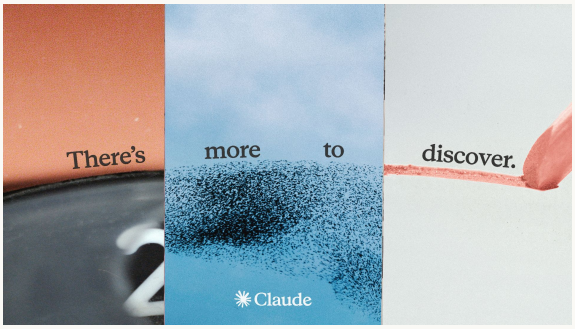
Claude BY ANTHROP\C

Haiku 5.5 : fait pour les gros volumes



Claude Haiku 5.5 est le modèle le plus rapide et efficace de la famille Claude 5.5, conçu pour le travail agentique à grand volume, sensible à la latence.

Il coûte aussi environ 75 % de moins que Claude Haiku 4.5 pour la plupart des tâches. Jusqu'à 100K tokens, il coûte 0,10 $ / 0,50 $ par million de tokens en entrée / sortie. Au-delà de 100K tokens, 0,50 $ / 2,50 $.

Usages :

Sous-agents

Un sous-agent rapide et économique pour le code et les tâches bien définies. Des modèles plus grands comme Opus 5.5 planifient et confient les sous-tâches à Haiku 5.5.

Code

Soutient des agents de code plus complexes en résumant les transcriptions, en classant et routant les requêtes, et en compactant le contexte.

Travail intellectuel

Extrait les champs clés de documents courts ou moyens et fait un premier tri pour signaler ce qui mérite un examen plus attentif.

Applis interactives

Assez rapide pour le support et le chat client en direct : réponses simples, rapides et utiles, sans attente.

Gros volumes

Gère classification, résumé et génération de texte à une vitesse et un coût qui rendent réaliste le traitement de gros volumes de contenu.

Utilisation de l'ordinateur

Un sous-agent d'utilisation de l'ordinateur pour les tâches répétitives : formulaires, saisie de données, transfert d'informations entre applis.

Quel modèle choisir

Modèle	Usages
Claude Fable 5.1	Problèmes de pointe : agents pour tâches longues, nouveaux usages, code et raisonnement les plus durs, autonomie totale
Claude Opus 5.5	Votre modèle au quotidien pour le travail d'entreprise complexe et exigeant en raisonnement, et pour le code complexe et de production
Claude Sonnet 5.5	Une alternative économique à Opus pour le travail bien cadré : créer des fonctionnalités, corriger des bugs, documents courts et soignés, présentations, diapositives et tableurs.
Claude Haiku 5.5	Travail à grand volume et sensible à la latence : sous-agents, classification, résumé

Selon les premiers utilisateurs, Claude Haiku 5.5 marque un net progrès sur Claude Haiku 4.5 dans leurs propres évaluations, avec des réponses plus rapides. Ils l'ont testé sur le travail où volume et latence comptent le plus : questions-réponses sur documents, routage, compaction et sous-agents de leurs produits. Plusieurs ont aussi souligné sa fiabilité. Il repère ses propres erreurs et dit clairement ce qu'il a fait ou non au lieu d'inventer une réponse.

Pourquoi passer votre agent à Claude Haiku 5.5

Plus capable

Un progrès sur Haiku 4.5 en code, outils, utilisation de l'ordinateur et agents.

Économique

Le client moyen de Haiku 4.5 économise 75 % sur sa facture Haiku.

Réglage de l'effort

Premier Haiku avec réglage de l'effort. Arbitrez coût et intelligence pour chaque tâche.

Pour migrer, remplacez la chaîne de modèle par claude-haiku-5-5. Le skill claude-api, préinstallé dans Claude Code, peut faire le changement pour vous avec /claude-api migrate.

Claude Haiku 5.5

	Haiku 5.5	Haiku 4.5	GPT-6 Luna	Sonnet 5.5 For reference
Knowledge work GDPval-AA v2.1	**1620**	735	1437	1840
Knowledge work AA-Briefcase v1.1	**1578**	614	1336	1824
Computer use OSWorld 2.1 (Offline subset)	**72.4%**	15.7%	48.9%	83.9%
Multidisciplinary reasoning Humanity's Last Exam (no tools)	**45.9%**	10.2%	—	56.9%
Multidisciplinary reasoning Humanity's Last Exam (with tools)	**57.4%**	18.7%	—	64.5%
Agentic coding Terminal-Bench 4.0	**39.2%**	0.0%	16.4%	70.6%
Agentic coding FrontierCode 1.1 (Main)	**46.4%**	—	42.4%	52.1% Xhigh
Visual reasoning Chartography (no tools)	**46.4%**	6.4%	29.1%	61.6%

For details on how we run our evaluations, see the Haiku 5.5 System Card.