Évaluation des modèles pour WeeScribe V2

Mémo avec intégration de Gemini Argon
Test Marie 12 mn
3 novembre 2026

Recommandation : utiliser Claude Opus 5.5 en effort élevé comme choix courant. Il offre le meilleur compromis observé entre qualité, prudence, coût et durée. Gemini Argon devient l'option économique intermédiaire. Astra reste le choix privilégiant la finition rédactionnelle ; DeepSeek représente le coût minimal avec une relecture de fond.

Classement et paramètres de cet essai

Rang	Modèle	Coût IA	Durée	Usage recommandé
1	OpenAI Astra	0,693757 €	04:22	Priorité à la finition rédactionnelle
2	Claude Opus 5.5	0,342168 €	01:57	Choix courant recommandé
3	Gemini Argon	≈ 0,173104 €	02:38	Option économique intermédiaire
4	DeepSeek V4 Pro	≈ 0,106164 €	03:10	Coût minimal avec relecture de fond

Le rang concerne la qualité rédactionnelle observée, sans certifier la fidélité audio. Durées en minutes et secondes. Le signe ≈ désigne une estimation. Sources : S1 à S4, L1 à L4 et D1.

Objectif éditorial

Le niveau 3 Intervention écrite doit produire un français soutenu, sobre et naturel à lire. Les hésitations vides, les familiarités, les bégaiements et les répétitions inutiles doivent disparaître. Les phrases maladroites peuvent être reconstruites, en conservant les informations, les réserves, les conditions, les exemples et le déroulement du dialogue.

La priorité reste la fidélité : la qualité de rédaction ne doit ajouter ni raisonnement, ni intention, ni accusation, ni conclusion. Une phrase inachevée ne doit pas être complétée par une suite plausible. Les incertitudes doivent rester visibles lorsque la source ne permet pas de les résoudre.

Périmètre du test

Les quatre essais indiquent le moteur weescribe-editorial-v2.0, le niveau 3 Intervention écrite et un effort élevé. Les sources sélectionnées sont les transcriptions Grok xAI et Soniox ; les repères des interlocuteurs viennent de Soniox. Une seule sortie par modèle a été examinée. [L1 à L4]

Le coût Gemini, absent du journal initial, est désormais intégré à partir du complément fourni par l'utilisateur. Les quatre durées proviennent du même complément. Le mémo distingue les observations de cet essai des recommandations d'usage. [D1]

Analyse de la qualité des quatre résultats

Astra construit le texte le plus abouti

Astra organise mieux les paragraphes et intègre l'autocorrection sur les recettes françaises : « La France peut éventuellement récupérer des recettes grâce à la règle d'inclusion du revenu, mais pas si l'on applique le "side by side" avec les États-Unis. » La formulation est plus claire que la succession d'une affirmation absolue et d'un « Enfin si ». [S1, 05:17]

Des défauts subsistent : la recherche de notes du début est conservée ; « je me souviens, en montant » reste incompréhensible ; « on veut juste faire plaisir à Trump » devient « nous voulons seulement faire plaisir à Trump ». Ce changement de personne peut attribuer une intention au locuteur et à ses interlocuteurs. [S1, début, 05:17 et 04:33]

Claude offre le meilleur compromis observé

Claude commence directement par le sujet fiscal et balise le passage douteux de la Barbade au lieu de maintenir « en montant ». Cette prudence est utile avec une source dégradée. Elle ne constitue toutefois pas une certification de fidélité. [S2, début et 05:17]

Le nettoyage reste incomplet : « Enfin si », « En gros », certaines reprises orales et des paragraphes chargés subsistent. Au coût mesuré, Claude reste proche d'Astra sur la qualité pour environ la moitié du prix ; il est aussi le plus rapide dans les durées communiquées. [S2 ; L1 et L2 ; D1]

Gemini Argon constitue une option économique crédible

Gemini préserve les principales réserves, mais son premier paragraphe rassemble trop de sujets. Il laisse « en montant » et « j'en ai encore échangé ce matin avec ma collaboratrice », construction qui demanderait « j'en ai encore discuté » ou « j'ai encore échangé à ce sujet ». Il demeure moins abouti que Claude et Astra. [S3, début, 05:17 et 10:05]

Avec un coût estimé de 0,173104 € et un traitement de 2 min 38 s, il offre une alternative économique sérieuse. Pour ce besoin, je le privilégierais à DeepSeek si l'on veut réduire la facture tout en conservant une qualité rédactionnelle intermédiaire. Une relecture des passages défectueux reste nécessaire. [D1]

DeepSeek demande une relecture de fond

DeepSeek comporte « Que l'on comprenne qui peut », ferme trop tôt la citation attribuée aux États-Unis et transforme « le système qui doit être… » en « je suis parfaitement d'accord avec le système ». Il parle aussi des « pays » là où les autres résultats parlent de « la Commission ». Ces écarts concernent la portée et l'attribution des propos, au-delà de la grammaire. [S4, début, 03:15, 05:17 et 08:50]

Le coût estimé est le plus faible, mais le contrôle doit porter sur les noms, les chiffres, les réserves et les attributions. Je le retiendrais comme brouillon économique avec relecture, plutôt que comme compte rendu définitif automatique.

Coûts et durées avec Gemini Argon

Les montants comparables sont ceux de la réécriture seule. Le premier essai, Claude, comprend par ailleurs les transcriptions initiales ; les essais suivants les réutilisent sans nouvel appel. Comparer le total de Claude au seul coût IA des autres le pénaliserait artificiellement. [L1 à L4]

Modèle	Coût de cet extrait	100 extraits comparables
Claude Opus 5.5	0,342168 €	34,22 €
Gemini Argon	≈ 0,173104 €	≈ 17,31 €
DeepSeek V4 Pro	≈ 0,106164 €	≈ 10,62 €
OpenAI Astra	0,693757 €	69,38 €

Projection arithmétique à coût unitaire identique, hors transcription, serveur, taxes, frais de change et relecture humaine. Ce ne sont pas des prix garantis pour des enregistrements plus longs. [L1 à L4 ; D1]

Ce que change le prix désormais connu de Gemini

Claude coûte 1,98 fois Gemini ; Astra coûte 4,01 fois Gemini. Le supplément de Gemini par rapport à DeepSeek est d’environ 0,067 € par extrait, soit 6,69 € pour 100 extraits comparables. Pour la qualité observée, ce supplément paraît raisonnable si l’on cherche une option économique sans retenir le résultat le plus fragile.

Le choix courant reste Claude lorsque la qualité et le temps de reprise priment. Gemini devient le choix économique intermédiaire. Astra se justifie lorsque sa meilleure construction écrite compte davantage que le supplément de prix. Ces recommandations ne chiffrent pas le temps de correction humaine, qui n’a pas été mesuré.

Durées observées

Claude termine en 117 secondes, Gemini en 158 secondes, DeepSeek en 190 secondes et Astra en 262 secondes. Gemini est donc 41 secondes plus lent que Claude, mais 32 secondes plus rapide que DeepSeek. Astra a la durée la plus élevée dans cette série. [D1]

Ces durées décrivent les traitements communiqués. Elles ne constituent pas une mesure moyenne de rapidité des modèles et n’isolent pas nécessairement toutes les étapes techniques.

Limites à conserver avec le classement

Un seul extrait et une seule sortie par modèle ont été examinés. Il n’y a pas eu de nouvelle vérification auditive intégrale. L’accord de plusieurs résultats ne prouve pas l’exactitude d’un mot. Les prompts exacts envoyés ne figurent pas dans les journaux ; le niveau élevé ne garantit pas un effort équivalent entre fournisseurs, et le mode Pro d’Astra n’est pas confirmé.

Les 3 793 mots indiqués comme sources additionnent deux transcriptions du même audio. Ils ne représentent pas un contenu unique et ne permettent pas de calculer un taux d’omission. Les coûts consignés ne constituent pas un audit des factures des fournisseurs. [L1 à L4]

Consignes pour poursuivre le travail sur la V2

Conserver la V2 comme base du niveau 3 Intervention écrite et améliorer sa finition sans résumer le contenu. Les essais soutiennent le choix courant Claude, une option économique Gemini Argon, une option Astra privilégiant la rédaction et une option DeepSeek avec contrôle approfondi.

Contrôles éditoriaux à renforcer

1. Préserver les personnes et les acteurs. Ne pas transformer un « on » ambigu en « nous » ni remplacer une institution par une autre.

2. Maintenir les réserves, les exceptions et le degré de certitude. « De mémoire », « pour moi » et « éventuellement » doivent rester explicites lorsqu'ils portent du sens.

3. Vérifier les citations et les interruptions. Une ponctuation différente ne doit pas changer la voix rapportée ; une phrase coupée ne doit pas devenir un accord ou une conclusion complète.

4. Traiter les fragments incompréhensibles. Vérifier la source ou signaler le doute ; ne pas laisser une formulation sans sens ni la compléter par plausibilité.

5. Nettoyer la forme. Supprimer les recherches de notes et les reprises vides, intégrer les autocorrections explicites et organiser des paragraphes cohérents, tout en gardant les détails utiles.

Configuration consignée dans les essais

Tous les essais utilisent le niveau 3 Intervention écrite et un effort élevé. Identifiants indiqués : Claude claude-opus-5-5 ; Gemini gemini-4-argon ; DeepSeek deepseek-v4-pro ; OpenAI gpt-6-astra. Ces noms sont repris des données de l'essai et ne constituent pas une nouvelle vérification des offres disponibles. [L1 à L4 ; D1]

Sources et données complémentaires

Les repères cités désignent les débuts d'intervention dans les Word. Les liens ouvrent les pièces analysées.

[S1] Astra V2 — Test Marie 12mn OpenAI Astra élevé élevé.docx

[S2] Claude V2 — Test Marie 12mn Claude Opus 5.5 élevé.docx

[S3] Gemini V2 — Test Marie 12mn Gemini Argon élevé.docx

[S4] DeepSeek V2 — Test Marie 12mn Deepseek V4 Pro élevé.docx

[L1] Journal Astra — Test Marie 12mn OpenAI Astra élevé élevé.txt

[L2] Journal Claude — Test Marie 12mn Claude Opus 5.5 élevé.txt

[L3] Journal Gemini — Test Marie 12mn Gemini Argon élevé.txt

[L4] Journal DeepSeek — Test Marie 12mn Deepseek V4 Pro élevé.txt

[D1] Complément utilisateur du 3 novembre 2026 : Claude 117 s et 0,342168 € ; DeepSeek 190 s et ≈ 0,106164 € ; Gemini 158 s et ≈ 0,173104 € ; Astra 262 s et 0,693757 €. Le prix Gemini complète le journal initial, qui ne le renseignait pas.

Référence éditoriale : prompt « Réécriture soutenue », niveau 3, avec priorité à la fidélité, interdiction de synthétiser et traitement explicite des incertitudes. Les conclusions restent limitées à cet essai du 3 novembre 2026.