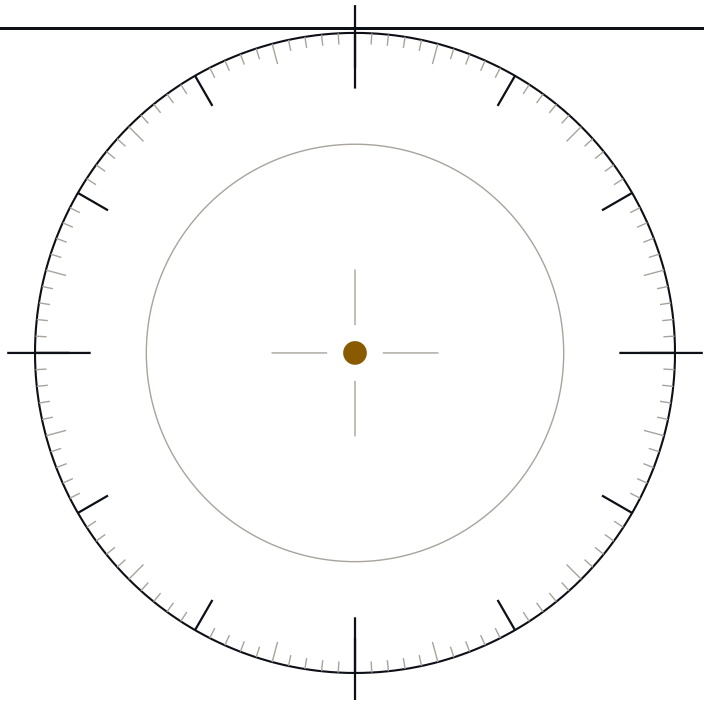
DEEO.AI

FICHE D'ACTUALITÉ · ACCÈS LIBRE



01 · IA

VEILLE IA · FICHE D'ACTUALITÉ · 3 OCTOBRE 2026

L'IA fabrique l'IA : puces conçues par agents, prix en chute libre, disjoncteur NVIDIA

Fiche d'actualité DEEO.AI · Veille IA. Situation au 23.10.2026.

PUBLICATION	3 octobre 2026
SITUATION	23.10.2026 · faits vérifiés à cette date
DOMAINE	01 · IA · Intelligence artificielle
VIDÉO	8 min 26, sur deeo.ai et YouTube : la vidéo et la fiche disent la même chose.

Vidéo · 8 min 26

deeo.ai/videos/veille-ia-2026-10-03/

PARTAGE

Fiche en accès libre : vous pouvez la partager, en citant DEEO.AI.

CHAPÔ

L'IA fabrique l'IA. Fin septembre 2026, trois annonces l'ont montré chacune à sa façon. Le 28 septembre, Synopsys a présenté des agents d'IA capables d'enchaîner seuls des étapes entières de la conception d'une puce, sous le contrôle d'humains. Le 22 septembre, l'institut Epoch AI a calculé qu'en moins de dix-huit mois, le même niveau d'IA était devenu 725 fois moins cher sur un examen scientifique, et environ 13 fois moins cher chaque année. Le 28 septembre encore, NVIDIA a présenté un dispositif qui surveille un agent depuis l'extérieur et, selon l'entreprise, peut l'arrêter en quelques millisecondes. Une IA plus capable, moins chère, et qu'il faut pouvoir couper : voici ce qu'il faut en retenir.

Sources numérotées en fin de fiche, avec leur fiabilité. Les trois sujets ont été repérés dans le récapitulatif hebdomadaire de Tom's Hardware Premium du 3 octobre 2026, puis vérifiés sur les sources primaires, et revérifiés le 23.10.2026 ; aucune de ces annonces ne date du 3 octobre. Les chiffres annoncés par une entreprise sont présentés comme tels, et les mises en contexte de DEEO.AI sont signalées.

EN UNE MINUTE

- **Des puces conçues par des agents** : Synopsys, éditeur de logiciels qui servent à dessiner les puces, a présenté le 28 septembre une plateforme, Synopsys Autopilot, et une gamme d'agents d'IA, AgentEngineer. Ces agents enchaînent seuls des étapes complètes de conception, mais les validations cruciales restent humaines [1] [2].
- **Des chiffres annoncés, pas encore mesurés par d'autres** : jusqu'à 50 fois plus rapide pour boucler le contrôle d'une puce, selon des travaux menés avec NVIDIA ; 10 à 30 % de productivité en plus, selon Fujitsu. Disponibilité générale prévue fin 2026 [1].
- **Le prix de la pensée s'effondre** : selon l'étude d'Epoch AI du 22 septembre, à niveau égal, le coût d'une réponse d'IA baisse d'environ 47 % par trimestre depuis 2023, soit un prix divisé par environ 13 en un an. C'est six fois plus rapide que la baisse historique du coût du calcul informatique [3].
- **725 fois moins cher** : sur des questions scientifiques de haut niveau, un niveau qui coûtait environ 30 centimes de dollar par question en janvier 2025 s'obtenait, moins de dix-huit mois plus tard, pour 725 fois moins. C'est le prix d'un niveau sur des tests, pas la facture totale de l'IA [3].
- **Un disjoncteur hors du modèle** : NVIDIA a présenté le 28 septembre une plateforme de sécurité pour les agents. Sa partie matérielle, Sentry, surveille l'agent de l'extérieur et, selon NVIDIA, l'isole en quelques millisecondes s'il sort du périmètre fixé [4] [5].

1. Synopsys : des agents d'IA qui conduisent la conception de puces

1.1 **Ce qui a été annoncé.** Le 28 septembre 2026, Synopsys, une entreprise qui édite les logiciels avec lesquels on dessine les puces électroniques, a présenté **Synopsys Autopilot**. C'est une plateforme qui sert de socle à une gamme d'agents d'IA appelée **AgentEngineer** [1]. Un agent d'IA, c'est un programme qui ne se contente pas de répondre à une question : il reçoit un objectif, découpe le travail en étapes et les exécute les unes après les autres.

1.2 Synopsys parle d'agents « à long horizon ». Ils raisonnent, planifient et enchaînent seuls des étapes entières d'ingénierie [2]. Deux exemples cités :

- **la vérification**, c'est-à-dire le contrôle que la puce fera bien ce qu'on attend d'elle, avant de la fabriquer ;
- **l'implémentation physique** : placer les blocs de la puce, tracer les connexions entre eux et s'assurer que les signaux arrivent à l'heure.

1.3 L'organisation compte un agent par grand domaine, qui s'appuie sur des agents plus spécialisés pour des tâches précises [2].

1.4 **Pourquoi c'est un travail colossal.** Une puce haut de gamme compte aujourd'hui des dizaines de milliards de transistors, ces minuscules interrupteurs qui s'allument et s'éteignent pour calculer. Les plus grosses puces d'IA dépassent 80 milliards de transistors [1].

1.5 Pour se le représenter : dessiner une puce, c'est comme établir le plan de circulation d'une immense ville, avec des milliards de carrefours. Il faut placer chaque quartier, tracer chaque rue, et vérifier qu'aucun embouteillage ne retarde les livraisons. Les agents de Synopsys proposent de tracer et de contrôler ce plan par étapes entières, là où les outils précédents n'aidaient que sur des points précis.

1.6 **Autonomie, mais sous contrôle.** Synopsys place ces agents au niveau le plus élevé, le niveau 5, de **sa propre** échelle d'autonomie, une échelle qu'il a lui-même définie en mars 2025. Il précise pourtant que **les points de validation cruciaux restent humains** [1] [2]. Autrement dit, l'IA ne conçoit pas les puces sans humains : des ingénieurs approuvent les étapes décisives.

1.7 **Les chiffres, tels qu'ils sont annoncés** [1] :

- **jusqu'à 50 fois plus rapide** pour boucler la vérification ;

— **20 % de couverture en plus**, c'est-à-dire une part plus grande des situations possibles effectivement contrôlées.

1.8 Ces deux chiffres viennent de travaux menés avec NVIDIA, présentés lors d'un salon professionnel en juillet 2026, et comparés aux méthodes de Synopsys sans agents.

— **10 à 30 % de productivité en plus**, selon Fujitsu, pour écrire le texte informatique qui décrit le fonctionnement de la puce (le communiqué ne précise pas la base de comparaison).

1.9 Ce sont des chiffres annoncés par Synopsys et ses partenaires, pas des mesures indépendantes.

1.10 **Où en est-on ?** Plus de 50 entreprises testent déjà la plateforme avec Synopsys. La disponibilité générale, c'est-à-dire la mise en vente à tous, est **prévue pour fin 2026**. TSMC, Intel, Samsung, MediaTek, NVIDIA et Fujitsu figurent parmi les partenaires qui ont soutenu l'annonce [1]. Soutenir une annonce n'est pas l'utiliser en production : ces noms ne sont pas présentés comme des clients du déploiement. Deux jours plus tard, le 30 septembre, OpenAI et Synopsys ont annoncé GPT-Synopsys, un modèle d'IA dédié à la conception de puces et intégré à Autopilot, sans date de disponibilité ni chiffre [6].

1.11 **Ce qui est vraiment nouveau.** Synopsys fait de l'IA dans la conception de puces depuis plusieurs années. Son outil DSO.ai, lancé en 2020, optimisait déjà des étapes précises. La nouveauté, ce sont des agents qui enchaînent des parcours complets, d'un bout à l'autre d'une étape d'ingénierie [1] [2].

2. Le prix de la pensée s'effondre

2.1 **L'étude.** Le 22 septembre 2026, l'institut de recherche Epoch AI a publié « The plunging price of thought » (« Le prix de la pensée en chute libre »), signé Luke Emberson et David Roodman [3].

2.2 **La question posée.** Combien coûte, en dollars, le modèle d'IA **le moins cher** capable d'atteindre un score donné sur un test ? Le coût retenu est réel, calculé par question, en comptant aussi le « raisonnement » que le modèle produit avant de répondre [3].

2.3 Cinq tests servent d'examens : des questions scientifiques de niveau doctorat (le test GPQA Diamond), des mathématiques de compétition et de recherche, des problèmes d'échecs et des jeux de logique [3].

2.4 **Les chiffres** [3] :

- depuis 2023, à niveau de performance égal, ce coût baisse d'environ **47 % par trimestre** ;
- sur un an, cela fait un coût divisé par **environ 13** ;
- c'est **six fois plus rapide** que la baisse historique du prix du calcul informatique (sur 1940-2001, ce prix était divisé par environ 1,5 chaque année) ; l'étude ne parle pas de « loi de Moore ».

2.5 **L'exemple des 725 fois.** En janvier 2025, le modèle o3 d'OpenAI atteignait un niveau donné sur les questions scientifiques de GPQA Diamond (75 % au-dessus du hasard), pour environ 30 centimes de dollar par question. Moins de dix-huit mois plus tard, un autre modèle, GPT-5.6 Luna, faisait aussi bien pour une fraction de centime : **725 fois moins cher**, selon l'étude [3]. Ce facteur est calculé sur des valeurs exactes que l'étude ne publie pas : les prix arrondis cités ici ne permettent pas de le retrouver à l'unité.

2.6 **Comparé à d'autres technologies.** L'étude elle-même rapproche cette chute de deux autres baisses célèbres [3] :

- elle est **quatre fois plus rapide** que la baisse du coût de la lecture de l'ADN, le séquençage, sur 2001-2025 ;
- elle est **dix-huit fois plus rapide** que celle des batteries lithium-ion, celles de nos téléphones, sur 1991-2024 ;
- et cinquante-quatre fois plus rapide que celle de l'électricité, sur 1892-1973.

2.7 Pour se le représenter : le séquençage de l'ADN et les batteries sont deux exemples classiques de technologies devenues abordables à une vitesse spectaculaire. Selon l'étude, le prix d'un niveau d'IA baisse encore plus vite. Les auteurs comparent ces vitesses en comptant les divisions successives du prix, pas les montants en dollars.

2.8 **Une baisse qui ralentit.** La chute est la plus forte juste après qu'un niveau est atteint pour la première fois. Elle ralentit ensuite [3].

2.9 **Les réserves des auteurs eux-mêmes** [3] :

- les modèles peuvent être entraînés pour briller sur les tests, comme un élève qui bachote les annales ;
- réussir un test ne prouve pas qu'on fait un travail utile ;
- l'étude suppose qu'on passe toujours au modèle le moins cher ;
- les données ne couvrent qu'environ trois ans ;
- comparer l'IA aux batteries, c'est un peu « comparer des pommes et des oranges ».

2.10 **Ce que l'étude ne dit pas.** Elle mesure le prix d'un niveau de performance donné. Elle ne dit pas que l'IA est devenue gratuite, ni que la facture totale de l'IA baisse, ni que l'énergie qu'elle consomme diminue [3].

3. NVIDIA : un disjoncteur hors de la maison

3.1 **Ce qui a été annoncé.** Le 28 septembre 2026, NVIDIA a présenté l'**Open Agent Safety Platform**, une plateforme de sécurité pour les agents d'IA, du test jusqu'à la mise en service [4]. Elle a deux étages.

3.2 **Premier étage : OpenShell, le logiciel.** OpenShell fait tourner l'agent dans un « bac à sable » : un espace fermé où il peut agir sans toucher au reste de la machine. Il enregistre chacune de ses actions, comme un journal de bord, et applique les règles fixées par la personne qui le met en service [4] [5]. OpenShell est un logiciel ouvert : son code est public et réutilisable librement, sous la licence Apache 2.0. Il était déjà disponible depuis février 2026.

3.3 **Second étage : Sentry, le matériel.** Sentry est un système de référence matériel, c'est-à-dire un modèle d'installation que d'autres peuvent reproduire. Il fonctionne sur les cartes réseau programmables BlueField-4 de NVIDIA, des cartes qui gèrent le trafic d'un serveur. Sentry surveille l'agent **de l'extérieur**, en dehors du modèle d'IA et de son logiciel [4].

3.4 Selon NVIDIA, si un agent tente de sortir du périmètre qu'on lui a fixé, Sentry **le met en quarantaine et l'arrête en quelques millisecondes**. La règle est appliquée « au niveau du silicium », directement par le matériel, et non par le logiciel de l'agent [4]. Le déclencheur est bien une sortie du périmètre défini, pas un simple comportement suspect.

3.5 Pour se le représenter : un disjoncteur placé hors de la maison. Si l'installation électrique s'emballe, il coupe le courant de l'extérieur, quoi qu'il se passe à l'intérieur. Sentry joue ce rôle pour l'agent : il n'a pas besoin que l'agent coopère pour l'arrêter.

3.6 **Les chiffres, à prendre comme tels** [4] :

- **quelques millisecondes** pour isoler un agent qui sort de son périmètre : c'est une affirmation de NVIDIA, sans mesure indépendante publiée ;
- **plus de 100 organisations** soutiennent la plateforme, dont Anthropic, Microsoft, Cisco, CrowdStrike, Salesforce, SAP et Red Hat.

3.7 **Où en est-on ?** Le logiciel OpenShell est disponible en accès libre [5]. Sentry, lui, reste un système de référence : ni date de disponibilité générale ni prix n'ont été annoncés [4]. Le 21 octobre 2026, NVIDIA a lancé une « architecture de référence » OpenShell, présentée comme disponible, qui intègre Sentry ; Sentry lui-même n'a toujours ni date ni prix propres. NVIDIA y cite de nouveaux partenaires, dont OpenAI, AWS, Google Cloud et Meta [7].

4. La question qui reste

4.1 Mise en contexte de DEEO.AI : les trois annonces ne se citent pas entre elles, mais elles dessinent une même boucle. L'IA aide à concevoir les puces qui la font tourner. Le prix d'un niveau d'IA chute plus vite que tout ce que l'étude d'Epoch AI a pu comparer. Et les industriels commencent à installer des disjoncteurs hors du modèle, pour pouvoir arrêter un agent qui sort du cadre.

4.2 Chacune de ces annonces garde ses limites : des chiffres annoncés par les entreprises, des humains toujours aux commandes chez Synopsys, une étude qui mesure des tests et pas la facture totale, un disjoncteur matériel qui n'est pas encore en vente.

4.3 Si l'IA devient à la fois plus capable et presque gratuite, qui doit tenir la main sur l'interrupteur ?

Sources

Consultées pour l'épisode du 03.10.2026. Fiabilité indiquée en italique.

[1] Synopsys, communiqué présentant la plateforme Synopsys Autopilot et la gamme AgentEngineer, 28.09.2026, https://finviz.com/news/396005/synopsys-powers-autonomous-engineering-with-a-broad-portfolio-of-long-horizon-agents-and-autopilot-platform . *Source primaire (communiqué de l'entreprise) ; les chiffres de performance sont annoncés par Synopsys et ses partenaires, sans mesure indépendante.*

[2] Synopsys, billet de blog sur les agents « à long horizon » AgentEngineer, https://www.synopsys.com/blogs/chip-design/long-horizon-agentengineer-solutions.html . *Source primaire (présentation par l'entreprise).*

[3] L. Emberson et D. Roodman (Epoch AI), « The plunging price of thought », 22.09.2026, https://epoch.ai/publications/the-plunging-price-of-thought . *Étude d'un institut de recherche ; les réserves sont celles des auteurs.*

[4] NVIDIA, communiqué présentant l'Open Agent Safety Platform, 28.09.2026, https://investor.nvidia.com/news/press-release-details/2026/NVIDIA-Launches-Open-Agent-Safety-Platform-to-Secure-Agents-From-Testing-to-Deployment/ . *Source primaire (communiqué de l'entreprise) ; le délai de quelques millisecondes est une affirmation de NVIDIA.*

[5] NVIDIA, dépôt du logiciel OpenShell sur GitHub, https://github.com/NVIDIA/OpenShell . *Source primaire (code publié).*

DEEO.AI cite ces sources et ne garantit pas leurs chiffres ; les estimations et déclarations restent sous la responsabilité de leurs auteurs. 6. Synopsys et OpenAI, communiqué présentant GPT-Synopsys, 30.09.2026, https://news.synopsys.com/2026-09-30-OpenAI-and-Synopsys-Announce-GPT-Synopsys-Frontier-Intelligence-to-Revolutionize-Chip-Design . *Annonce des entreprises, sans date ni chiffre.* 7. NVIDIA, communiqué présentant l'architecture de référence OpenShell, 21.10.2026, https://nvidianews.nvidia.com/news/openshell-reference-architecture . *Source primaire (communiqué de l'entreprise).*

SITUATION	Faits vérifiés au 23.10.2026, sur les sources primaires.
VIDÉO	La vidéo de cette actualité (8 min 26) est sur sa page de deeo.ai : https://deeo.ai/videos/veille-ia-2026-10-03/
PARTAGE	Fiche en accès libre : vous pouvez la partager, en citant DEEO.AI.
CONTACT	admin@deeo.ai · deeo.ai

DEEO.AI

OBSERVATOIRE DE L'ÉMERGENCE DE L'IA