

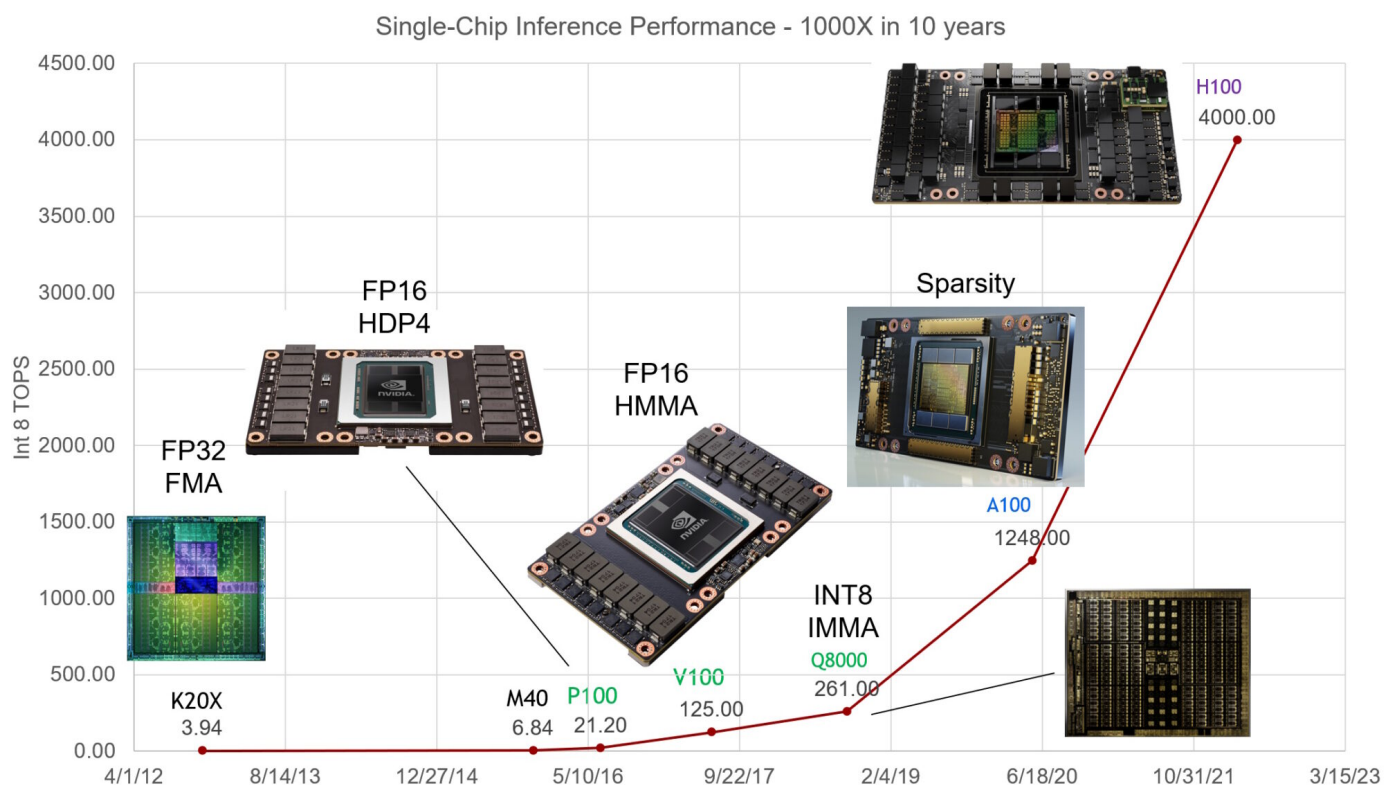
Veille Technologique sur l'intégration de l'IA dans l'informatique au quotidien

Depuis le début des années 2010, l'intégration de l'intelligence artificielle (IA) dans les composants informatiques tels que les CPU, GPU et autres processeurs spécialisés a connu une évolution significative. NVIDIA, en particulier, a été à l'avant-garde de ces innovations, transformant ses technologies pour répondre aux besoins croissants de l'IA. Voici un aperçu chronologique des principales avancées de NVIDIA dans ce domaine.

Début des années 2010 : Adoption des GPU pour l'IA

Au début des années 2010, les processeurs graphiques (GPU) de NVIDIA, initialement conçus pour le rendu graphique, ont été adoptés pour l'entraînement de modèles d'apprentissage profond en raison de leur capacité à effectuer des calculs parallèles massifs. Cette transition a été facilitée par la plateforme CUDA de NVIDIA, permettant aux développeurs de tirer parti de la puissance de calcul des GPU pour des applications d'IA.

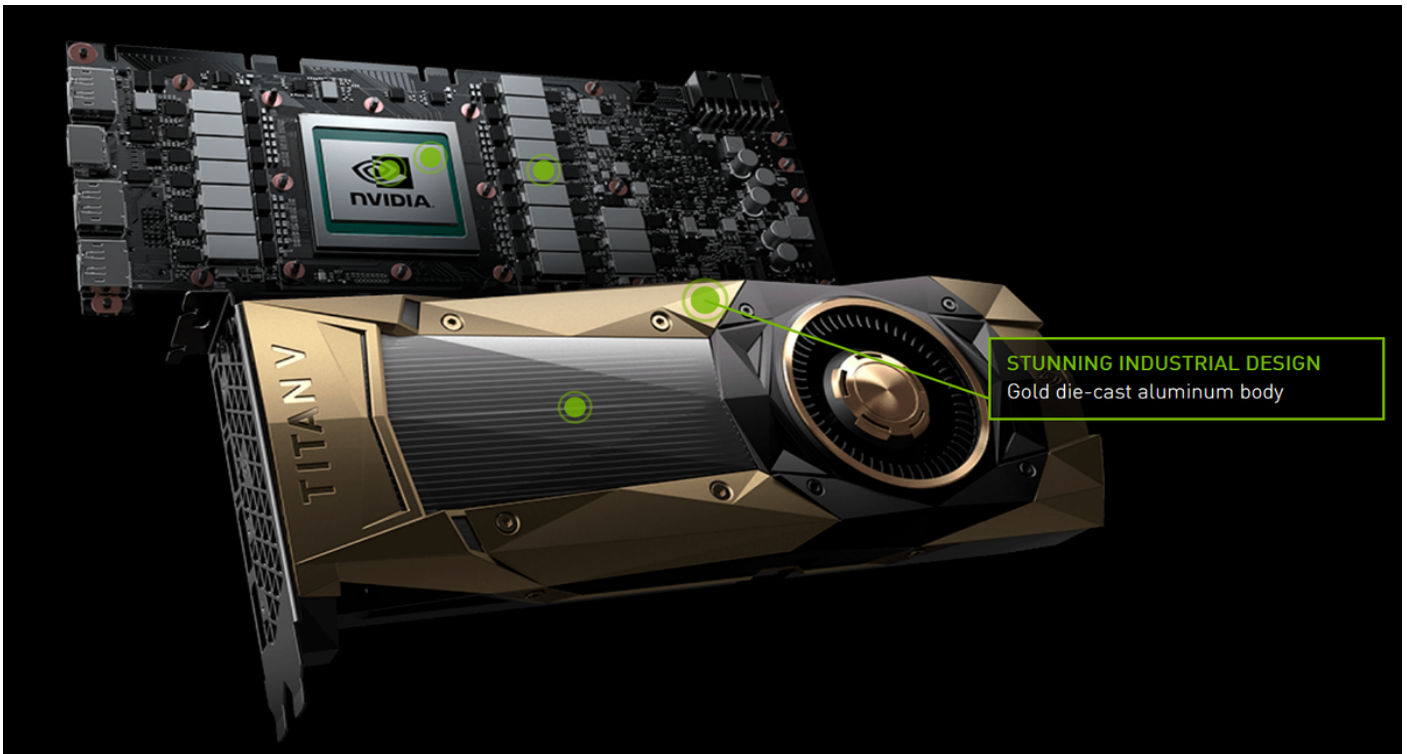
Source : <https://medium.com/%40ArunPrakashAsokan/nvidias-ai-revolution-story-behind-how-it-became-a-trillion-dollar-company-620e7235fdc>



2017 : Introduction de l'architecture Volta par NVIDIA

En 2017, NVIDIA a dévoilé l'architecture Volta, introduisant des cœurs Tensor dédiés pour accélérer les calculs liés à l'IA. La carte graphique NVIDIA Titan V, basée sur cette architecture, a offert des performances accrues pour les tâches d'apprentissage profond.

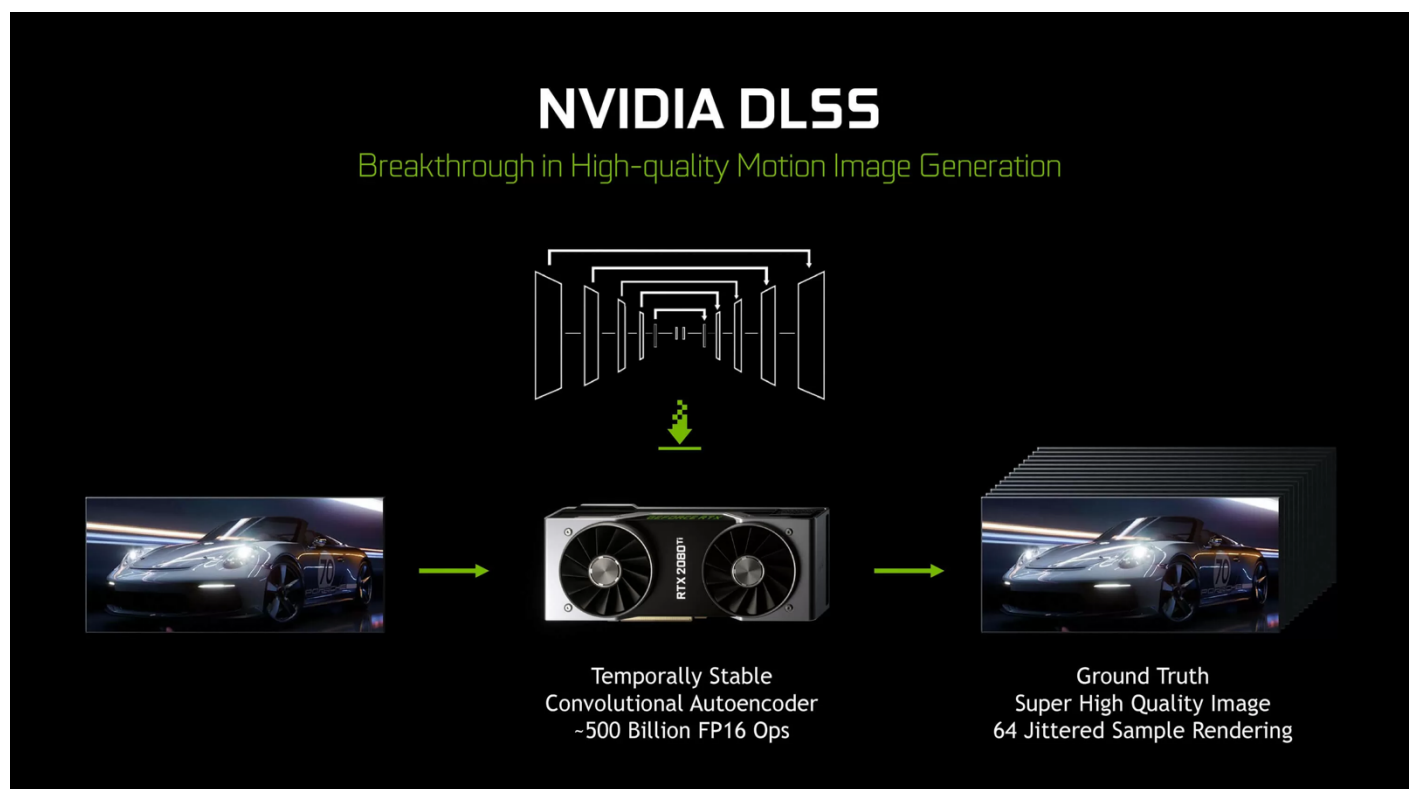
Source : <https://nvidianews.nvidia.com/news/nvidia-titan-v-transforms-the-pc-into-ai-supercomputer>



2018 : Introduction du Deep Learning Super Sampling (DLSS)

En 2018, NVIDIA a lancé le Deep Learning Super Sampling (DLSS) avec ses GPU GeForce RTX série 20. Le DLSS est une technologie d'IA qui améliore la qualité d'image et les performances dans les jeux vidéo en générant des images haute résolution à partir de résolutions inférieures, réduisant ainsi la charge sur le GPU tout en offrant une expérience visuelle supérieure.

Source : [Tech_sama - Tech_Sama Show #70 : RTX 20x0, SteamPlay](#)

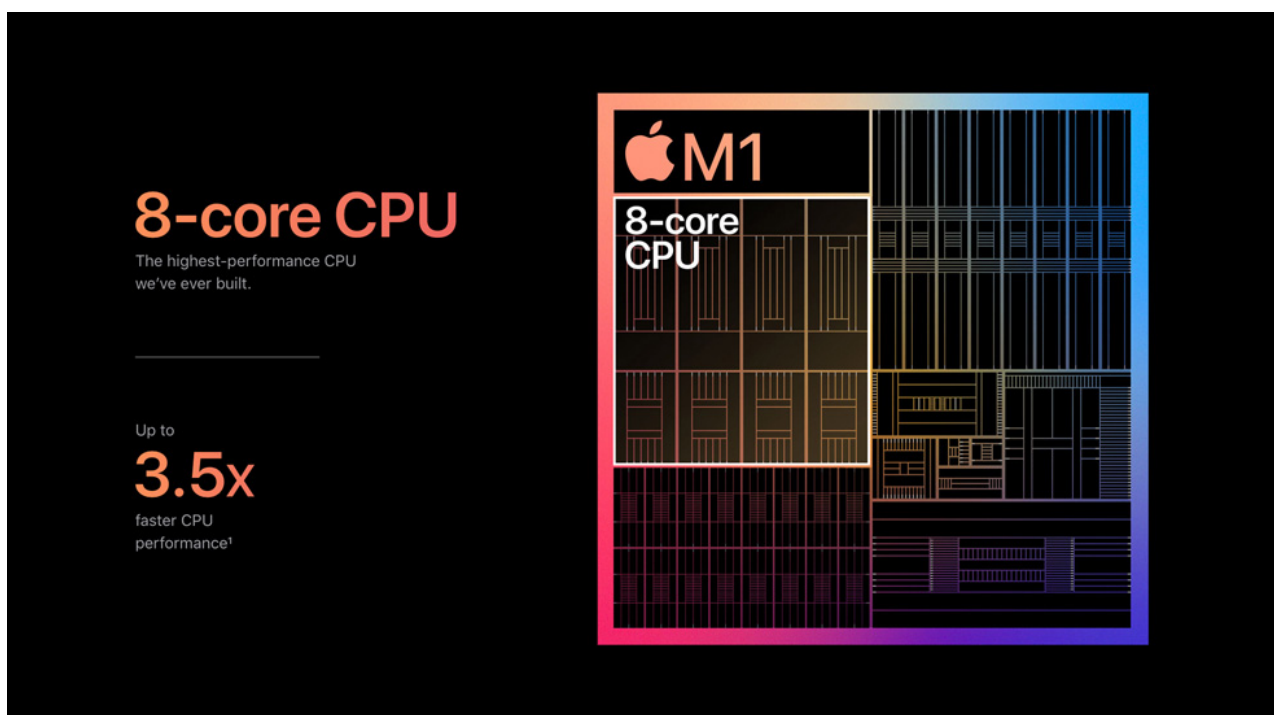


2020 : Le début de la fin pour les processeurs x86 initié par Apple

En novembre 2020, Apple a introduit la puce **M1**, amorçant la transition des Mac de l'architecture **x86** d'Intel vers sa propre architecture **ARM**, connue sous le nom d'Apple Silicon. Cette évolution a permis une amélioration significative des performances et de l'efficacité énergétique des Mac. L'architecture ARM, avec son jeu d'instructions RISC, favorise des instructions plus simples et une exécution plus rapide, contrairement à l'architecture x86, qui utilise un jeu d'instructions CISC plus complexe et gourmand en énergie.

Un atout majeur de la puce M1 est l'intégration d'un **Neural Engine** à 16 cœurs, capable de traiter jusqu'à 11 000 milliards d'opérations par seconde. Cette capacité permet d'exécuter localement des tâches d'intelligence artificielle (IA) telles que la reconnaissance vocale, l'analyse d'images et le traitement du langage naturel. Cela améliore la réactivité des applications et renforce la confidentialité des données en évitant leur transfert vers des serveurs externes. De plus, le traitement local des tâches d'IA réduit la dépendance à une connexion Internet, offrant une utilisation plus fluide et continue des fonctionnalités basées sur l'IA.

Source : [Tech Sama Show #172 : 3080 ti?! Apple M1, Space X Crew 1](#)



2021 : L'IA dans les smartphones via le Qualcomm Snapdragon AI Engine



Source : <https://cambrian-ai.com/qualcomm-ups-the-snapdragon-ai-game/>

2022 : L'IA se perfectionne dans les SoC grâce au MediaTek APU

En 2022, MediaTek a intensifié l'intégration de l'intelligence artificielle au sein de ses System on Chip (SoC) via son APU (Unité de Traitement IA). Cette technologie, présente dans les dernières séries de chipsets destinés aux smartphones grand public, permet d'exécuter des algorithmes d'IA directement sur l'appareil. Cela se traduit par une amélioration notable dans des domaines tels que l'optimisation de la qualité d'image, la reconnaissance de scènes et l'efficacité énergétique, offrant ainsi une performance accrue et une réactivité améliorée pour les utilisateurs.

Source : <https://www.lesnumeriques.com/telephone-portable/mwc-2022-dimensity-8000-et-8100-mediatek-elargit-sa-gamme-de-soc-premium-n177637.html>

Incredible Performance - 2nd Generation TSMC 4nm-class

Packed with industry firsts, this chip powers a new era of flagship smartphones.

- 1st Arm Cortex-X3 and Cortex-A715
- 1st LPDDR5X-8533, with 13% more bandwidth
- 1st UFS 4.0 + MCQ - fastest-ever smartphone storage
- TSMC N4P + MediaTek-optimized IC design

Cameras & Imaging - MediaTek Imagiq 890

Capture brighter, sharper images with the industry's leading AI-ISP architecture.

- 1st RGBW sensor native support for 30% brighter captures
- Second-Gen AI-Video Engine with APU+ISP fusion
- Capture and stream HDR video from multiple cameras
- Up to 12.5% power savings for 8K30 video with EIS
- Fastest AI-NR photo capture

AI - MediaTek APU 690

Enjoy better AI performance with the new 6th-Gen AI featuring xTreme Power Saving Technology.

- 35% higher performance (ETHZ 5.0)
- 25% lower power 4K AI-NR video vs 5th Gen APU
- 45% improved video SR efficiency
- Upgraded DLA processor with mixed precision mode
- Upgraded shared memory efficiency

Gaming - MediaTek HyperEngine 6.0

Hardware-based raytracing and exclusive gaming technologies ensure incredible visuals with smooth high frame-rate gameplay.

- 1st Arm Immortalis-G715 GPU with hardware raytracing
- MediaTek Adaptive Game Technology
- Motion Blur Reduction
- Frame Rate Smoother 2.0

Connect Everywhere

Never skip a beat with wireless audio for premium earbuds, the fastest Wi-Fi, and the most reliable 5G connectivity.

Wi-Fi 7

- 1st Wi-Fi 7 ready smartphone chip with up to 6.5Gbps download
- MediaTek Wi-Fi UltraSave provides up to 70% power savings

Bluetooth 5.3

- 1st Studio-grade Bluetooth Audio (24-bit / 192kHz)
- LE Audio Auracast with hearing aid support

5G

- Sub-6GHz + mmWave ready
- 4CC-CA sub-6GHz 5G R16 modem
- MediaTek 5G UltraSave 3.0
- Industry-leading Multimode Dual SIM Dual Active

Display - MediaTek MiraVision 890

We added the most impressive display technologies, including extreme performance gaming displays to ensure movies and videos always show their best.

- MediaTek Intelligent Display Sync 3.0
- Motion Blur Reduction
- EnergySmart Screen 2.0
- MediaTek AI-SR/MEMC video

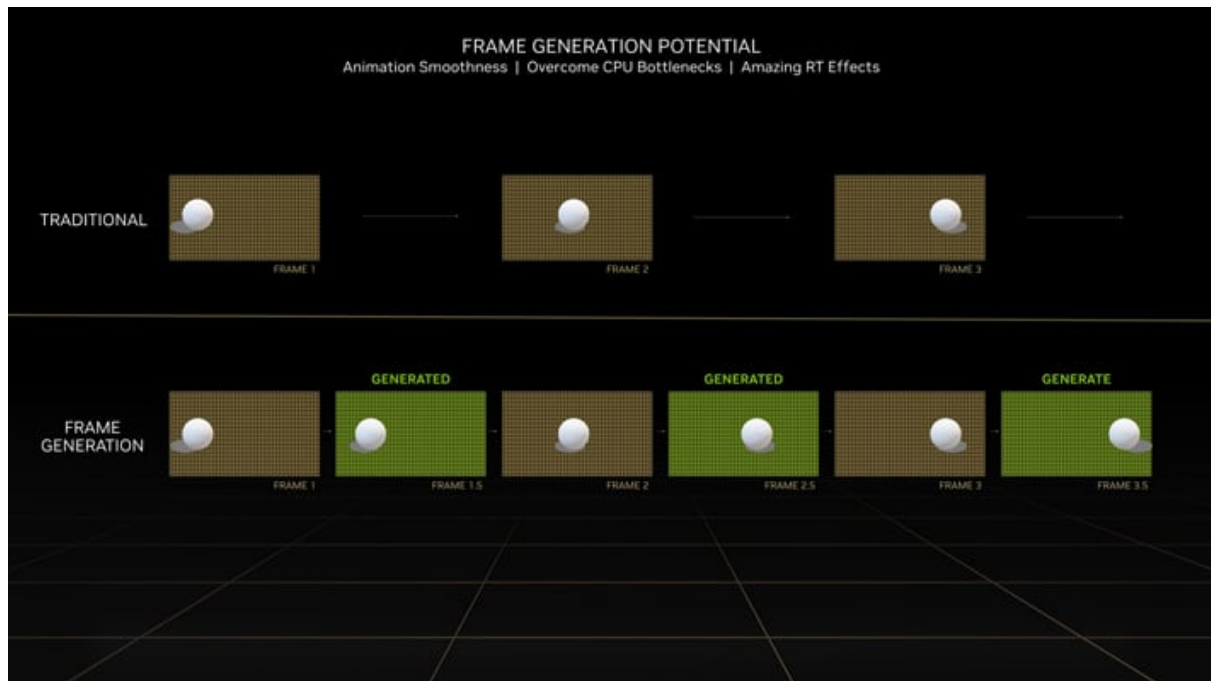
5G

MediaTek Dimensity 9200

2023 : Lancement du DLSS 3.0 et présentation de l'architecture Ada Lovelace

En 2023, NVIDIA a lancé DLSS 3.0, introduisant la génération d'images par IA « Frame Generation » ayant pour optique d'améliorer les performances et de potentiellement améliorer la qualité visuelle dans les jeux. Cette version a permis de générer des images supplémentaires en utilisant des modèles d'IA avancés, offrant ainsi une expérience de jeu plus fluide et plus immersive. La même année, l'architecture Ada Lovelace a été présentée, apportant des améliorations supplémentaires aux performances GPU et aux capacités d'IA.

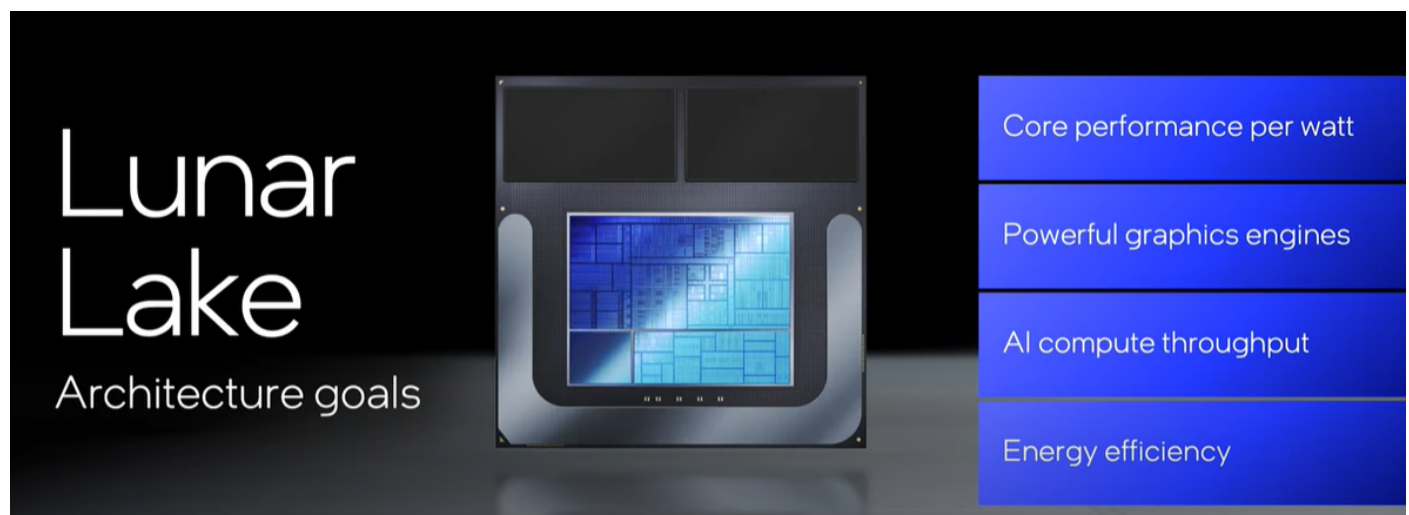
Source : [Linus Tech Tips - Double your FPS for FREE!*** - Nvidia DLSS 3.0](#)



2024 : Dévoilement de l'architecture Lunar Lake par Intel

En 2024, Intel a annoncé Lunar Lake, une architecture de processeur pour ordinateurs portables intégrant des capacités d'IA avancées. Lunar Lake a été conçu pour exécuter des opérations d'IA localement, offrant des performances accrues pour les applications d'IA sur les appareils portables.

Source : <https://www.comptoir-hardware.com/actus/processeurs/47719-lunar-laken-quand-intel-promet-la-lune.html>



2025 : Annonce des nouveaux GPU Blackwell et RDNA 4

Les GPU GeForce RTX 50, basés sur l'architecture Blackwell, apportent des avancées majeures en IA. Avec DLSS 4, de nouveaux cœurs Tensor et les "Neural Shaders", ils améliorent la fluidité des jeux, l'inférence IA et les graphismes en temps réel. Ces innovations font de la série RTX 50 un choix idéal pour les gamers et créateurs recherchant performances et capacités d'IA accrues.

De son côté, AMD a présenté au CES 2025 ses nouveaux GPU Radeon RX 9070 et 9070 XT, basés sur l'architecture RDNA 4. Ces modèles, attendus au premier trimestre 2025, intègrent des accélérateurs d'IA de deuxième génération et un moteur de ray-tracing amélioré. Ils introduisent également le FidelityFX Super Resolution 4, une technologie d'upscaling basée sur l'intelligence artificielle, visant à offrir une qualité d'image supérieure, notamment en 4K.

Ces annonces illustrent les efforts continus d'AMD et de NVIDIA pour dominer le marché des GPU dédiés à l'intelligence artificielle et au ray-tracing.

Source : <https://iphonesoft.fr/2025/01/07/ces-2025-amd-nouvelle-architecture-rdna4-upscaling-fsr-4>

INTRODUCING

AMD RADEON™ RX 9070 XT
AMD RADEON™ RX 9070

Available
Q1 2025

Logos des partenaires : acer, ASUS, SAPHIRE, XFX, ASRock, GIGABYTE, PowerColor, 華強電子, YESTON 盈通

The image displays a collection of AMD Radeon RX 9070 and RX 9070 XT graphics cards from various manufacturers, arranged on a dark surface. The cards feature different cooling solutions, including dual and triple fans with RGB lighting. The AMD logo is visible on the side of each card.

RTX 5090 3,400 AI TOPS \$1,999	RTX 5080 1,800 AI TOPS \$999	RTX 5070 Ti 1,400 AI TOPS \$749	RTX 5070 1,000 AI TOPS \$549
--	--	---	--

Availability Starting January

The image shows a close-up of the NVIDIA GeForce RTX 5090 graphics card. It is a black, rectangular card with a prominent silver-colored metal shroud. The shroud features a large, stylized 'X' logo in the center. The text 'GEFORCE RTX' is visible on the right side of the shroud.

Mais à quoi ça sert tout ça ?

1. Statistiques et performances :

- Des rapports récents indiquent que les performances des GPU ont augmenté de plus de 40% en moins d'une décennie, tandis que l'efficacité énergétique des nouvelles architectures a permis une réduction de la consommation électrique jusqu'à 30 % par rapport aux générations précédentes.
- Une entreprise spécialisée dans la domotique a récemment intégré des solutions basées sur des processeurs équipés de NPU. Résultat : une amélioration de 25 % de la réactivité de leurs systèmes de contrôle, ainsi qu'une réduction significative des coûts de traitement des données, permettant une gestion plus fluide des dispositifs connectés.

2. Avantages et Limites dans le contexte professionnel

Bénéfices :

- **Amélioration de la performance et de la réactivité :**
L'adoption des GPU et des NPU a permis d'accélérer les calculs et de réduire le temps de réponse des systèmes informatiques, un atout majeur pour des applications exigeantes telles que la domotique ou la gestion de réseaux complexes.
- **Optimisation énergétique :**
Les nouvelles architectures offrent une meilleure gestion de l'énergie, réduisant les coûts opérationnels et l'impact environnemental, un enjeu crucial pour les entreprises soucieuses de leur empreinte écologique.

Limites :

- **Risques de sécurité :**
Bien que ces technologies apportent des performances accrues, elles introduisent également de nouveaux vecteurs de piratage. La multiplication des dispositifs connectés nécessite une vigilance renforcée en matière de cybersécurité.
- **Obsolescence rapide et formation continue :**
L'évolution constante des technologies impose aux professionnels de se former régulièrement pour maîtriser ces outils. L'obsolescence des matériels et des logiciels peut également représenter un frein pour certaines organisations, nécessitant une planification stratégique sur le long terme.

3. Impact sur les compétences et les enjeux métiers

- Les innovations technologiques analysées ici influencent directement la gestion des infrastructures IT. Par exemple, la mise en œuvre d'un système de domotique optimisé grâce à l'IA demande aux techniciens de comprendre et d'intégrer ces nouvelles architectures dans leur environnement de travail.
- La comparaison entre différentes solutions (GPU vs NPU, architectures traditionnelles vs modernes) permet de mieux orienter les choix stratégiques des entreprises, en prenant en compte des critères de performance, de coût et de pérennité.

Conclusion

L'intelligence artificielle s'impose de plus en plus dans l'informatique, optimisant les performances des CPU et GPU tout en réduisant leur consommation énergétique. Des technologies comme les NPU et les architectures dédiées permettent d'accélérer des tâches complexes, rendant l'IA plus accessible et efficace.

Cependant, cette évolution rapide soulève des défis : l'obsolescence du matériel, les risques de cybersécurité et la nécessité d'une veille constante pour suivre les innovations. Alors que les avancées en IA se multiplient, leur intégration devra être maîtrisée pour garantir un équilibre entre performance, sécurité et durabilité.

Outils utilisés pour construire ma veille :

- Feedly
- ChatGPT Research
- YouTube, Twitter/X (ex : LinusTechTips & Tech_Sama)
- Connaissances personnelles