Claude Fable 5.1 : ce qui change vraiment

Les benchmarks lus honnêtement, les 7 changements de comportement avec le correctif de prompt de chacun, les 3 ruptures techniques, le vrai coût et la grille pour choisir ton modèle.

BLYX-AGENCY.COM · 2 SEPTEMBRE 2026

Anthropic a sorti Claude Fable 5.1 le 1er septembre 2026. En quelques heures, tout le monde a relayé la même chose : les scores explosent, le modèle est le plus puissant du monde, il coûte 25 à 45 % moins cher. Ces trois affirmations viennent de la même page, écrite par le fabricant.

Ce guide fait le travail que le communiqué ne fait pas. Il sépare ce qu'Anthropic annonce de ce que des tiers ont mesuré, il liste les sept comportements qui changent sans prévenir avec le correctif exact de chacun, il explique les trois ruptures techniques, et il te donne la grille pour savoir quand ce modèle est le bon choix. Parce que la réponse n'est pas toujours oui, et Anthropic l'écrit lui-même dans sa documentation.

Tout ce qui suit a été vérifié le 1er et le 2 septembre 2026 sur les sources primaires. Quand un chiffre vient d'Anthropic, c'est écrit. Quand il vient d'un tiers, c'est écrit aussi.

1. Ce qui a réellement changé

Fable 5.1 est presque un remplacement direct de Fable 5. Même interface de programmation, même façon de compter les tokens, même prix affiché, même fenêtre de 1 million de tokens, même limite de 128 000 tokens en sortie.

Ce qui ne bouge pas	Ce qui bouge
Prix d'entrée : 10 $ par million de tokens	Lecture de cache : de 1 $ à 0,25 $
Prix de sortie : 50 $ par million	Sept comportements par défaut
Fenêtre de contexte : 1 million de tokens	Trois règles techniques qui renvoient une erreur
Sortie maximum : 128 000 tokens	Cinq nouveautés, toutes en version d'essai
Façon de compter les tokens	Un filigrane invisible sur tout le texte produit

Un point de vocabulaire, parce que la presse mélange tout. Anthropic a sorti **deux** modèles le même jour : Fable 5.1, accessible à tous, et Mythos 5.1, réservé aux participants d'un programme d'accès restreint aujourd'hui limité à des organisations américaines. C'est le **même modèle**, avec des garde-fous différents. Le score de 60,9 % sur Terminal-Bench 4.0 qui circule partout est celui de Mythos, pas celui de Fable 5.1, qui plafonne à 55,8 %.

> ***Mythos 5.1 n'est pas une version plus puissante de Fable 5.1. C'est le même modèle avec moins de verrous, et la documentation précise mêm***

2. Les benchmarks, et comment les lire

Voici les scores publiés par Anthropic le jour du lancement. Tous sont mesurés par Anthropic sur ses propres bancs d'essai.

Benchmark	Fable 5.1	Fable 5	Opus 5	GPT-5.6 Sol
Terminal-Bench-Science 0.1	52,6 %	24,7 %	29,0 %	22,4 %
Terminal-Bench 4.0 (code)	55,8 %	42,0 %	52,3 %	37,3 %
AutomationBench (workflows)	31,4 %	17,1 %	26,9 %	19,6 %
CursorBench 3.2.0	73,4 %	70,5 %	70,0 %	67,2 %
OSWorld 2.0 (notation stricte)	41,7 %	36,1 %	39,6 %	non publié
Humanity's Last Exam (sans outils)	60,9 %	57,8 %	56,6 %	non publié

Trois choses à savoir avant d'utiliser ce tableau dans une réunion.

Anthropic a évalué son modèle avec ses garde-fous de production activés. Quand un garde-fou est intervenu pendant un test, le modèle a reçu un zéro sur OSWorld 2.0. Sur les autres tests, les tâches de cybersécurité ont été traitées par un autre modèle. Anthropic écri

sur blanc que cela réduit probablement ses propres scores. C'est honnête. C'est rarement repris.

La marge d'erreur est parfois plus grande que l'écart annoncé. Sur Terminal-Bench-Science, Anthropic annonce lui-même une erreur type de 3,5 à 4,5 points par modèle. Le classement public de ce benchmark donne d'ailleurs des chiffres différents des siens : Opus 5 à 30,0 % et Fable 5 à 21,4 %, contre 29,0 % et 24,7 % dans la reproduction interne.

Il manque un benchmark, et son absence est un signal. Le tableau de lancement ne contient aucun score SWE-bench, ni Verified ni Pro. C'était pourtant l'argument central du lancement de Fable 5. Au 2 septembre 2026, aucun score SWE-bench pour Fable 5.1 n'existe nulle part.

2.1 Ce que mesurent les tiers

Trois organisations indépendantes ont publié des mesures. Elles racontent une histoire plus nuancée.

Vals AI place Fable 5.1 premier de son classement général sur 51 modèles, avec un indice de 67,87 %, devant Opus 5 à 67,21 % et Fable 5 à 66,04 %. La marge d'erreur annoncée est de plus ou moins 1,10 point.

> ***L'écart entre Fable 5.1 et Opus 5 sur le classement Vals est de 0,66 point, pour une marge d'erreur de 1,10. Autrement dit, l'écart n'est pas statistiquement significatif, alors que Fable 5.1 coûte deux fois plus cher au token.***

Vals AI publie aussi deux choses que personne ne relaie. D'abord des résultats nettes : sur son banc juridique agentique, Fable 5.1 tombe à 6,67 % contre 11,25 % pour Fable 5, et il recule aussi sur le code médical, sur SAGE et sur l'évaluation fiscale. Ensuite un effet de mesure : les refus fréquents sur les sujets biologie et cybersécurité ont obligé Vals à activer des modèles de repli. En comptant ces tâches comme des échecs, l'indice passe de 67,87 % à 66,85 %, et sur un banc de rétro-ingénierie le score chute de 22,90 % à 10,69 %, avec 74 % des tâches assistées par un autre modèle.

Artificial Analysis classe Fable 5.1 premier sur 192 modèles pour l'intelligence, et mesure en même temps un coût de 3,69 $ par tâche contre 2,34 $ pour Opus 5, qui ne perd que trois points. À effort maximum, Fable 5.1 coûte environ 20 % de plus par tâche que Fable 5, parce qu'il consomme environ 1,7 fois plus de tokens de sortie. Le délai avant le premier mot passe de 121 secondes à 285 secondes.

CodeRabbit l'a testé sur 45 tâches de revue de code couvrant 105 défauts connus. Résultat : un défaut trouvé en moins que Fable 5, une précision qui monte de 32,8 % à 37,3 %, 87 commentaires en moins, et une latence de 18 minutes 38 par tâche contre 12 minutes 32, soit 48,7 % de plus.

Source	Ce qu'elle mesure	Le verdict
Anthropic	Ses propres bancs, garde-fous actifs	Gains larges, surtout sur le scientifique
Vals AI	51 modèles, effort maximum	Premier, mais à 0,66 point et avec des régressions
Artificial Analysis	Intelligence et coût par tâche	Premier en intelligence, 20 % plus cher par tâche

Source	Ce qu'elle mesure	Le verdict
CodeRabbit	Revue de code réelle	Plus précis, plus lent, un défaut de moins

3. Les six domaines où le gain est réel

Anthropic annonce six axes de progrès : le code agentique sur longue session, le travail documentaire, la recherche multi-étapes, la vision, le travail en contexte long et le pilotage d'un ordinateur.

Les cas ci-dessous viennent des 22 témoignages publiés par Anthropic le jour du lancement. Ce sont des partenaires en accès anticipé, sur des bancs d'essai internes non reproductibles. Ce sont des indices d'usage, pas des preuves.

- **Millennium** a identifié la cause d'un plantage survenant une fois sur un million d'exécutions, que l'équipe n'avait pas expliqué en quatre à cinq ans
- **Ramp** rapporte une session de 38 heures sans surveillance sur un problème d'apprentissage automatique, avec six expériences lancées en parallèle pendant la nuit
- **Browserbase** mesure 82 % de tâches réussies sur son banc de navigation le plus difficile, contre 74 % pour Opus 5 et 57 % pour Fable 5
- **Plaid** a fait cartographier un changement traversant huit services répartis sur trois bases de code

- **Crosby**, sur la relecture de contrats, passe de 47,9 à 57,0 sur son benchmark interne
- **Rakuten** a fait relire un projet de recherche clinique déjà validé par trois autres modèles, et Fable 5.1 y a trouvé une faille qu'aucun n'avait vue

> ***Astuce*** *: pour la vision, Anthropic recommande explicitement de donner au modèle un outil de recadrage et d'agrandissement d'image. Sans cet outil, le gain annoncé sur les graphiques denses et les tableaux imbriqués dans les PDF ne se matérialise pas.*

Le contrepoint mérite d'être lu avec la même attention. La revue **Every**, qui a eu un accès avant le lancement, documente des dépassements systématiques de consignes : 1 288 mots pour une limite de 1 000, huit thèmes pour une demande de trois à six, 43 citations pour une demande de huit à douze, dont cinq introuvables dans le texte source. Sur les mêmes tâches, Opus 5 est resté dans toutes les limites.

---

4. Les sept changements de comportement, et leur correctif

C'est la partie la plus utile de ce guide, et la moins reprise ailleurs. Anthropic documente sept dérives de comportement qui apparaissent **sans aucune modification de ton côté**. Ce ne sont pas des pannes : ce sont des défauts qui ont changé, et chacun a un correctif publié

Le changement	Ce que tu constates	Le geste
1. Appels d'outils dégroupés	L'agent est plus lent, coûte plus	Ajouter la phrase de regroupement à chaque tour
2. Moins de messages d'avancement	L'agent semble muet pendant des minutes	Activer l'affichage, puis demander les points d'étape
3. Répond de mémoire à effort bas	Réponses périmées sur l'actualité	Monter l'effort sur le tour, ou ajouter la règle de vérification
4. Prose plus dense	Phrases longues, moins d'air	Définir l'anti-modèle, pas la brièveté
5. Moins de mise en forme	Moins de gras, de titres, de listes	Supprimer tes vieilles règles anti-formatage
6. Citations non marquées	Passages recopiés sans guillemets	Donner un exemple complet, pas une règle
7. Fichiers réécrits en entier	Plus de tokens, plus de temps	Demander l'édition chirurgicale

4.1 Le regroupement des appels d'outils

Fable 5.1 peut n'émettre qu'un appel d'outil par tour là où Fable 5 en groupait plusieurs. Cela ne concerne que les boucles d'agent longues où les lectures suivantes sont seulement sous-entendues. Quand ta demande nomme explicitement plusieurs éléments à récupérer, le parallélisme fonctionne toujours. Anthropic précise que la qualité de la réponse n'est pas affectée : le coût est en tokens, en allers-retours et en temps.

```
First privately list what you need next; then request every item
that
doesn't depend on another's result in this one response.
```

Cette phrase doit être renvoyée après **chaque** retour de résultats d'outils, et surtout les copies précédentes doivent rester en place à l'octet près. Les supprimer revient à modifier l'historique, ce qui casse le cache et invalide les blocs de raisonnement suivants (voir la section 5).

4.2 Le silence de l'agent

Le modèle écrit moins de texte pendant les longues séquences d'outils, d'autant plus que l'effort est élevé. Attention : dans la majorité des cas, ce n'est pas le modèle qui se tait, c'est ton affichage qui ne demande pas ses messages. Les points d'étape reviennent dans des blocs de raisonnement qui sont **vides par défaut**.

Anthropic donne trois gestes, dans l'ordre. Activer l'affichage des mises à jour. Supprimer de ton prompt les vieilles lignes du type « garde tes conclusions pour la réponse finale ». Puis seulement demander la narration :

```
Before you start, say in a line what you're about to do; brief
updates
while you work help the user follow along. Close with a short recap
that
stands on its own — what you found, what you did, and what's next —
so a
reader who only sees the last message has the full picture.
```

4.3 Les réponses de mémoire

À effort bas, le modèle appelle moins souvent un outil de recherche et répond plus volontiers de mémoire. Sur un sujet qui bouge vite, comme les modèles d'IA justement, c'est le meilleur moyen d'obtenir une réponse fausse avec assurance. Le correctif tient en une règle à placer dans tes instructions :

```
When a query centers on a name you do not confidently recognize, or
recognize from a fast-moving area like AI models and developer
tools where
the landscape shifts within months, the name itself is the thing to
verify:
search before answering, and include the name as the user wrote it
in at
least one query alongside any reformulations. This holds even when
you have
some background on it — partial background is exactly what makes an
out-of-date answer sound authoritative, so familiarity is not a
reason to
skip the search.
```

4.4 La prose plus dense

Anthropic dit que l'écriture de Fable 5.1 est globalement meilleure, avec moins de formules toutes faites, mais parfois plus dense que celle de Fable 5. Le conseil est contre-intuitif : ne demande pas d'être plus bref, décris le défaut à éviter. La version courte, qu'Anthropic donne comme fonctionnant la plupart du temps, tient en cinq mots :

```
Please remove all mannered prose.
```

4.5 La mise en forme

Celui-là piège tout le monde. Les modèles précédents abusaient du gras et des listes, alors beaucoup de gens ont écrit des règles pour les en empêcher. Fable 5.1 penche dans l'autre sens : il met moins en forme. Résultat, tes vieilles règles anti-formatage suppriment maintenant une structure dont le contenu a besoin. Le correctif commence par **supprimer** ces règles, puis à les remplacer par une règle conditionnelle :

```
Use lists and bullet points when asked to, or when the content is
multifaceted enough that they help with clarity. If the person
explicitly
requests minimal formatting, always format your responses without
bullet
points, headers, lists, or bold emphasis, as requested. In
conversational,
personal, or emotional exchanges, keep to plain prose.
```

4.6 Les citations non marquées

En résumé de documents, le modèle reproduit plus souvent des passages du texte source sans les marquer comme citations. Le correctif recommandé n'est pas une règle mais **un exemple complet** ajouté à tes instructions : la demande, la réponse attendue, et une phrase qui explique pourquoi cette réponse est correcte. C'est le seul des sept correctifs qui demande un peu de travail de ta part.

4.7 Les réécritures de fichiers

Le modèle réécrit plus volontiers un fichier entier plutôt que de modifier trois lignes. Le résultat est généralement le même, mais ça

coûte des tokens de sortie et du temps. Anthropic affirme que l'instruction suivante ramène Fable 5.1 au niveau de Fable 5 sur les modifications petites et moyennes :

```
The number of tokens used to edit files is best minimized, all else
being
equal. Therefore, when it will not affect the end result, try to
surgically
edit a file rather than rewrite the entire thing.
```

> ***Attention*** *: ces trois comportements-là, le dégroupage des appels d'outils, la prose plus dense et la réécriture de fichiers entiers, augmentent la facture à prix affiché identique. C'est la contrepartie officielle de l'économie annoncée sur le cache, et c'est exactement le mécanisme que mesure Artificial Analysis avec ses 1,7 fois plus de tokens de sortie.*

---

5. Les trois ruptures techniques

Cette section ne concerne qu'une catégorie précise de gens : ceux dont le code construit lui-même la conversation envoyée au modèle. Si tu utilises Claude via l'application, via Claude Code ou via la trousse à outils officielle, tu n'as **rien** à faire : ces outils gèrent déjà l'historique correctement.

Première rupture : l'outil forcé disparaît. Obliger le modèle à appeler un outil précis renvoie désormais une erreur. La raison est logique une fois expliquée : le raisonnement est maintenant toujours actif, et un

appel forcé le court-circuiterait. Le modèle écrirait alors son raisonnement dans les arguments de l'outil, ce qui dégrade leur qualité. La solution consiste à nommer l'outil dans la consigne plutôt que dans le paramètre.

Deuxième rupture : le raisonnement est signé. Chaque bloc de raisonnement enregistre quel modèle l'a produit, et la compatibilité ne va que dans un sens. Fable 5.1 lit le raisonnement des modèles antérieurs, aucun modèle antérieur ne lit le sien. Un système qui bascule automatiquement vers un modèle plus ancien perd donc le raisonnement, **silencieusement**, sans erreur et sans surcoût visible.

Troisième rupture : modifier l'historique invalide le raisonnement. La signature couvre aussi tes instructions, tes outils et tous les messages précédents. Changer l'un des trois entre deux requêtes fait échouer la suivante. Ce contrôle n'est appliqué par défaut que pour les comptes créés le 31 août 2026 ou après, ce qui est l'inverse de l'intuition : les nouveaux venus sont contraints en premier.

La rupture	Qui est touché	Le remplacement officiel
Outil forcé	Ceux qui imposaient un appel d'outil	Nommer l'outil dans la consigne
Raisonnement signé	Les systèmes de repli entre modèles	Accepter la perte, ou rester sur le même modèle
Historique figé	Ceux qui injectent puis retirent du texte	Le message système limité à un tour

Le pourquoi est le plus intéressant. La documentation technique dit que le contrôle empêche qu'un raisonnement produit sous un jeu d'instructions soit rejoué sous un jeu d'instructions adverse : c'est u

protection contre la manipulation. L'annonce commerciale, elle, parle d'une mesure anti-distillation, qui ferme une technique publiquement documentée permettant d'extraire le raisonnement de Claude. Les deux raisons sont officielles et coexistent.

Et les cinq nouveautés livrées en même temps ne sont pas du confort : ce sont les remplacements exacts des trois manipulations d'historique désormais interdites. Changer le niveau d'effort en cours de conversation sans casser le cache, poser un message système valable pour un seul tour, récupérer les messages d'avancement en texte lisible, payer les lectures de cache quatre fois moins cher, et signer le contenu produit.

---

6. Le vrai prix

Voici la grille complète, en dollars par million de tokens.

Modèle	Entrée	Cache 5 min	Cache 1 h	Lecture de cache	Sortie
Fable 5.1	10 $	12,50 $	20 $	**0,25 $**	50 $
Fable 5	10 $	12,50 $	20 $	1 $	50 $
Opus 5	5 $	6,25 $	10 $	0,50 $	25 $
Sonnet 5	2 $	2,50 $	4 $	0,20 $	10 $
Haiku 4.5	1 $	1,25 $	2 $	0,10 $	5

La baisse ne porte que sur une ligne : la lecture de cache, divisée par quatre. Le reste est identique à Fable 5. Le « 25 à 45 % moins cher » est une **estimation** d'Anthropic sur des charges typiques, pas un tarif.

> ***Astuce*** *: cette économie n'existe que si ta charge de travail relit beaucoup un contexte déjà mis en cache. Une utilisation conversationnelle, qui ne met jamais rien en cache, n'économise strictement rien. Et à effort maximum, Artificial Analysis mesure au contraire 20 % de coût en plus par tâche. Les deux chiffres sont exacts en même temps : tout dépend de la part du cache dans ta facture.*

6.1 Ce que ça change si tu paies un abonnement

C'est le point que la plupart des gens découvrent en cliquant.

Ton plan	Accès à Fable 5.1
Gratuit	Aucun accès
Pro, Team standard	Non inclus, uniquement avec des crédits payés en plus
Max, Team premium, Enterprise premium	Inclus, dans la limite de 50 % de tes plafonds hebdomadaires

Dans Claude Code, Fable 5.1 n'est le modèle par défaut d'aucun plan : il faut le choisir explicitement, et la version 2.1.255 au minimum. Une confirmation s'affiche avant qu'une requête ne consomme des cré mais **uniquement en session interactive**. En mode automatisé, la

facturation part sans rien demander. C'est le piège le plus concret de

7. Quel modèle pour quelle tâche

Anthropic écrit dans sa propre documentation la phrase que personne n'a relayée le jour du lancement : pour la plupart des charges de travail, commence par **Opus 5**. Fable 5.1 est réservé au raisonnement exigeant et à l'agentique de longue haleine, ou au cas où tes tests sur Opus 5 à effort élevé échouent encore.

Ton besoin	Le modèle
La capacité maximale disponible, coûte que coûte	Fable 5.1
Code agentique complexe, travail d'entreprise	Opus 5
Équilibre vitesse et intelligence au quotidien	Sonnet 5
Latence et prix les plus bas	Haiku 4.5

Anthropic ajoute un conseil plus utile encore : **régler le niveau d'effort est souvent un meilleur levier que changer de modèle**. Il existe cinq niveaux, de bas à maximum, et le défaut est élevé. Passer au niveau intermédiaire donne des résultats qui égalent à peu près Fable 5, pour un coût nettement inférieur.

Une mesure d'Anthropic illustre à quel point l'intuition trompe : sur un banc de résolution de bugs, Fable 5.1 à effort **bas** résout 88,6 % de tâches pour 0,54 $ par tâche résolue, quand Sonnet 5 à son réglag

par défaut en résout 77,4 % pour 0,84 $. Onze points de mieux pour 35 % de plus cher, alors que le prix au token est cinq fois supérieur.

Et à l'inverse, CodeRabbit mesure qu'en revue de code, le réglage élevé donne un rappel **inférieur** au réglage bas, pour trois minutes de plus par tâche. Plus d'effort ne veut pas dire meilleur résultat.

---

8. Trois points qui comptent plus qu'un benchmark

La rétention des données. Fable 5.1 impose 30 jours de rétention et n'est pas disponible en rétention zéro sans autorisation expresse d'Anthropic. Si tu es sous contrainte contractuelle avec un client, ce seul point peut disqualifier le modèle quel que soit son score. Sur certains outils tiers, l'activer impose d'ailleurs cet engagement à toute ton équipe.

Le traitement par lots. L'option qui permet 300 000 tokens en sortie existe pour Opus 5 et Sonnet 5, mais pas pour Fable 5.1, plafonné à 128 000. Pour du traitement de masse, c'est un argument concret en faveur d'Opus 5.

Les refus. Les garde-fous bloquent 60 % de faux positifs en moins qu'avant, ce qui reste insuffisant pour certains usages. Un utilisateur intensif rapporte qu'un blocage à la cinquième heure d'une tâche de six heures fait perdre tout le travail. Si tes sujets touchent à la sécurité informatique ou à la biologie, teste avant de basculer.

---

9. La checklist

- [ ] Vérifier si ton plan donne accès à Fable 5.1, ou si tu paies des crédits en plus
- [ ] Retester tes niveaux d'effort, sans supposer que ceux de Fable 5 se transposent
- [ ] Commencer par le niveau intermédiaire avant de monter
- [ ] Supprimer de tes instructions les vieilles règles anti-formatage
- [ ] Ajouter la règle de vérification si tu travailles sur des sujets d'actualité
- [ ] Ajouter la consigne d'édition chirurgicale si le modèle touche à des fichiers
- [ ] Mesurer ta part de lecture de cache avant de croire à l'économie annoncée
- [ ] Vérifier tes contraintes de rétention de données avant tout usage client
- [ ] Ne jamais citer un score sans dire qui l'a mesuré

Un mot pour finir, parce qu'il vaut plus que tout ce qui précède.

Ce guide contient une trentaine de chiffres. Aucun ne dit ce que tu dois automatiser dans ton entreprise. Un modèle qui gagne trois points sur un banc d'essai ne change rien à une organisation où personne ne sait quelle tâche mérite d'être confiée à une machine en premier. La sortie d'un modèle est un événement de calendrier chez un fourni[illegible] américain, pas un événement dans ton activité.

La bonne question, quand un modèle sort, n'est pas « est-ce que je bosse ou pas ». C'est « est-ce que le travail que je lui confie est cadré ». Si la réponse est non, le meilleur modèle du monde produira des réponses justes et inutilisables, plus vite qu'avant.
