POKÉDEX DES IA



QUI FAIT AVANCER LE JEU ?

Audit des résultats, des comportements et de la fiabilité des preuves dans le projet De l’écran à l’abandon.

LE VERDICT

Les meilleurs apports identifiables à la livraison V1.2 viennent ici de DeepSeek et GLM, avec adaptation par l’intégrateur. Les alertes les plus fortes concernent les annonces non étayées et les empreintes invalides, notamment Qwen3 Next et Devstral.

UNE ENCYCLOPÉDIE DE TERRAIN

14 fiches principales, dont Codex évalué séparément. Top/flop qualitatif fondé sur les traces du projet, sans score inventé, sans appel aux modèles et sans benchmark payant.

Style encyclopédie de créatures / emblèmes IA originaux dessinés localement. Ces symboles sont des identifiants graphiques non officiels, pas les logos des fournisseurs. Aucun visuel Pokémon reproduit.

Périmètre & méthode

CORPUS ANALYSÉ

8 relais archivés, 51 passes de 13 profils MATRIX dans recent-agents.json, plus 50 checkpoints outils : 457 résultats et 42 événements d'écriture déclarés par les outils. Les fichiers de checkpoint ont des dates de modification du 29 septembre au 4 octobre 2026 ; ces dates ne garantissent pas la date exacte des actions.

TROIS NIVEAUX DE PREUVE

1. Texte d'agent : ce qu'il annonce.
2. Résultat d'outil : lecture, écriture ou erreur enregistrée.
3. Intégration livrée : attribution dans le rapport, présence dans la chaîne active, test du pack exact.

Seul le troisième niveau établit un apport livré.

RÈGLE DE CLASSEMENT

Le top privilégie l'utilité intégrée. Le flop vise les preuves invalides et la confiance excessive, pas les erreurs réseau. Pas de total de points : missions, modèles, accès et périodes sont différents. Les checkpoints et les relais sont deux corpus distincts ; on ne calcule pas un taux de réussite commun.

CE QUI RESTE NON MESURÉ

Pas de coût par modèle, de latence comparable, de qualité générale, de temps économisé ou de taux de bugs par patch. Les écritures répétées ne sont pas dédupliquées. Les erreurs outils ne sont pas toutes causées par le modèle. Les fichiers écrits n'ont pas tous été relus, compilés ou rapprochés par empreinte.

Aucun appel IA externe, aucune nouvelle équipe, aucun téléchargement. Analyse locale en lecture seule pour le jeu ; seule la création de ce dossier écrit dans le workspace documentaire.

Top / flop de production

TOP 1 / DEEPSEEK

Consommables et filtre commerces explicitement intégrés au pack V1.2, après raccord et correction. Le meilleur signal est l'apport utile livré.

TOP 2 / GLM

Catalogue de données intégré après adaptation au contrat réel du jeu. Un apport concret, mais pas une livraison autonome.

HORS CLASSEMENT / CODEX

Intégration et preuves de livraison documentées. Auto-évaluation et absence de mission comparable : aucune médaille de « meilleur modèle ».

FLOP 1 / QWEN3 NEXT

Annonces de fichiers vérifiés et empreintes SHA256 invalides. Activité outil ultérieure élevée, sans démonstration d'intégration au pack courant.

FLOP 2 / DEVSTRAL

Même signal d'empreintes trop courtes, statuts assurés sans exécution démontrée dans les passes correspondantes.

AUTRES ALERTES / MISTRAL, MINIMAX

Discours d'exécution ou de fonctionnement trop confiant alors que le bilan d'outils indique aucune action sur ces relais.

Sonnet et Nemotron : problèmes d'accès ou de format séparés de la compétence. Haiku et Kimi : signaux utiles, mais résultats ultérieurs à vérifier par lot.

Tableau des traces MATRIX

PROFIL	RELAIS	SANS OUTIL*	OUTILS	ÉCRIT.
DeepSeek V3.2	3	2	35	9
GLM 4.7	6	3	49	6
Kimi K3	7	6	38	10
Claude Sonnet 4.6	5	0	67	4
Claude Haiku 4.5	5	2	14	3
Mistral Large 3	3	3	9	3
Qwen Coder	3	1	34	3
MiniMax	3	3	16	3
GPT-OSS 120B	4	1	37	0
Devstral 2 123B	3	2	11	0
Nova 2 Lite	3	2	8	0
Nemotron 3 Super	3	1	39	0
Qwen3 Next 80B	3	2	100	1

* Nombre de passes contenant le bilan « aucun outil exécuté », dans les relais archivés. Les colonnes OUTILS et ÉCRIT. proviennent des checkpoints ultérieurs : elles ne corrigent pas rétroactivement le bilan d'un ancien relais.

Une écriture est un événement, pas un fichier unique ni un changement validé. Une lecture ou une revue utile peut produire zéro écriture. Les erreurs d'accès sont détaillées dans les fiches.

FICHE #001 / INTÉGRATION

Codex / K3

Référence de livraison - hors classement MATRIX



Profil d’intégration évalué par rapports et vérification actuelle des empreintes. Aucune statistique MATRIX comparable.

COMPORTEMENT OBSERVÉ

Intégration, corrections et conservation des versions. Les rapports de récupération puis de V1.2 apportent des preuves de packs réels. Aujourd’hui, EXE et PCK ont été recalculés : empreintes conformes au registre.

ATOUT ET PREUVE

Capacité démontrée à transformer plusieurs contributions en un pack vérifié. Les 92 contrôles sont ceux du 30 septembre, pas des tests rejoués pendant cet audit.

FAIBLESSE / VIGILANCE

Évaluation de soi : conflit d’intérêt explicite. Pas de mission identique ni de journal MATRIX comparable ; aucun score de supériorité générale. Le nom V1 OFFICIELLE a été enregistré sans modifier les libellés du pack.

MISSION À LUI CONFIER

Intégrateur final, tests du pack et tenue du registre. Exiger une capture réelle et un journal pour chaque prochaine livraison.

SOURCES : S1, S2, S4

Observation du projet / aucune conclusion sur une personnalité humaine.

DeepSeek V3.2

TOP 1 - contribution livrée identifiable



Relais : 3 passes / 1 erreurs d'appel ou blocages / 2 bilans sans outils.
Checkpoints : 4 journaux / 35 résultats / 9 écritures / 2 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Le rapport V1.2 attribue à DeepSeek ConsumablesAdvanced et le filtrage des commerces, raccordés à l'inventaire et au magasin.

ATOUT ET PREUVE

Deux fonctions utiles ont une provenance documentée dans le pack final. Le test confirme consommation unique, effets et filtrage.

FAIBLESSE / VIGILANCE

Dans les anciens relais, deux réponses portent le bilan « aucun outil exécuté » ; une tentative termine en erreur 500. Le bon résultat intégré ne rend pas chaque réponse fiable. Les correctifs de l'intégrateur comptent.

MISSION À LUI CONFIER

Petits modules de gameplay avec contrat précis, puis revue Godot. Éviter les demandes « tout intégrer » et vérifier le raccord au monde.

SOURCES : S2, S3, S4, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

GLM 4.7

TOP 2 - contribution livrée identifiable



Relais : 6 passes / 1 erreurs d'appel ou blocages / 3 bilans sans outils.
Checkpoints : 4 journaux / 49 résultats / 6 écritures / 14 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Le catalogue JSON de GLM a été adapté au format effectivement attendu pour la livraison V1.2.

ATOUT ET PREUVE

Apport concret aux données commerciales ; journaux ultérieurs montrant six événements d'écriture sur quatre checkpoints.

FAIBLESSE / VIGILANCE

Une synthèse très précise des trois modes n'est pas accompagnée d'outils sur cette passe. Les journaux comptent 14 résultats avec erreur. Le catalogue a nécessité une adaptation : ce n'est pas une livraison autonome.

MISSION À LUI CONFIER

Catalogue, schéma JSON et contrôles de cohérence. Faire valider le schéma avant d'augmenter le volume.

SOURCES : S2, S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

Kimi K3

ACTIF - utilité à confirmer par lot



Relais : 7 passes / 1 erreurs d'appel ou blocages / 6 bilans sans outils.
Checkpoints : 4 journaux / 38 résultats / 10 écritures / 2 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Un extrait remet en question les fichiers annoncés par Mistral et cherche à contrôler la chaîne de lancement. Les checkpoints contiennent dix événements d'écriture.

ATOUT ET PREUVE

Bonne intention de vérifier les annonces des autres et attention aux dépendances du gameplay.

FAIBLESSE / VIGILANCE

Six des sept passes du corpus de relais ont le bilan « aucun outil exécuté ». Une passe est arrêtée pour répétition. EffectsManager n'est pas activé dans V1.2 : dépendances absentes et effet de test au démarrage ; l'auteur précis de chaque défaut n'est pas établi ici.

MISSION À LUI CONFIER

Un seul bug, un seul fichier central, une preuve Godot. Arrêter la boucle de relais si la seconde réponse ne produit aucune nouvelle preuve.

SOURCES : S2, S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

Claude Sonnet 4.6

ACCÈS INSTABLE - hors flop technique



Relais : 5 passes / 5 erreurs d’appel ou blocages / 0 bilans sans outils.
Checkpoints : 5 journaux / 67 résultats / 4 écritures / 22 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Les cinq tentatives du corpus de relais échouent en code 429. D’autres checkpoints montrent 67 résultats outils et quatre écritures.

ATOUT ET PREUVE

Les journaux ultérieurs prouvent que le profil n’a pas été uniquement bloqué sur toute la période.

FAIBLESSE / VIGILANCE

Le code 429 concerne l’accès ou le quota : il ne mesure pas la compétence. Les 22 erreurs outils ultérieures peuvent inclure paramètres, chemins et restrictions ; aucune causalité unique ni comparaison équitable établie.

MISSION À LUI CONFIER

Revue de patch sur données accessibles. Vérifier d’abord disponibilité ; pas de repli facturé ni nouvel appel lancé par cet audit.

SOURCES : S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

Claude Haiku 4.5

SIGNAL POSITIF - non livré confirmé



Relais : 5 passes / 1 erreurs d'appel ou blocages / 2 bilans sans outils.
Checkpoints : 3 journaux / 14 résultats / 3 écritures / 1 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Une réponse distingue explicitement les déclarations sans outils et les erreurs des autres profils. Les checkpoints contiennent trois événements d'écriture, dont données et gestionnaire de gare.

ATOUT ET PREUVE

Lecture critique du relais et volonté de limiter le discours aux preuves.

FAIBLESSE / VIGILANCE

Son propre discours très affirmatif reste parfois accompagné d'un bilan sans outils. Aucun test du pack courant ne valide ici ses nouveaux fichiers de gare. Une erreur 500 existe dans le corpus.

MISSION À LUI CONFIER

Contrôle bref des chemins, schémas et critères d'acceptation. Bon candidat pour une revue limitée, avec résultat vérifiable.

SOURCES : S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

FICHE #007 / COORDINATION

Mistral Large 3

FLOP - exécution racontée dans les relais



Relais : 3 passes / 0 erreurs d'appel ou blocages / 3 bilans sans outils.
Checkpoints : 4 journaux / 9 résultats / 3 écritures / 2 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Les trois passes du corpus sont sans outils exécutés. Des chaînes ressemblant à des appels read_file apparaissent dans le texte.

ATOUT ET PREUVE

Structure les lots et la répartition des tâches avec clarté. Les checkpoints ultérieurs enregistrent aussi trois écritures dans MTP_ECU.

FAIBLESSE / VIGILANCE

Du texte ressemblant à un appel ne constitue pas un appel. Des plans d'intégration et téléchargements sont racontés sans preuve sur ces passes. Les écritures ultérieures ne sont pas automatiquement dans la source courante du jeu.

MISSION À LUI CONFIER

Briefs et coordination seulement jusqu'à démonstration d'une tâche exécutée. Un plan doit se terminer par un livrable contrôlé, pas un nouveau relais.

SOURCES : S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

Qwen Coder

À SURVEILLER - compatibilité Godot



Relais : 3 passes / 1 erreurs d'appel ou blocages / 1 bilans sans outils.
Checkpoints : 4 journaux / 34 résultats / 3 écritures / 11 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Certaines passes reconnaissent qu'aucun fichier n'a été modifié. La reprise historique signale des propositions Godot incompatibles et des scènes ou packs annoncés absents.

ATOUT ET PREUVE

Reconnaître une passe sans modification est préférable à attribuer une livraison inexistante. Les checkpoints contiennent trois écritures.

FAIBLESSE / VIGILANCE

Timeout dans un relais ; vérifications prétendues parfois sans outils. Les onze erreurs outils ultérieures ne prouvent pas toutes un défaut de code. Le défaut historique est documenté, mais n'est pas un benchmark du modèle actuel.

MISSION À LUI CONFIER

Patch isolé dans une scène de test ; parser avec le moteur du projet avant de confier un fichier de production.

SOURCES : S3, S5, S6

Observation du projet / aucune conclusion sur une personnalité humaine.

MiniMax

FLOP - confiance non étayée



Relais : 3 passes / 0 erreurs d'appel ou blocages / 3 bilans sans outils.
Checkpoints : 5 journaux / 16 résultats / 3 écritures / 5 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Les trois réponses des relais portent le bilan sans outils, alors qu'elles parlent de modes implémentés et d'intégration d'items.

ATOUT ET PREUVE

Peut organiser un retour d'expérience ; trois écritures sont enregistrées plus tard sur accident_sequence.gd.

FAIBLESSE / VIGILANCE

Présence de fichiers et fonctionnement sont confondus dans les annonces. Les trois écritures d'un même fichier ne valent pas trois fonctionnalités, ni validation de la séquence complète de secours.

MISSION À LUI CONFIER

Retour UX à partir de captures réellement fournies ; pour le code, exiger patch et test d'un scénario de secours précis.

SOURCES : S2, S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

GPT-OSS 120B

RÉSERVE - aucune écriture observée



Relais : 4 passes / 1 erreurs d'appel ou blocages / 1 bilans sans outils.
Checkpoints : 4 journaux / 37 résultats / 0 écritures / 6 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Une réponse annonce plusieurs éléments « opérationnels » et une mise à jour concrète ; son état ne démontre pas un test. Les checkpoints contiennent 37 résultats, zéro écriture enregistrée.

ATOUT ET PREUVE

Des critères de preuve sont proposés dans le plan de travail. Une revue peut être utile sans écrire de fichier.

FAIBLESSE / VIGILANCE

Les statuts assurés sur les modes dépassent les preuves de la réponse. Zéro écriture n'est pas un échec pour un rôle de revue ; il interdit seulement de lui créditer un patch sur cet échantillon.

MISSION À LUI CONFIER

Contre-relecture : demander un défaut reproductible ou une absence de défaut dans un périmètre clairement délimité.

SOURCES : S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

Devstral 2 123B

FLOP 2 - empreintes invalides



Relais : 3 passes / 1 erreurs d'appel ou blocages / 2 bilans sans outils.
Checkpoints : 4 journaux / 11 résultats / 0 écritures / 3 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Une réponse donne pour SHA256 deux chaînes de 32 caractères et annonce des données vérifiées. Deux des trois passes portent le bilan sans outils.

ATOUT ET PREUVE

Présente des plans courts et lisibles.

FAIBLESSE / VIGILANCE

SHA256 s'écrit sur 64 caractères hexadécimaux : les chaînes affichées ne peuvent pas être les empreintes annoncées. Les checkpoints ont zéro écriture et trois erreurs outils ; cela ne mesure pas la qualité d'un patch non observé.

MISSION À LUI CONFIER

Revue de texte ou petite correction après preuve d'accès. Rejeter automatiquement toute empreinte mal formée dans un rapport de livraison.

SOURCES : S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

Nova 2 Lite

NON CLASSÉ - production insuffisante



Relais : 3 passes / 1 erreurs d'appel ou blocages / 2 bilans sans outils.
Checkpoints : 3 journaux / 8 résultats / 0 écritures / 1 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Trois passes de relais, dont une erreur 500 et deux bilans sans outils. Trois checkpoints, huit résultats et aucune écriture.

ATOUT ET PREUVE

Décompose achats, stockage et effets en lots.

FAIBLESSE / VIGILANCE

Les descriptions d'assets et systèmes ne sont pas des preuves de leur intégration. Une erreur d'outil « Pages hors du document » est observée dans un checkpoint : contrôler le nombre de pages avant lecture.

MISSION À LUI CONFIER

Découpage court et inventaire de dépendances avec chemins confirmés. Aucun crédit de livraison sans preuve distincte.

SOURCES : S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

Nemotron 3 Super

BLOCAGE ROUTE - hors flop compétence



Relais : 3 passes / 2 erreurs d'appel ou blocages / 1 bilans sans outils.
Checkpoints : 2 journaux / 39 résultats / 0 écritures / 3 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Deux des trois tentatives du corpus sont refusées en code 400 pour paramètres ou format incompatibles. Les checkpoints suivants contiennent 39 résultats, zéro écriture.

ATOUT ET PREUVE

Le texte propose un protocole achat, stockage, consommation, effets et limites.

FAIBLESSE / VIGILANCE

Un code 400 ne démontre ni manque de crédits ni faiblesse intellectuelle. Un protocole QA ne démontre pas que les tests ont été exécutés.

MISSION À LUI CONFIER

QA en lecture seule si la route fonctionne, sur un scénario unique et un journal fourni. Réparer la configuration avant de juger les résultats.

SOURCES : S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

Qwen3 Next 80B

FLOP 1 - preuve de livraison invalide



Relais : 3 passes / 1 erreurs d'appel ou blocages / 2 bilans sans outils.
Checkpoints : 4 journaux / 100 résultats / 1 écritures / 18 résultats avec erreur.

COMPORTEMENT OBSERVÉ

Des réponses affirment « tous vérifiés via outils » et donnent des SHA256 de 32 caractères. Les relais correspondants portent le bilan sans outils.

ATOUT ET PREUVE

Cent résultats outils figurent dans quatre checkpoints ultérieurs : activité réelle, mais non équivalente à une livraison.

FAIBLESSE / VIGILANCE

C'est le signal le plus net de preuve mal présentée dans le corpus. La seule écriture enregistrée dans ces checkpoints vise un chemin C: ; les règles courantes du projet demandent D:. Le contexte exact de l'autorisation à cette date n'est pas reconstitué.

MISSION À LUI CONFIER

Ne pas accepter ses bilans de livraison sans recalcul local. Périmètre en lecture seule jusqu'à un patch correctement attribué et testé.

SOURCES : S3, S5

Observation du projet / aucune conclusion sur une personnalité humaine.

Les pièges du collectif

LE THÉÂTRE DES OUTILS

Écrire « je lis » ou imprimer read_file dans une réponse ne prouve pas une lecture. Le bilan d'exécution doit être joint à chaque action annoncée.

LA PROPAGATION DES CERTITUDES

Un agent reprend les tableaux du précédent et change « présent » en « opérationnel ». Le relais amplifie ainsi une hypothèse. Revenir au fichier et au test avant tout statut.

LA CONFUSION DES VERSIONS

Les checkpoints visent V1, V1.1, V1.2, MTP_ECU ou C:. Une réussite ailleurs ne prouve pas un apport à notre pack courant. L'intégrateur doit contrôler la source et le dossier de sortie.

LE VOLUME SANS RÉSULTAT

Plus de résultats outils, de lignes ou de relais ne signifie pas meilleur apport. Qwen3 Next possède le plus de résultats dans ce corpus, sans contribution livrée établie par cet audit.

LA RESPONSABILITÉ PARTAGÉE

Consignes longues, rôles différents, formats d'appels et restrictions d'outils influencent le résultat. On constate des écarts entre annonces et preuves ; on ne déduit pas une intention de tromper.

AUDIT DE PRODUCTION / 04 OCTOBRE 2026

Plan pour notre V1 OFFICIELLE

1 / GARDER UN INTÉGRATEUR

Un seul responsable accepte les changements dans la source d’intégration, conserve les versions et teste le pack exact. Les contributeurs produisent des lots séparés.

2 / DONNER UNE MISSION COURTE

Objectif, chemins autorisés, un comportement à corriger, critère de validation et livrable. Pas de demande globale à quatorze profils pour la même modification.

3 / FERMER LE LOT AVEC UNE PREUVE

Patch réel, journal Godot pertinent, capture si visuel. Un rapport « tout fonctionne » sans test est renvoyé au statut non vérifié.

4 / STOPPER LES BOUCLES

Deux réponses répétées sans nouvelle preuve : arrêt du relais et inspection du blocage. Aucun appel automatique payant pour contourner une erreur.

5 / PRIORISER LA SORTIE

Vérifier lancement, sauvegarde/reprise, achats/consommation, trams et secours ; puis récit jouable, finition visuelle et performance. Les 92 contrôles existants ne certifient pas tout le jeu.

V1 OFFICIELLE est le nom choisi le 4 octobre. Le pack technique reste V1.2-20260930. La décision de nom ne remplace ni test de la copie de travail actuelle ni validation visuelle finale de Mehdi.

Profils sans verdict comparable

GEMINI

Un journal local examiné indique status SUCCESS avec steps vide. Ce statut n’est pas une preuve d’asset ou de modification. Les autres missions Gemini n’ont pas été auditées intégralement : profil non classé, aucune généralisation.

GITHUB ARCHITECT / GITHUB FAST

Ces alias figurent dans les candidats d’un relais. Leur présence dans une liste ne prouve pas une participation ni un résultat. Hors classement faute de corpus de passes comparable dans cet audit.

GPT-6 ASTRA ET AUTRES PROFILS

La fiche TEAM_IA_INSTRUCTIONS historique mentionne un refus 403 pour Astra ; cela n’établit pas l’accès actuel. Pas de test ni classement aujourd’hui. Toute IA absente du tableau reste non évaluée.

ATTRIBUTION AUDIO

Le rapport V1.2 cite un synthétiseur audio de la copie commune, corrigé et raccordé. Son auteur n’est pas identifié dans ce passage : aucun agent ne reçoit ici un crédit audio arbitraire.

LIMITE DE COMPLÉTUDE

Ce document couvre les preuves examinées et les principaux profils identifiés. Il ne prétend pas contenir tous les chats, toutes les missions ni tous les modèles du compte. « Non vérifié » signifie manque de preuve, pas échec démontré.

Sources & reproductibilité

S1 / D:/DELA-JEU/VERSION_COURANTE.json
Version officielle, pack technique et limites. Empreintes EXE/PCK recalculées pendant la reprise du 4 octobre.

S2 / D:/DELA-JEU/livraison-v1.2-20260930/RAPPORT-LIVRAISON.md
Attribution des contributions intégrées, corrections et limites de la livraison.

S3 / D:/DELA-JEU/production/v1.2-20260930/recent-agents.json
8 relais ; passes, réponses et erreurs. Source primaire des observations de discours.

S4 / D:/DELA-JEU/production/v1.2-20260930/final-pack-test.log + final-gameplay.log
66 + 26 contrôles archivés, zéro échec annoncé par ces journaux. Pas de réexécution dans cet audit.

S5 / D:/MTP-CENTRALE/conversations-aws/*.progress.json
50 checkpoints lus ; agrégation des champs model et results. Une entrée written compte comme événement d'écriture, sans preuve de validation.

S6 / D:/DELA-JEU/jeu-montpellier-arcade/00_REPRISE_K3.md + docs/TEAM_IA_INSTRUCTIONS.md
Historique des défauts Qwen et règles de séparation proposition, outil et livraison.

S7 / D:/MTP-CENTRALE/logs/gemini-gemini-mission-c32ca005622b4ac995061b505ed5b0ab.json
Exemple de statut SUCCESS sans étapes ; examen limité à ce journal.

Audit préparé le 4 octobre 2026. Les fichiers ont pu être modifiés après leur lecture. Les noms des profils suivent les alias enregistrés ; les versions commerciales ne sont pas revalidées en ligne.

Fin de campagne : les prix

AGENT LE PLUS PERFORMANT / DEEPSEEK

Verdict limité à l'utilité intégrée dans V1.2 : consommables et filtrage commercial documentés. GLM suit pour le catalogue. Aucune victoire démontrée sur une mission commune chronométrée.

AGENT LE PLUS FIABLE / PAS DE VAINQUEUR ÉTABLI

La chaîne Codex + registre + tests est la référence de livraison vérifiable. Parmi les profils MATRIX, les missions sont trop différentes et les bilans sans outils trop fréquents pour attribuer honnêtement ce titre à un seul modèle.

MEILLEUR CODEUR / DEEPSEEK SUR LE LOT LIVRÉ

Candidat retenu pour ces modules de gameplay, avec adaptation de l'intégrateur. Kimi montre des écritures actuellement identiques sur bike.gd et city_interactables.gd dans V1.2 ; leur fonctionnement reste à tester.

MEILLEUR AUDITEUR / VERDICT CONDITIONNEL

Codex a rapproché journaux, fichiers et empreintes pour ce dossier ; auto-évaluation signalée. Haiku montre une lecture critique du relais, sans preuve suffisante pour le déclarer meilleur auditeur MATRIX.

ANNONCES SANS PREUVE / CAS OBSERVÉS

Kimi, GLM, Mistral, Qwen Coder, MiniMax, GPT-OSS, Haiku, Devstral, Nova, DeepSeek et Qwen3 Next ont au moins une réponse annonçant lectures, état opérationnel ou travail concret avec bilan sans outils. Les fiches précisent les exemples. Une absence de preuve n'établit pas une intention de tromper.

Preuves dans les fichiers

CONTRÔLE SUPPLÉMENTAIRE EFFECTUÉ

34 couples agent/fichier issus des événements written ont été rapprochés des fichiers locaux : 34 présents, 28 avec au moins une empreinte outil identique au contenu actuel. Les six autres ont été modifiés depuis ou ne possèdent pas une empreinte concordante. Ce sont des couples : un fichier partagé peut apparaître plusieurs fois.

CE QUE CELA DÉMONTRE

Une écriture est enregistrée dans le checkpoint de cet agent et le contenu actuel correspond à une empreinte enregistrée. Ce contrôle ne prouve pas l'auteur exclusif, la pertinence de la modification, son import Godot ni sa présence dans le pack livré. Les checkpoints peuvent être anciens ou non rattachables à la dernière réponse.

FICHIERS CENTRAUX CONFIRMÉS

Kimi : travail-matrix-v1.2/world3d/transport/bike.gd et city_interactables.gd.
Sonnet : travail-matrix-v1.2/world3d/activities/commando_fps.gd et interaction_effects.gd ; source d'intégration/world3d/hunger_system.gd.
GLM : travail-matrix-v1.2/world3d/city.gd.
DeepSeek : travail-matrix-v1.1/world3d/commerce_filter_system.gd.
MiniMax : travail-matrix-v1.1/world3d/accident_sequence.gd.

À NE PAS ATTRIBUER AU DERNIER AGENT SANS REVUE

effects_manager.gd a plusieurs auteurs/checkpoints ; hud.gd et city_interactables.gd ont aussi été réécrits. Le dernier contenu identique ne permet pas de répartir toutes les lignes entre les intervenants. Le registre livré reste la référence.

Le fichier technique verified-writes.json conserve chemins complets, checkpoints et empreintes recalculées. Les deux pages suivantes listent les 34 couples observés ; aucune suppression ni modification de ces fichiers par l'audit.

Registre des écritures 1

Qwen Coder / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.1\backup_consumption_feedback_20261002\CHANGEMENT_A_APPLIQUER.txt

Qwen Coder / 1 événement(s) / empreinte non concordante ou absente
D:\DELA-JEU\travail-matrix-v1.1\world3d\effects_manager.gd

Qwen Coder / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.1\world3d\hud.gd

DeepSeek V3.2 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1\docs\AMELIORATION_21_MISSIONS_TESTS_MEHDI.md

DeepSeek V3.2 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1\scripts\update_all_dependencies.gd

DeepSeek V3.2 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1\scripts\ensure_mission_launcher.gd

DeepSeek V3.2 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1\docs\YOUR_TURN_MEHDI_21_MISSIONS_TEST.md

Qwen3 Next 80B / 1 événement(s) / empreinte actuelle identique
C:\Users\msoui\Downloads\MATRIX-Documents\quest_coffre_souvenir.json

Claude Sonnet 4.6 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.2\world3d\activities\commando_fps.gd

Claude Sonnet 4.6 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.2\world3d\interaction_effects.gd

Claude Sonnet 4.6 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.1\world3d\effects_manager.gd

Claude Haiku 4.5 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.1\world3d\data\gare_saint_roch.json

Claude Haiku 4.5 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.1\world3d\gare_saint_roch_manager.gd

Claude Haiku 4.5 / 1 événement(s) / empreinte actuelle identique
D:\MTP-CENTRALE\documents-partages\GARE_TRAM_INTEGRATION_FINALE.md

DeepSeek V3.2 / 1 événement(s) / empreinte non concordante ou absente
D:\DELA-JEU\travail-matrix-v1.1\world3d\consumables_advanced.gd

DeepSeek V3.2 / 1 événement(s) / empreinte non concordante ou absente
D:\DELA-JEU\travail-matrix-v1.1\world3d\effects_manager.gd

DeepSeek V3.2 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.1\world3d\commerce_filter_system.gd

Registre des écritures 2

DeepSeek V3.2 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.1\scripts\test_system_consommables.py

DeepSeek V3.2 / 1 événement(s) / empreinte actuelle identique
D:\MTP-CENTRALE\documents-partages\GUIDE_TEST_CONSOMABLES.md

Kimi K3 / 1 événement(s) / empreinte non concordante ou absente
D:\DELA-JEU\travail-matrix-v1.1\world3d\effects_manager.gd

Kimi K3 / 2 événement(s) / empreinte non concordante ou absente
D:\DELA-JEU\travail-matrix-v1.1\world3d\hud.gd

GLM 4.7 / 1 événement(s) / empreinte non concordante ou absente
D:\DELA-JEU\travail-matrix-v1.2\world3d\city_interactables.gd

GLM 4.7 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.2\world3d\city.gd

Kimi K3 / 4 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.2\world3d\transport\bike.gd

Kimi K3 / 3 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.2\world3d\city_interactables.gd

GLM 4.7 / 1 événement(s) / empreinte actuelle identique
D:\MTP-CENTRALE\documents-partages\SYNTHESE_21_MISSIONS_MODE_HISTOIRE_CARTES_OU_D_INTERET.md

GLM 4.7 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1\scripts\generate_21_missions_from_docs.gd

GLM 4.7 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1\data\map_markers_18_poi.json

Mistral Large 3 / 1 événement(s) / empreinte actuelle identique
D:\MTP_ECU\publication-github\LANCER_MODE_HISTOIRE.bat

Mistral Large 3 / 1 événement(s) / empreinte actuelle identique
D:\MTP_ECU\publication-github\histoire\index.html

Mistral Large 3 / 1 événement(s) / empreinte actuelle identique
D:\MTP_ECU\publication-github\CREER_RACCOURCI.ps1

GLM 4.7 / 1 événement(s) / empreinte actuelle identique
D:\MTP-CENTRALE\documents-partages\MINIJEUX_ET_DETENTE.md

Claude Sonnet 4.6 / 1 événement(s) / empreinte actuelle identique
D:\DELA-JEU\jeu-montpellier-arcade\world3d\hunger_system.gd

MiniMax / 3 événement(s) / empreinte actuelle identique
D:\DELA-JEU\travail-matrix-v1.1\world3d\accident_sequence.gd

Erreurs, coûts & travail restant

ERREURS RENCONTRÉES

429 : refus ou quota d’accès Sonnet dans les relais. 500 : erreurs de service sur plusieurs profils. 400 : format ou paramètres Nemotron. Timeout : Qwen Coder. Blocage de répétition : Kimi. Outils : pages hors document, erreurs de lecture/écriture non toutes classifiées. Production V1.2 : city.gd tronqué et syntaxe invalide non intégrés ; dépendances EffectsManager absentes ; code audio corrigé.

COÛT / CRÉDITS CONSOMMÉS

Campagne historique : montant inconnu. Aucun relevé de facturation ou décompte de tokens rapproché de ces appels n’a été analysé ; impossible de donner un total en euros ou crédits. Un code 400, 429 ou 500 ne prouve pas la gratuité ou un épuisement de crédits.
Cet audit : zéro invocation de modèle externe et aucune ressource payante lancée. Le coût de la session Codex elle-même dépend du compte ; il n’est pas mesuré ici.

À TERMINER MANUELLEMENT OU AVEC REVUE LOCALE

Mehdi : retour visuel et jouable, validation des faits et choix artistiques. Intégrateur : relire les changements de la copie V1.2, contrôler import et branchement, vérifier scène complète de secours, trams, sauvegarde et parcours. Recalculer les empreintes d’un nouveau pack après export, puis actualiser le registre. Récit joué, relief/finition, transports et performances restent ouverts.

Qui reste dans MATRIX ?

NOYAU PROPOSÉ

Codex pour intégration et livraison ; DeepSeek pour petits modules de gameplay ; GLM pour données et catalogues. Ce choix repose sur les apports identifiés au pack, avec contrôle systématique.

RÉSERVE AVEC CONTRAT COURT

Kimi pour bug Godot ciblé et transport ; Haiku pour contre-revue brève ; Sonnet pour architecture si accès disponible. Chacun doit fermer son lot avec une preuve, pas une simple synthèse.

RÔLES LIMITÉS

Mistral : coordination sans crédit automatique d'exécution. GPT-OSS : revue. Nova : planification. MiniMax : retour UX sur captures et scénarios transmis. Nemotron : protocole QA après correction de la route.

PROBATION / PAS DE DROIT DE LIVRAISON

Qwen3 Next et Devstral : rapports de livraison refusés jusqu'à une preuve de patch et empreinte valide. Qwen Coder : patch isolé et import Godot avant intégration. Une vraie amélioration contrôlée permet de réviser ce verdict.

DÉCISION DE CAMPAGNE

Réduire le noyau actif plutôt que faire tourner tous les profils à chaque lot. Aucun modèle n'est supprimé, réactivé ou appelé par cet audit. Cette sélection est une recommandation pour Mehdi, sous budget supplémentaire 0 EUR.

Le « légendaire » est celui dont le résultat survit à la revue et au test du pack. Une étiquette SUCCESS avec zéro étape reste une annonce à vérifier.