NL-Q · FICHE-MÉMO

Quatre fiches demandées, quarante modifiées : un cas imaginé, pas à pas

Ce qui a été demandé.
Corriger l'adresse d'un client qui déménage : quatre fiches dans le logiciel de gestion, une seule nouvelle adresse.

Ce qui s'est passé.
L'assistant a choisi les fiches par un mot du nom, pas par le numéro client. Quarante fiches modifiées, et un compte rendu sans aucun nombre.

Ce que contient la fiche.
Le cas pas à pas : la demande, l'écriture, le signal, le retour arrière, la nouvelle règle, puis huit questions de contrôle.

Pour les dirigeants et responsables administratifs de TPE et PME, non développeurs, dont un assistant IA peut déjà modifier un tableur, une base ou un logiciel de gestion

Édition du 19/11/2026 · NLQ

01 Lisez la demande telle qu'elle a été tapée

Ce cas est imaginé pour cette fiche : l'entreprise, les personnes, les fiches et les nombres du cas (4, 36, 40) sont inventés et ne sont pas des statistiques. Les recommandations citées, elles, viennent de l'OWASP, de l'ANSSI et de Cybermalveillance.gouv.fr.

Une PME de négoce de matériaux : le gérant, une assistante commerciale, un préparateur qui organise les livraisons. Les fiches clients vivent dans le logiciel de gestion. Quelques mois plus tôt, un prestataire y a branché un assistant IA pour un chantier de nettoyage : doublons, majuscules, codes postaux. Pour ce chantier, le compte de l'assistant a reçu le droit de modifier les fiches. Le chantier fini, personne ne le lui a retiré.

Un client déménage. Il a quatre fiches : son siège, son dépôt, deux adresses de chantier. Le gérant tape : « Mets à jour l'adresse de [nom du client], il a déménagé zone artisanale du Moulin. » Le nom du client contient celui de sa ville, comme beaucoup d'autres fiches de la base.

1 **La demande** — un client, quatre fiches → 2 **L'écriture** — quarante fiches touchées → 3 **Le compte rendu** — « c'est fait » → 4 **Le signal** — une adresse, trois clients → 5 **Le retour** — trente-six fiches rétablies

02 Regardez ce que l'assistant a fait

L'assistant n'a été ni piraté ni piégé : il a interprété une demande ambiguë. L'OWASP range l'hallucination et les consignes mal posées parmi les déclencheurs courants d'actions dommageables, et précise que le risque existe quelle que soit la cause de l'erreur. Ce qui fait la taille des dégâts, c'est ce que l'assistant a le droit de faire.

Pour retrouver le client, l'assistant a cherché les fiches dont le nom contenait le mot de la ville. Il en a trouvé quarante : les quatre du client, et trente-six autres clients de la même ville. Il a écrit la nouvelle adresse dans les quarante, puis a répondu : « C'est fait, l'adresse est à jour. » Aucun nombre, aucune liste.

L'OWASP ramène ce risque à trois causes. Le cas les réunit toutes.

CAUSE NOMMÉE PAR L'OWASP	DANS LE CAS	CE QUI L'AURAIT LIMITÉE
Fonctions excessives	L'outil de nettoyage savait modifier n'importe quelle fiche	Une action nommée : changer l'adresse d'une fiche désignée
Droits excessifs	Le droit d'écrire, accordé pour le chantier, était resté	Un compte en lecture seule une fois le chantier fini
Autonomie excessive	Aucune validation avant l'écriture	Une personne qui valide la liste des fiches avant

03 Repérez le signal, puis comptez

Personne n'a relu les fiches après l'écriture. Le signal est venu du terrain, le lendemain matin ; c'est le compte, pas la réponse de l'assistant, qui a mesuré l'erreur.

Le signal.
Le préparateur voit la même adresse sur trois bons de livraison de clients différents. Il prévient le gérant au lieu de corriger lui-même : trois corrections à la main auraient brouillé ce qu'il fallait retrouver.

Comptez sans l'assistant.
Première question, posée directement au logiciel : combien de fiches portent la nouvelle adresse ? Quarante, pour quatre attendues.

Ne lui demandez pas de vérifier son travail.
Son compte rendu était déjà faux par omission. L'OWASP met en garde contre la confiance excessive dans un contenu produit par l'IA, accepté sans vérification, et recommande un contrôle humain pour les informations critiques.

Remerciez celui qui a signalé.
Une anomalie signalée tôt se répare ; une anomalie corrigée en silence ne se retrouve plus.

04 Revenez en arrière sans perdre la journée

Rétablir toute la base dans son état de la veille aurait effacé les commandes saisies le matin même. Il fallait défaire l'écriture de l'assistant, et seulement elle. Les étapes ci-dessous illustrent l'arbre de la fiche « Avant qu'une IA écrive dans vos données : l'arbre du retour arrière », sur nl-q.com.

Coupez l'assistant d'abord.
Le gérant fait retirer au compte de l'assistant le droit d'écrire, avant toute réparation. L'équipe travaille comme avant l'assistant : l'ANSSI recommande de prévoir une procédure de contournement du système d'IA pour répondre aux besoins métier (R15).

Ouvrez la sauvegarde à part.
La sauvegarde de la veille est restaurée dans un dossier d'essai, jamais sur la base en service. Cybermalveillance.gouv.fr conseille de tester régulièrement ses sauvegardes, par exemple en les copiant dans le système original : dans le cas, l'essai avait déjà été fait, et l'on savait l'ouvrir.

Comparez une seule colonne.
On liste les fiches dont l'adresse diffère entre la sauvegarde et la base : quarante. Quatre sont celles du client qui déménage. Les trente-six autres reprennent l'adresse lue dans la sauvegarde.

Recomptez, puis refaites les documents.
Quatre fiches portent désormais la nouvelle adresse, et les commandes du matin sont intactes. Les bons de livraison édités entre-temps avec une mauvaise adresse sont refaits.

05 Changez la règle, pas seulement les fiches

La réparation a rendu les bonnes adresses ; elle n'empêche pas la prochaine erreur. Le gérant a écrit une règle nouvelle, réglée ensuite par le prestataire.

Lecture seule par défaut.
L'OWASP décrit l'erreur exacte du cas : un compte qui n'a besoin que de lire et qui peut aussi modifier et supprimer. Le détail du réglage est dans la fiche « Donner à une IA l'accès à vos données : le moindre accès en pratique ».

Une proposition chiffrée, pas une écriture.
L'assistant rend un tableau : numéro client, valeur actuelle, valeur proposée, et le nombre de lignes en tête. L'assistante commerciale le relit et l'applique. L'OWASP recommande l'accord d'une personne avant toute action à fort impact ; l'ANSSI, qu'un système d'IA ne puisse pas exécuter seul des actions critiques (R9).

Le numéro client, jamais le nom.
Une fiche se désigne par son identifiant. Quatre numéros listés, quatre lignes modifiées, pas une de plus.

Des petits lots et une trace.
L'OWASP propose de limiter le nombre d'actions possibles sur une période, pour repérer une dérive avant qu'elle ne s'étende. L'ANSSI recommande de journaliser les requêtes des utilisateurs et les appels de l'assistant aux données, pour reconstituer entièrement un événement (R29).

Un droit vérifié par le logiciel.
« Ne modifie rien sans me demander » n'est pas un réglage. L'OWASP recommande que l'autorisation soit vérifiée par le système qui reçoit la demande, au lieu de laisser l'assistant décider.

Comment savoir si c'est en place

Huit questions fermées. Chaque « non » renvoie à une section de la fiche.

- [ ] Le compte de votre assistant est-il en lecture seule sur vos fiches clients et vos stocks ?
- [ ] Une modification proposée désigne-t-elle les fiches par leur numéro, pas par leur nom ?
- [ ] Une personne nommée valide-t-elle la liste avant qu'elle soit appliquée ?
- [ ] Pouvez-vous dire qui a modifié une fiche hier, et ce qu'elle valait avant ?
- [ ] Un droit d'écrire accordé pour un chantier est-il retiré à la fin du chantier ?
- [ ] La proposition annonce-t-elle le nombre de lignes touchées avant toute écriture ?
- [ ] Savez-vous restaurer une sauvegarde à part, sans écraser la base en service ?
- [ ] Votre équipe sait-elle travailler sans l'assistant le jour où on le coupe ?

VOTRE PROCHAINE ÉTAPE

Venez avec la dernière modification faite par votre assistant, ou celle que vous voulez lui confier : on regarde ensemble comment vous la verriez et comment vous la déferiez.

www.nl-q.com/rendez-vous/

Réserver un échange gratuit

Sources et portée

Toutes les sources ont été ouvertes le 19/11/2026.

- OWASP Gen AI Security Project, *LLM06:2025 Excessive Agency* : déclencheurs, trois causes, droits de modifier et de supprimer inutiles, accord humain, autorisation vérifiée par le système, journal, limitation du nombre d'actions. https://genai.owasp.org/llmrisk/llm062025-excessive-agency/ (consulté le 19/11/2026)
- OWASP Gen AI Security Project, *LLM09:2025 Misinformation* : hallucination, confiance excessive, contrôle humain. https://genai.owasp.org/llmrisk/llm092025-misinformation/ (consulté le 19/11/2026)
- ANSSI, guide ANSSI-PA-102 *Recommandations de sécurité pour un système d'IA générative*, version 1.0 du 29/04/2024 : R9, R15, R29. https://messervices.cyber.gouv.fr/documents-guides/Recommandations_de_s%C3%A9curit%C3%A9_pour_un_syst%C3%A8me_d_IA_g%C3%A9n%C3%A9rative.pdf (consulté le 19/11/2026)
- Cybermalveillance.gouv.fr, *Pourquoi et comment bien gérer ses sauvegardes ?* : version la plus récente, test de la sauvegarde. https://www.cybermalveillance.gouv.fr/tous-nos-contenus/bonnes-pratiques/sauvegardes (consulté le 19/11/2026)

Portée : le cas est imaginé, aucune entreprise réelle n'est décrite, et ses nombres ne mesurent rien. Aucun logiciel n'est nommé. Les obligations liées aux données personnelles ne sont pas traitées. Ni audit, ni avis juridique.

Édité par NLQ — NL-Q (www.nl-q.com), avec le soutien d'ALMA, sponsor. Édition du 19/11/2026. Exemples fictifs ; ce qui est dit des textes cités vient des sources datées ci-dessus.