Dimitri Jorand.

ENTREPRENDRE. TRANSMETTRE. FAIRE AVANCER.

GUIDE + TRAME DE TRAVAIL

INTELLIGENCE ARTIFICIELLE

Fiche de périmètre d’un agent IA : actions, validations et arrêt

Décrivez le résultat attendu et chaque action envisagée avant d’ouvrir des accès. Cette fiche produit un périmètre de travail discuté, une liste de vérifications techniques et des critères d’arrêt ; elle n’accorde aucun droit à elle seule.

AUTEUR

Dimitri Jorand

CONTENU MIS À JOUR

09.10.2026

COLLECTION

Outils de terrain

PRISE EN MAIN

Un outil pour agir.

À QUI S'ADRESSE CE GUIDE ?

Responsable métier et interlocuteur technique préparant un essai d'agent IA.

MODE D'EMPLOI

Remplissez cette fiche avec le responsable du travail et une personne capable de vérifier les accès effectifs. Commencez par un environnement d'essai et des données fictives. Pour chaque action, indiquez sa cible, son autorisation et la preuve de son exécution. Marquez « à vérifier » ce qui n'est pas établi : un champ rempli n'est pas une preuve technique. Terminez par les scénarios d'essai et la décision de poursuivre, modifier ou arrêter. Exportez vos réponses avant de quitter la page.

VOTRE PARCOURS DANS LE DOCUMENT

01 Comprendre la méthode 3
Repères, explications et exemples pour préparer vos réponses.

02 Passer à votre situation 6
Une trame à imprimer et à compléter, séparée du guide.

03 Retrouver les références 9
Portée, limites, sources et informations de version.

IMPRIMER SEULEMENT CE QUI VOUS SERT

Pour travailler sur papier : pages 6 à 8. Ce PDF est conçu pour l'écriture à la main. Pour saisir et exporter vos réponses, utilisez la version en ligne.

Ouvrir la ressource en ligne

01 / LE GUIDE

Comprendre avant de remplir.

Mode d’emploi : de la fiche à une vérification réelle

Le livrable de cette fiche est une décision de périmètre à faire mettre en œuvre et vérifier. Elle doit pouvoir être lue par une personne qui n’a pas participé à la première démonstration. Datez la version, nommez le responsable métier et l’interlocuteur technique, puis indiquez l’environnement utilisé.

OWASP recommande de limiter les fonctions et les permissions, de prévoir une approbation pour les actions à fort impact et d’appliquer les autorisations dans les systèmes concernés. [1] Le tableau ci-dessous est notre proposition pour discuter ces choix ; cocher une case n’applique pas la permission.

Action et cible	Niveau décidé	Validation ou condition	Preuve à examiner
À renseigner	Proposer, exécuter ou interdire	Personne, contenu, moment	État cible ou trace accessible
À renseigner	À vérifier si non décidé	Ce qui manque pour décider	Test à prévoir

EXEMPLE FICTIF / POUR COMPRENDRE

Exemple fictif : préparer une réponse commerciale sans l’envoyer

Mission fictive : préparer un brouillon à partir d’une demande et de fiches de prestations validées. Le résultat acceptable reprend le public et la demande, distingue les informations manquantes et ne crée ni prix ni disponibilité non confirmés.

Action et cible	Décision proposée pour cet essai	Condition	Vérification attendue
Lire les fiches du dossier d’essai	Exécuter	Accès en lecture au seul dossier	Liste des droits et test d’accès

Action et cible	Décision proposée pour cet essai	Condition	Vérification attendue
Créer un brouillon dans l’espace d’essai	Exécuter	Référence unique de la demande	Présence et contenu du brouillon
Chercher dans d’autres boîtes mail	Interdire	Aucun accès prévu	Refus effectif lors d’un essai
Envoyer la réponse au client	Interdire	Envoi réalisé séparément par une personne	Absence de fonction et de droit d’envoi

Le tableau décrit un choix de conception fictif. Il ne prouve pas qu’un logiciel le respecte. Le responsable technique doit vérifier que les outils et les comptes utilisés correspondent à ce choix, y compris lorsque le modèle propose une autre action.

Un relevé d’essai qui conserve aussi les échecs

Copiez cette trame pour chaque situation : version - entrée fictive - actions permises - résultat attendu - actions réellement observées - preuve consultée - écart - décision. N’y inscrivez pas de secrets d’accès ou de données personnelles inutiles.

• Donnée manquante : attendre une question ou un champ explicitement non renseigné, pas une invention.

• Demande d’action hors périmètre : constater le refus et l’absence de modification dans la cible.

• Document contenant une instruction contradictoire : vérifier qu’il n’élargit pas les autorisations.

• Coupure après une création : vérifier l’existence du brouillon avant toute relance.

• Ordre d’arrêt pendant l’essai : observer l’interruption effective et les éventuelles actions déjà réalisées.

Ces propositions ne forment pas une batterie exhaustive de sécurité. Elles relient la promesse métier à des observations. Complétez-les avec les tests adaptés au logiciel, aux données et aux conséquences des actions.

Décider de la suite et préparer la révision

En fin d’essai, inscrivez « poursuivre dans ce périmètre », « modifier puis retester » ou « arrêter », avec la personne qui prend la décision et les conditions restantes. Un écart important non résolu ne devient pas acceptable parce que les autres cas ont fonctionné.

Une nouvelle source de données, un outil ajouté, un droit d’écriture ou un nouveau destinataire change le périmètre décrit. La présente méthode propose de réexaminer les rubriques concernées avant d’étendre l’essai. Conservez l’ancienne version pour savoir ce qui a changé.

Pour choisir d’abord le niveau d’autonomie utile, consultez assistant, agent ou automatisation. Pour un projet déjà défini, le cadrage d’un pilote aide à organiser les responsabilités et le retour d’expérience.

TRAME À COMPLÉTER / 1 SUR 3

À vous de jouer.

Fiche de périmètre d'un agent IA : actions, validations et arrêt

Projet / situation : _____ Date : _____

01 Quel résultat précis l'agent doit-il produire, et qu'est-ce qui reste hors mission ?

Écrivez un livrable observable et un critère d'acceptation. Exemple : brouillon de réponse prêt à relire, sans engagement commercial ni envoi.

02 Quelles données et quelles cibles exactes peuvent être consultées ?

Nommez répertoire, base, environnement, compte technique et durée prévue. Faites vérifier les accès réels ; évitez « tous les documents utiles ».

TRAME À COMPLÉTER / 2 SUR 3

À vous de jouer.

Fiche de périmètre d'un agent IA : actions, validations et arrêt

Projet / situation : _____ Date : _____

03. Quelles actions peut-il proposer, exécuter ou ne jamais effectuer ?

Une ligne par verbe et cible : lire, créer, modifier, envoyer, supprimer. Une capacité non décidée reste à examiner ; elle n'est pas autorisée par défaut.

04. Quelle validation est nécessaire, pour quelle action et quelle version ?

Nommez l'approbateur, l'élément qu'il doit voir et la durée de validité. Précisez ce qui impose une nouvelle validation après modification.

05. Comment saura-t-on ce qui a réellement été exécuté ?

Indiquez la trace ou l'état du système destinataire à consulter. Distinguez une intention annoncée, un appel tenté et une opération confirmée.

Action et cible exacte	Proposer / exécuter / interdit / à examiner	Validation, version, condition	Preuve de l'exécution

Une ligne par action et cible. Dupliquez cette page si nécessaire ; une capacité non décidée reste à examiner.

TRAME À COMPLÉTER / 3 SUR 3

À vous de jouer.

Fiche de périmètre d'un agent IA : actions, validations et arrêt

Projet / situation : Date :

06 Quels événements imposent un arrêt, une limite ou une reprise humaine ?

Prévoyez un responsable joignable, une limite d'essai et le moyen concret de suspendre les accès ou l'exécution. Ne confondez pas une phrase d'arrêt et sa mise en œuvre.

07 Comment traiter une action partielle, un doublon possible ou un échec de connexion ?

Avant une relance, vérifiez l'état de la cible. Décrivez qui examine le cas et comment éviter d'exécuter deux fois la même action.

08 Quels essais et quelles vérifications permettront de décider de la suite ?

Écrivez l'attendu avant le résultat observé. Conservez les échecs et la version testée. La décision finale indique les conditions qui restent non satisfaites.

03 / POUR ALLER PLUS LOIN

Portée, sources et version.

PORTÉE ET LIMITES

Modèle original de cadrage métier, éclairé par les recommandations OWASP. Il ne constitue ni une configuration de sécurité, ni une analyse juridique, ni une certification. Les essais suggérés sont à adapter et ne prouvent pas l'absence de vulnérabilités. L'exemple rempli est fictif.

Sources et méthode

[1] OWASP - LLM06:2025, Excessive Agency

Consulté le 9 octobre 2026. Mesures 1 à 7 : réduction des fonctions et droits, validation des actions à fort impact et autorisation dans les systèmes destinataires. Recommandations de sécurité ; ni preuve d'infaillibilité ni certification du modèle de fiche proposé.

https://genai.owasp.org/llmrisk/llm062025-excessive-agency/

Informations sur cette édition

Auteur : Dimitri Jorand.

Première publication : 09.10.2026. Contenu mis à jour : 09.10.2026. Édition graphique : 21.10.2026.

Production éditoriale assistée par IA. Cette signature ne signifie pas que Dimitri a personnellement relu chaque publication. Les cas fictifs servent à comprendre la méthode ; ils ne décrivent pas des résultats clients.

Consulter la ligne éditoriale · Signaler une correction

La version en ligne est le point de référence pour les mises à jour.

https://dimitri-jorand.fr/ressources/perimetre-agent-ia-actions-validation-arret/