ces informations à une intelligence artificielle externe, ils ont fait sortir du périmètre de contrôle de l'entreprise des données relevant directement de sa propriété intellectuelle et de ses activités stratégiques.

L'incident a mis en lumière une réalité : même une entreprise disposant de moyens considérables en cybersécurité peut voir ses données sensibles exposées par un simple usage non encadré de l'IA.

Face à cette perte de maîtrise de l'information, Samsung a finalement interdit l'usage des IA génératives en interne et accéléré la mise en place de solutions mieux contrôlées.

Ce cas illustre parfaitement que la fuite de données ne résulte pas toujours d'une attaque. Elle peut être la conséquence d'une recherche légitime de productivité, transformant un simple copier-coller en incident de sécurité.

Si même une entreprise comme Samsung, dotée d'importants moyens humains, techniques et juridiques, peut être confrontée à ce type d'incident, qu'en est-il des TPE et PME ?

Dans ces structures, le risque est souvent plus élevé. Chacun peut être tenté d'utiliser l'outil gratuit qui lui semble le plus pratique sans nécessairement connaître ses conditions d'utilisation.

Plus préoccupant encore, les TPE et PME disposent rarement des mêmes capacités qu'un grand groupe. Elles ne bénéficient pas toujours d'équipes juridiques, de DPO ou d'experts en cybersécurité capables d'encadrer ces usages et d'en gérer les conséquences.

Enfin, les données exposées sont souvent celles qui constituent le cœur même de leur activité : marges commerciales, fichiers clients, conditions d'achat, devis, savoir-faire ou projets en cours.

Pour une grande entreprise, une fuite de données est un incident. Pour une TPE ou une PME, elle peut toucher directement ce qui fait sa compétitivité, sa réputation et parfois même sa pérennité.

Que faire concrètement ?

Sans disposer des moyens d'un grand groupe, une TPE ou une PME peut déjà réduire significativement son exposition au risque grâce à quelques mesures simples, efficaces et peu coûteuses :

La sensibilisation des collaborateurs

Il s’agit de l'un des leviers les plus efficaces à mettre en place.

Avant de transmettre une information à une IA, chaque utilisateur devrait s'interroger sur la sensibilité des données qu'il s'apprête à partager et sur les conséquences qu'aurait leur divulgation à un tiers ou à un concurrent. Une règle simple peut être appliquée au quotidien : ne jamais communiquer à une IA une information que l'entreprise ne souhaiterait pas voir diffusée à l'extérieur de son périmètre de contrôle.

Cette vigilance doit également s’accompagner d’actions de sensibilisation et de formation régulières, afin d'ancrer les bonnes pratiques et de permettre aux collaborateurs d'identifier plus facilement les informations qui ne doivent pas être communiquées à une IA.

- ✓ **Utiliser une seule IA officielle (avec compte nominatif**

  Définir un outil unique validé par l'entreprise, permet d'éviter que chaque collaborateur utilise sa propre IA selon ses habitudes. L'utilisation de comptes nominatifs assure également une meilleure traçabilité des activités et facilite l'identification d'éventuels incidents.

- ✓ **Anonymiser les données avant de les intégrer dans un prompt**

Avant de transmettre une information à une IA, il est recommandé de remplacer les données permettant d'identifier une personne, un client ou un fournisseur ou autres informations sensibles, par des informations génériques.
Par exemple : Remplacer le nom d'un client par « Client A » ou supprimer toute donnée sensible dans le prompt.

- ✓ **Privilégier les offres professionnelles**

  Lorsque l'IA est utilisée régulièrement dans l'entreprise et par de nombreuses personnes, il est préférable de privilégier une offre professionnelle souscrite et gérée par l'entreprise, plutôt que des comptes personnels et des versions gratuites utilisés individuellement. Cela permet de mieux maîtriser les accès, les paramètres de confidentialité et les conditions d'utilisation des données.

- ✓ **Classifier les données et mettre en place un DLP**

  Toutes les données ne présentent pas le même niveau de sensibilité. Une classification simple, distinguant par exemple les données **publiques**, **internes** et **confidentielles**, aide les collaborateurs à savoir ce qui peut ou non être partagé avec une IA. L'ajout d'un dispositif de prévention des fuites de données (DLP) permet de détecter, voire de bloquer, l'envoi d'informations sensibles vers des services externes.

- ✓ **Vérifier que les données ne sont pas réutilisées pour entraîner l'IA**

  Avant d'utiliser un service d'IA, il est important de vérifier ses conditions de confidentialité et lorsque l'option existe, de désactiver la réutilisation des données pour l'entraînement des modèles.

- ✓ **Interdire ou auditer les extensions IA dans les navigateurs**

  Les extensions liées à l'intelligence artificielle sont souvent installées pour des raisons de confort ou de productivité. Pourtant, elles peuvent bénéficier d'autorisations étendues sur le navigateur et accéder à des informations sensibles. Leur déploiement devrait être limité aux extensions validées par l'entreprise ou faire l'objet d'un contrôle régulier.

- ✓ **Privilégier lorsque cela est possible un hébergement en France ou en Europe**

  Lors du choix d'une solution d'IA, il est recommandé de privilégier des modèles dont les données sont hébergées en France ou au sein de l'Union européenne. Cette démarche permet de garantir le respect réglementaire en matière de protection des données personnelles et de sécuriser les conditions de traitement, de stockage et de transfert des informations.

- ✓ **Envisager une IA hébergée dans l'entreprise pour les données les plus sensibles**

  Pour certaines activités impliquant des informations particulièrement sensibles ou stratégiques, il peut être pertinent d'envisager une solution d'IA hébergée directement dans l'infrastructure de l'entreprise (« on-premise »).

Cette approche permet de conserver les données au sein du périmètre de l'organisation, de limiter leur exposition à des tiers et de conserver une maîtrise de leur traitement. L'entreprise garde ainsi davantage de contrôle sur ses informations les plus sensibles et sur les conditions dans lesquelles elles sont utilisées.

- ✓ **Établir une liste rouge des informations à ne jamais communiquer à une IA**

  Certaines données ne devraient jamais être intégrées dans un prompt : mots de passe, données clients, marges, contrats, devis ou code source.

  Formaliser cette liste et la communiquer aux collaborateurs permet de fixer un cadre clair, compris de tous et applicable au quotidien.

- ✓ **Réaliser régulièrement un exercice de mise en situation**

  Une fois par semaine, par exemple le vendredi en fin de journée, il peut être utile de se poser une question simple : *« Si j'étais mon principal concurrent, quelles informations aimerais-je obtenir sur mon entreprise ? »* La réponse met souvent en évidence les données réellement stratégiques de l'activité et rappelle celles qui ne devraient jamais quitter le périmètre de l'entreprise.

Conclusion

Soyez vigilants et attentifs à votre usage de L'IA et elle sera en retour un formidable levier de performance pour vous et votre entreprise.

Sources

[1] Exemple en introduction https://efros.com/incident-stories/ai-data-exposure-at-cpa-firm/?utm

[2] OWASP GenAI LLM TOP 10 https://genai.owasp.org/resource/owasp-genai-llm-top-10-2026/

[3] Exemple ChatGPT : https://openai.com/index/march-20-chatgpt-outage/

[4] Samsung Leak https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-ai-use-by-staff-after-leak