IAVENIR. NEWSLETTER 09.02.2026

FOCUS OUTILS IA



INTRODUCTION

Cette édition fait un arrêt sur image sur les grands modèles d’IA réellement utilisables aujourd’hui en entreprise. Nous passons en revue GPT 5.2, Claude Opus 4.6, Gemini 3 Pro, Mistral Large via Le Chat, DeepSeek V3.2 et Grok pour rappeler leurs fonctions concrètes en matière de projets, recherche approfondie, images, vidéo, audio et agents.

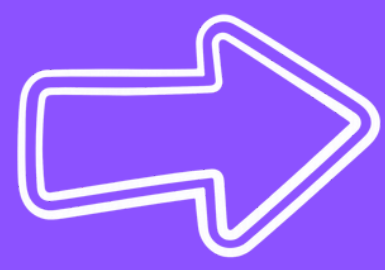
L’objectif est de clarifier quels outils servent le mieux vos cas d’usage plutôt que de courir après la dernière annonce marketing.

OPENAI - CHATGPT GPT 5.2

GPT 5.2 est un modèle propriétaire d’OpenAI orienté travail professionnel, avec un mode de raisonnement approfondi et un contexte étendu adapté aux dossiers volumineux. Dans ChatGPT, il alimente les Projets qui regroupent fichiers et conversations par thème, Deep Research pour produire des notes sourcées à partir du web et de vos documents, et les expériences audio temps réel pour des agents vocaux.

Concrètement, il permet de coder, d’analyser des contrats, de préparer des présentations et de piloter des suites de tâches dans un même espace de travail.

Sources : OpenAI · OpenAI release notes · Digital Applied



ANTHROPIC - CLAUDE OPUS 4.6

Claude Opus 4.6 est un modèle propriétaire d’Anthropic avec une fenêtre de contexte en bêta jusqu’à un million de tokens et une très bonne maîtrise du code.

Il introduit les Agent Teams qui permettent à plusieurs instances de Claude de se répartir un travail multi étapes, ainsi qu’un panneau d’artéfacts pour afficher documents, tableurs ou slides générés à côté de la conversation. En pratique, il sert bien pour la refonte de grands codebases, la revue de fichiers denses et la préparation de présentations directement à partir de données brutes.

Sources : Anthropic · The Verge · ITPro

GOOGLE - GEMINI 3 PRO

Gemini 3 Pro est un modèle propriétaire de Google qui alimente l'application Gemini, Workspace et Vertex AI, avec un long contexte et une compréhension multimodale du texte, des images, de l'audio, de la vidéo et du code.

Il sert de moteur au mode Deep Research qui planifie et exécute des recherches longues sur le web et dans vos documents, et se combine à Veo 3.1 pour générer des vidéos de quelques secondes avec son dans l'interface Gemini. Dans un contexte entreprise, il permet par exemple de résumer des corpus entiers, de générer des visuels et d'automatiser des workflows sans sortir de l'écosystème Google. .

Sources : Google DeepMind · Gemini Enterprise docs · Gemini vidéo

MISTRAL - LE CHAT MISTRAL LARGE

Mistral Large, en version Large 3 open weight, est le grand modèle généraliste de Mistral avec un contexte étendu et une architecture mixte pensée pour le raisonnement, le code et la multimodalité.

Dans Le Chat, il combine questions classiques, le mode Deep Research, compréhension de documents, génération et édition d’images, ainsi que transcription et conversation audio via Voxtral. Concrètement, il permet d’accéder à toutes les fonctionnalités IA utiles, dans un environnement aligné avec les attentes de souveraineté.

Sources : Mistral Large 3 · Le Chat Deep Research · Voxtral Mistral

XAI - GROK 4.1

Grok 4.1 est un modèle propriétaire de xAI accessible dans X et via API, avec une variante Grok 4.1 Fast optimisée pour l’appel d’outils et une fenêtre de contexte pouvant atteindre deux millions de tokens. L’Agent Tools API donne accès aux données temps réel de X, à la recherche web et à l’exécution de code, tandis que Grok Imagine 1.0 permet de générer des vidéos d’environ dix secondes en 720p avec audio synchronisé via une API dédiée.

Ensemble, ces briques servent à construire des agents de support ou de finance branchés sur X et à produire rapidement des contenus vidéo courts, dans un contexte de forte vigilance réglementaire sur les deepfakes.

Sources : xAI Grok 4.1 Fast · Grok Imagine API · eWeek

DEEPSEEK - DEEPSEEK V3.2

DeepSeek V3.2 est un modèle open weight pensé pour les agents, qui équilibre efficacité de calcul et performances de raisonnement grâce à son mécanisme “DeepSeek Sparse Attention” et à un cadre de renforcement très poussé.

Il gère de longs contextes et intègre la réflexion directement dans l’usage des outils, ce qui le rend adapté à des agents capables d’appeler des API, des bases de connaissances et des environnements d’exécution tout en expliquant leurs étapes. Les poids et guides publics permettent de le déployer sur vos clusters ou via des fournisseurs compatibles, avec un contrôle fin sur les coûts et la gouvernance technique.

Sources : DeepSeek V3.2 arXiv · DeepSeek Hugging Face · vLLM recipes

CONCLUSION

Le panorama actuel se structure autour de quelques modèles propriétaires très intégrés et de quelques modèles ouverts puissants, chacun avec des points forts bien identifiés en projets, recherche approfondie, images, audio, vidéo ou agents. La bonne approche consiste à choisir un portefeuille réduit de modèles et à les relier à des cas d'usage précis.

Quels outils retiendriez vous pour devenir vos standards de travail sur les projets, les artéfacts, la recherche, l'image, la voix et la vidéo dans les prochains mois ?

ABONNE TOI !