Discussion :
Les diligents anticipent sur une extension de la pénurie de RAM jusqu'en 2028 via la demande liée à l'IA
Sujet : Intelligence artificielle

02/10/2025, 18h30 #101

Patrick Ruiz
Chroniqueur Actualités
Inscrit le : Février 2017
Messages : 2 636

Les diligents anticipent sur une extension de la pénurie de RAM jusqu'en 2028 via la demande liée à l'IA

Les diligents anticipent sur une extension de la pénurie de RAM jusqu'en 2028 via la demande liée à l'IA. Micron rapporte avoir déjà vendu sa production de RAM dynamique pour 2027 plus cher qu'en 2026

L'industrie traverse une crise majeure marquée par l'extension des cycles de production et l'explosion des tarifs. Cette situation a principalement causée par une demande pour les infrastructures d'IA, incitant les fabricants comme Samsung, Micron et SK Hynix à privilégier les puces à haute performance dédiées aux data centers au détriment du produit grand public. Les diligents anticipent sur une extension de cette pénurie de RAM jusqu'en 2028.

Lors de la conférence de résultats de Micron du 30 septembre 2026, son PDG Sanjay Mehrotra a officiellement déclaré que la pénurie de mémoire vive (DRAM) allaient s'aggraver en 2027 et 2028. De son côté, la direction de Samsung Electronics maintient son directeur financier et ses vice-présidents à également confirmé lors de ses récents bilans financiers que la pénurie s'intensifierait, excluant tout retour à la normale avant 2028.



Les diligents anticipent sur une aggravation de la situation en raison du shrinkage de la production par l'intelligence artificielle (entre autres facteurs)

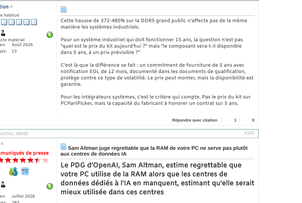
Pour faire tourner les modèles d'intelligence artificielle, les géants de la filière technologique s'arrachent la mémoire à haute bande passante (HBM), beaucoup plus lucrative pour les fabricants. En témoigne de la mémoire HBM3E ou HBM4 consomme environ 3 fois plus de plaquettes de silicium (wafers) que la RAM classique (DDR5). Chaque ligne de production allouée à l'intelligence artificielle détruit donc le triple de capacité pour les PC et smartphones grand public.

Les limites physiques des usines (salles blanches) font partie des autres raisons mises en avant par les responsables. Augmenter la production prend énormément de temps. Micron et Samsung investissent des centaines de milliards de dollars, mais les nouvelles usines et zones hautement contrôlées (salles blanches) n'entreront en service qu'entre fin 2028 et 2030.

Certains de ces entreprises se retirent même déjà de certains de leurs segments grand public. Devant l'explosion des marges sur les serveurs, Micron a purement et simplement abandonné une partie du marché grand public (en fermant notamment sa marque grand public Crucial) pour dédier 100 % de sa production de pointe aux centres de données.

La réalité actuelle est que les carnets de commandes sont déjà complets. Micron a annoncé avoir déjà vendu la quasi-totalité de sa production de 2027 à des tarifs très élevés à des entreprises comme Nvidia, OpenAI ou Microsoft. Cette rareté artificielle engendre des hausses de prix massives (le prix de certaines barrettes a été multiplié par 5 en un an), ce qui va se répercuter par une hausse du prix des ordinateurs et smartphones d'ici à la fin des années 2027-2028.





Cette flambée des prix de la mémoire vive crée un contexte économique et technique particulièrement favorable aux systèmes d'exploitation open source, principalement Linux

Le marché informatique grand public et professionnel fait face au problème de l'obsolescence logicielle de Windows et à cette indicte des PC censés tourner sous Windows :

- Les PC neufs deviennent accessivement chers en raison du coût de la mémoire vive
- L'arrêt définitif du support de Windows 10 pousse vers Windows 11, un système qui exige au minimum 4 Go de RAM (et en pratique 8 à 16 Go pour tourner de manière fluide avec les outils d'IA natifs comme Copilot). De nombreuses entreprises et particuliers se retrouvent avec des parcs informatiques parfaitement fonctionnels mais jugés obsolètes par Microsoft.

C'est là qu'entre en jeu Linux. Là où Windows a besoin de toujours plus de mémoire pour faire tourner ses services en arrière-plan, une distribution Linux moderne peut redonner vie à un ordinateur équipé de seulement 2 Go ou 4 Go de RAM. Face à l'impossibilité financière de rajouter des barrettes de RAM (devenues hors de prix), basculer sur Linux devient la solution la plus rationnelle pour garder sa machine performante.

Pour les gestionnaires de parcs informatiques, l'équation devient purement financière. Acheter des milliers de licences Windows et mettre à niveau la RAM des ordinateurs au prix fort est un gouffre financier. Migrer vers un écosystème Linux permet de gérer les investissements matériels en attendant la fin de la pénurie vers 2028.

Ironiquement, la pénurie de RAM est causée par l'infrastructure Cloud et l'IA, des secteurs où Linux règne déjà en maître absolu (plus de 90 % des serveurs mondiaux). Moins de RAM disponible pour le grand public signifie que le calcul va davantage se déporter vers le Cloud et ce, sur des serveurs tournant sous Linux.

En somme, la crise de la RAM agit en principe comme un frein violent à la course à l'accès à Windows. Cela offre à Linux une opportunité historique de s'imposer chez les particuliers et les professionnels comme l'alternative de la résilience économique.
Et vous ?

Que pensez-vous de la situation actuelle de la filière IT faite de pénurie de mémoire vive ? Y voyez-vous aussi une opportunité pour des systèmes d'exploitation comme Linux de décoller dans la filière des ordinateurs de bureau ?

Voir aussi :

Apple a fortement augmenté ses prix, jusqu'à +300 $ sur l'ensemble de ses gammes Mac et iPad, alors que la forte hausse de la demande liée à l'IA entraîne une pénurie mondiale de mémoire et de stockage

L'IA asséche le marché du matériel informatique : quand RAM, GPU et stockage deviennent des ressources sous tension permanente. Les prix des disques durs ont augmenté en moyenne de 46 % depuis septembre

La mémoire représente désormais 63 % du coût total des composants des puces d'IA. La bulle n'est vraiment les prix de la mémoire et force des fabricants comme Microsoft à augmenter le prix de leurs appareils

Contribuez au club : Corrections, suggestions, critiques, ... : Contactez le service news et Rédigez des actualités

03/10/2025, 03h06 #102

lasagfilan
Membre habitué
Inscrit le : Août 2026
Messages : 13

Cette hausse de 372-485% sur la DDR5 grand public n'affecte pas de la même manière les systèmes industriels.

Pour un système industriel qui doit fonctionner 15 ans, la question n'est pas "quel est le prix d'une barrette ?" mais "le composant sera-t-il disponible dans 5 ans, à un prix prévisible ?"

C'est là que la différence se fait : un commitment de fourniture de 5 ans avec notification EOL de 12 mois, documentation dans les documents de qualification, protège contre ce type de volatilité. Ce prix peut monter, mais la disponibilité est garantie.

Pour les intégrateurs systèmes, c'est le critère qui compte. Pas le prix du kit sur PCPartPicker, mais la capacité du fabricant à honorer un contrat sur 5 ans.

Elora
Communiquée de presse
Inscrit le : Juillet 2026
Messages : 283

Sam Altman juge regrettable que la RAM de votre PC ne serve pas plutôt aux centres de données IA

Le PDG d'OpenAI, Sam Altman, estime regrettable que votre PC utilise de la RAM alors que les centres de données dédiés à l'IA en manquent, estimant qu'elle serait mieux utilisée dans ces centres

Face à la pénurie mondiale de mémoire vive qui perdure, Sam Altman, PDG d'OpenAI, estime que le fait que les particuliers disposent de grandes quantités de mémoire vive sur leurs ordinateurs personnels, où cela ci reste souvent inutilisé, est regrettable. Il estime que cette mémoire serait mieux utilisée dans les centres de données, où elle pourrait servir à de nombreux utilisateurs. *« Beaucoup de gens veulent une inférence locale », a déclaré Altman. « Mais d'un point de vue de répartition mondiale, c'est en réalité assez regrettable que chacun dispose d'une partie de l'offre mondiale limitée de DRAM dans son propre ordinateur, alors que celle-ci n'est utilisée que 5 % du temps. »* Ses propos interviennent à un moment où les entreprises spécialisées dans l'IA grandissent leurs centres de données et se disputent les puces mémoire.

En janvier 2026, l'essor de l'IA accaparait une grande part de la production mondiale de mémoires. Cette situation avait provoqué une pénurie et perturbait les autres marchés (smartphones, objets connectés, consoles, PC...). De nouvelles prévisions indiquent que les centres de données consommeraient 70 % des puces mémoire fabriquées en 2026 et que la pénurie de RAM durerait au moins jusqu'en 2029, car la capacité mondiale de production de RAM en 2028, qui n'avait même pas encore été fabriquée, était déjà vendue. Les prix de la DDR5 avaient plus que quadruplé depuis septembre 2025, ce qui suggérait que les consommateurs pouvaient faire face à des prix encore plus insensés d'ici à 2030.

OpenAI est une société spécialisé d'intérêt public (PBC) américaine spécialisée dans l'intelligence artificielle (IA), dont le siège social est situé à San Francisco, en Californie. Elle développe des modèles d'IA générative *« de pointe »*, notamment la série de grands modèles de langage « Generative Pre-trained Transformer » (GPT). Le lancement de ChatGPT en novembre 2022 a accéléré l'essor de l'IA, contribuant à une utilisation généralisée de l'IA générative ; en septembre 2026, ChatGPT est le cinquième site web le plus visité au monde.peNAI commercialise également les modèles GPT intégré et Codex, un agent de codage basé sur l'IA.

Samuel Harris Altman est un entrepreneur et investisseur américain qui occupe depuis 2019 le poste de directeur général (PDG) de la société d'IA OpenAI. Auparavant, il a été président de l'accélérateur de start-up Y Combinator pendant ses études et de cofonder Loopt, une application de réseau géolocalisé pour smartphones. Loopt a été rachetée par Green Dot Corporation pour 43,4 millions de dollars en mars 2012. En 2015, Altman a rejoint Y Combinator, un accélérateur de start-ups technologiques et une société de capital-risque, dont il a été le président de 2014 à 2019. Il est milliardaire grâce à ses investissements dans plusieurs entreprises.

Face à la pénurie mondiale de mémoire vive (RAM) qui perdure, Sam Altman, PDG d'OpenAI, affiche un point de vue plutôt inhabituel. Il estime que le fait que les particuliers disposent de grandes quantités de mémoire vive sur leurs ordinateurs personnels, où cela ci reste souvent inutilisé la majeure partie du temps, ne constitue pas la meilleure façon d'exploiter les ressources limitées disponibles à l'échelle mondiale. Il soutient au contraire que cette mémoire serait mieux utilisée dans les centres de données cloud, où elle pourrait servir à de nombreux utilisateurs.

Altman a fait ces remarques lors d'une intervention dans le podcast « *Cheeky Pint* », où on l'a interrogé sur les contraintes liées à la chaîne d'approvisionnement et sur la demande croissante en mémoire. Evoquant la tendance à exécuter des modèles d'IA en local, plutôt que d'y accéder via des services cloud, il s'est demandé si cette approche était judicieuse alors que les stocks de mémoire sont déjà limités.



Watch on

« Beaucoup de gens veulent une inférence locale », a déclaré Altman, faisant référence au processus consistant à exécuter un modèle d'IA sur un ordinateur personnel ou un autre appareil local. Il a fait valoir que les modèles open source ont incité les gens à écrire plus efficacement l'IA sur leurs propres machines et réduire ainsi la besoin de grands centres de données. *« Mais d'un point de vue de répartition mondiale, c'est en réalité assez regrettable que chacun dispose d'une partie de l'offre mondiale limitée de DRAM dans son propre ordinateur, alors que celle-ci n'est utilisée que 5 % du temps »*, a-t-il déclaré.

Les raisons pour lesquelles Sam Altman souhaite davantage de RAM dans les centres de données

L'argument d'Altman porte principalement sur le taux d'utilisation. Un ordinateur personnel peut disposer de 16 Go, 32 Go, voire davantage de mémoire vive, mais il n'utilise pas nécessairement toute cette mémoire tout au long de la journée. Dans un centre de données cloud, les ressources peuvent être mises en commun et allouées à différents utilisateurs et charges de travail en fonction de l'évolution de la demande.

« On s'inscrit à optimiser l'utilisation dans le cloud lorsqu'on doit composer avec un environnement où l'offre est très limitée », a déclaré Altman. Ses propos interviennent à un moment où les entreprises spécialisées dans l'IA argumentent depuis longtemps en faveur de la disparition de la puces mémoire. OpenAI veut elle-même engager dans d'importants projets d'infrastructure. En octobre 2025, l'entreprise a annoncé des partenariats avec Samsung et SK Hynix afin de développer jusqu'à 900 000 plaquettes de DRAM par mois dans le cadre de ce qu'elle appelle Stargate.

Les entreprises ont indiqué qu'elles souhaitaient une capacité de production pouvant atteindre 900 000 plaquettes de DRAM par mois dans le cadre de ce développement.

Les fabricants d'électronique grand public ressentent déjà cette pression. TrendForce prévoit une hausse de 10 à 15 % des prix contractuels des DRAM classiques au quatrième trimestre 2026. De son côté, IDC estime que la pénurie de mémoire pourrait entraîner une baisse de 16,7 % des livraisons mondiales de smartphones cette année, tandis que les prix de vente moyens augmenteraient de 27,6 % pour atteindre 581 dollars.

Commente le MONDE entier s'est fait avoir sur la RAM (e

Xavier et moi PC

Watch on

Cependant, il existe un autre aspect à prendre en compte. L'exécution locale de l'IA peut offrir aux utilisateurs davantage de confidentialité, un accès hors ligne et un meilleur contrôle sur leurs données. De plus, la mémoire utilisée dans les PC grand public ne peut pas être simplement retirée et transférée vers un serveur d'IA. Différentes charges de travail nécessitent différents types de mémoire, même si les fabricants réorientent leurs capacités de production vers les serveurs et les produits d'IA, pour lesquels la demande est plus forte.

Au début d'octobre 2026, une sortie de Sam Altman, PDG d'OpenAI, mettait en lumière les désaccords entre les entreprises concernant la régulation de l'IA. Alors que certains concurrents préconisaient des mesures strictes pour éviter des scénarios catastrophes, Sam Altman soutenait qu'il fallait tolérer « certaines choses négatives » afin de maximiser les avantages globaux de la technologie. Cette déclaration intervenait peu après qu'il avait qualifié les inquiétudes concernant la consommation d'eau des centres de données d'IA de « *même sans fondement* » et traitait les humains « *d'ordinateurs de viande inefficaces* ». Il prêchait la régulation de l'IA, tout en la combattant en coulisses.

Et vous ?

➜ Pensez-vous que cette information est crédible ou pertinente ?
➜ Quel est votre avis sur le sujet ?

Voir aussi :

➜ L'IA assèche le marché du matériel informatique : quand RAM, GPU et stockage deviennent des ressources sous tension permanente. Les prix des disques durs ont augmenté en moyenne de 46 % depuis septembre

➜ Les prix de la RAM ont explosé en raison de la bulle IA », certaines barrettes DDR5 trimuant de prix de 500 % un kit de 128 Go se vendant désormais à 3 999 $, soit dix fois plus qu'avant

➜ La pénurie de RAM s'étend : Micron annonce la fermeture de sa marque Crucial pour se concentrer sur les clients professionnels, alors que la demande en mémoire explose

Publication de communiqués de presse en informatique. Contribuez au club : corrections, suggestions, critiques, ... **Contactez le service news** et **Rédigez des actualités**

Répondre avec citation 7 | 0

Aujourd'hui, 17h30 #164

kmedghalth ♀
Membre confirmé

Inscrit en: Décembre 2009
Messages: 177

Il estime aussi que votre argent n'a rien à faire sur votre compte en banque et que ca devrait financer l'IA et son yacht.

Répondre avec citation 4 | 0

Aujourd'hui, 18h27 #165

NotAbroad ♀
Membre confirmé

Développeur informatique
Inscrit en: Août 2024
Messages: 85

Ironiquement, prendre toute la RAM des particuliers, et donc les priver de PC/smartphone, détruirait en grande parti la base d'utilisateur d'OpenAI. Comment ferait-il tourner ses datacenters à 100% h24 si personne ne peut utiliser ChatGPT ?

Répondre avec citation 0 | 0

Aujourd'hui, 19h30 #192

Artemus24 ♀
Membre prolifique

Agent secret au service du gouvernement fédéral
Inscrit en: Février 2011
Messages: 7 681

Comment Sam Altman peut il dire cela ?
Est-il complètement idiot ou quoi ?
Je ne savais pas qu'un ordinateur pouvait fonctionner sans RAM.
Et la prochaine connerie, ne serait elle pas sur les disques ?
Ce sont des gens qui croient que l'économie fonctionne sans acheteur.

Répondre avec citation 1 | 0

« Répondre à la discussion | Signaler un problème

Page 9 sur 9 | ◂◂ Première | ◂ | … | 5 | 6 | 7 | 8 | 9

Forum | Général Développement | Algorithme & Mathématiques | Intelligence artificielle

Les dirigeants anticipent sur une extension de la pénurie de RAM jusqu'en 2028 vue la demande liée à l'IA

« Discussion précédente | Discussion suivante »

Discussions similaires

Le PDG de Google DeepMind, estime que les dipômes en informatique spécialisés dans l'IA resteront importants Par Etora dans le forum Intelligence artificielle	Réponses: 0 Dernier message: 21/07/2026, 14h14
USA : une cour d'appel estime que les clients des entreprises ne doivent pas nécessairement être informés Par Annonces dans le forum Actualités	Réponses: 0 Dernier message: 19/07/2017, 19h54
Facebook estime que les incidents liés à l'lash pourraient avoir un impact sur son activité Par Annonces dans le forum Flash	Réponses: 0 Dernier message: 24/06/2015, 10h38
Les prix des applications flambent sur l'App Store de 14 % en 2011 d'après les estimations de Piper Jaffray Par ibeliveau dans le forum Apple	Réponses: 1 Dernier message: 12/07/2011, 11h53
Les prix des applications flambent sur l'App Store de 14 % en 2011 d'après les estimations de Piper Jaffray Par ibeliveau dans le forum Actualités	Réponses: 0 Dernier message: 11/07/2011, 13h34

Partager

Nous contacter Developpez.com Haut de page

© 2000-2026 - www.developpez.com