

# Dossier de veille informationnelle sur les cartes graphique

Valentin Benard



## INTRODUCTION

Le sujet des cartes graphiques est incontournable car elles sont présentes dans notre quotidien, c'est à dire tous les ordinateurs et même nos smartphones et montres connectées et plus encore, via des micro puces appelées GPU (Graphics Processing Unit), **elles constituent également le futur avec l'IA qui dépend de la puissance de calcul des cartes graphique en grande partie.**

Elles sont utilisées pour le rendu 3D des jeux vidéo, les logiciels professionnels de modélisation, de conception 3D et d'AutoCAD, mais également, de plus en plus, pour l'Intelligence Artificielle et le deep learning.

Elles sont également utilisées pour le mining de crypto-monnaies comme l'Ethereum et le Bitcoin qui constituent des monnaies alternatives mais qui ont contribué plus tard à une pénurie à grande échelle.

**Le marché est composé de trois acteurs principaux:**



**NVIDIA:** Leader Historique qui est célèbre pour ses gammes GeForce, Quadro et Tesla.



**AMD:** Concurrent majeur avec ses séries Radeon et son rapport performance / prix.



**Intel:** Initialement présent uniquement avec les iGPUs présents dans ses CPUs, Intel est récemment entré sur le marché avec sa série Arc qui cible principalement les jeux vidéo et la création.

Les années récentes ont vu la rapide évolution de l'IA nécessitant des cartes graphiques dédiées à ce type de calcul, en plus de la tendance grandissante du Ray Tracing dans le domaine des jeux vidéo avec la série RTX de Nvidia.

L'IA est également de plus en plus utilisée dans le jeu vidéo pour des améliorations telles que DLSS qui sert, depuis une image de plus faible résolution, à l'améliorer en une résolution supérieure grâce à l'IA et aux cœurs Tensor des RTX.

# 1) Histoire et évolution

## 1. Les débuts des cartes graphiques

Dans les années 1980-1990, les cartes graphiques servaient principalement à afficher des images en 2D. Les premiers modèles étaient intégrés aux cartes mères, sans véritable puissance de calcul. Avec l'arrivée des jeux 3D dans les années 1990 (comme Doom), des cartes dédiées ont vu le jour. NVIDIA et ATI (aujourd'hui AMD) ont continué les avancées permettant de réaliser des graphismes toujours plus avancés.



### Chronologie pré années 2000:

- **1968** : Fondation d'**Intel** le 18 juillet.
- **1969** : Fondation d'**AMD** le 1er mai.
- **1981** : IBM lance la carte CGA (Color Graphics Adapter), capable d'afficher 16 couleurs en mode texte et 4 couleurs en mode graphique.
- **1987** : Introduction de la carte VGA (Video Graphics Array) par IBM, offrant une résolution de 640x480 pixels avec 16 couleurs.
- **Années 1990** : L'arrivée des cartes 3D marque un tournant majeur. Des entreprises comme 3dfx Interactive introduisent des cartes spécifiques pour le rendu 3D, comme la fameuse Voodoo.
- **1993** : Fondation de **Nvidia** le 5 avril.
- **1999** : NVIDIA révolutionne le secteur avec la première carte graphique utilisant un processeur graphique (GPU), la **GeForce 256**, qui intègre l'accélération hardware pour les calculs 3D, ce qui nous mène à l'ère moderne des GPUs.

## 2. Évolution des cartes graphiques depuis les années 2000

### **2006** : Le rachat d'ATI par AMD

Avant son **rachat par AMD en 2006**, ATI Technologies était l'un des principaux acteurs du marché des cartes graphiques. Fondée en 1985 au Canada, l'entreprise s'était spécialisée dans la conception de processeurs graphiques (GPU) et était particulièrement **réputée pour sa gamme Radeon**, qui **concurrentait directement NVIDIA**. ATI fournissait également des solutions graphiques pour les **consoles de jeux, notamment la Xbox 360, la Nintendo GameCube et la Nintendo DS**.



**En 2006, AMD annonça l'acquisition d'ATI pour environ 5,4 milliards de dollars**, une opération stratégique visant à intégrer les technologies graphiques dans ses processeurs et à renforcer sa position face à Intel. Cependant, ce rachat pesa lourdement sur les finances d'AMD, qui accumula des dettes importantes et traversa des années difficiles avant de rentabiliser cet investissement. Malgré ces défis, l'intégration d'ATI permit à AMD de développer des solutions innovantes, comme les **APU (Accelerated Processing Units), combinant CPU et GPU sur une même puce**. À long terme, cette fusion joua un rôle clé dans le renouveau d'AMD sur le marché des semi-conducteurs.



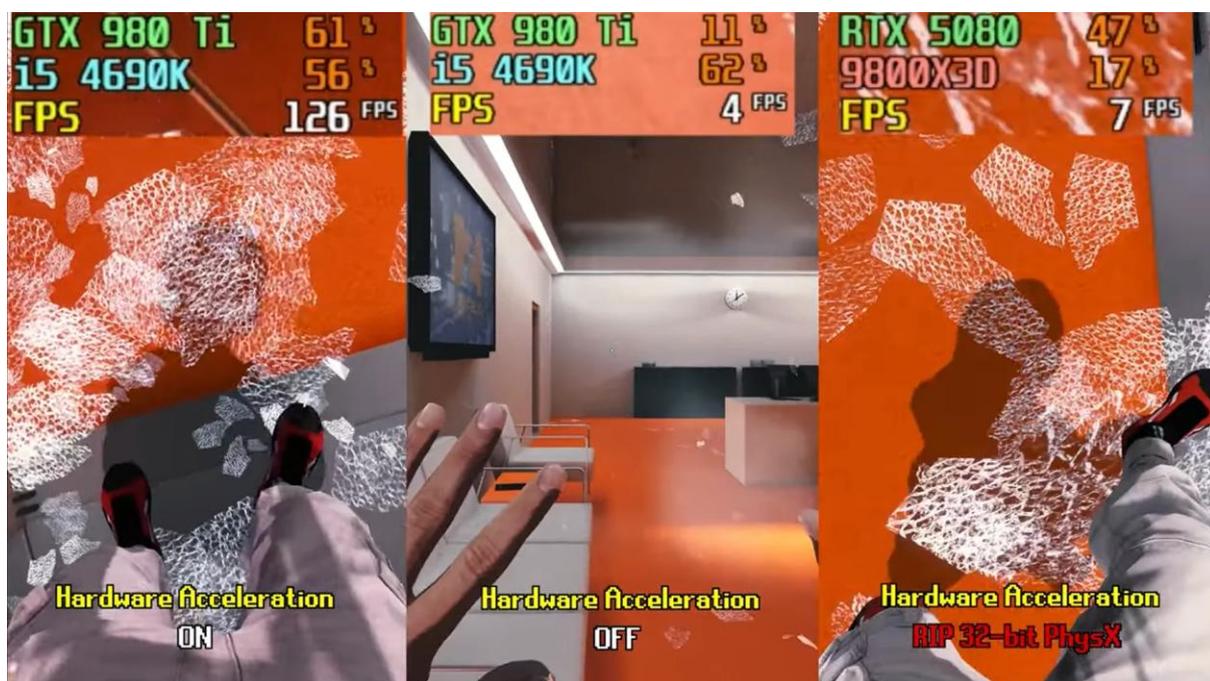
## **2008 : PhysX et les potentiels PPU**

PhysX est un moteur physique développé à l'origine par la société Ageia, fondée en 2002. **Son objectif était d'améliorer la simulation des interactions physiques dans les jeux vidéo** en introduisant un processeur dédié, appelé **PPU (Physics Processing Unit)**. Ce composant matériel permettait d'accélérer des calculs complexes, comme les destructions d'objets, les fluides ou les vêtements dynamiques, améliorant ainsi le réalisme des jeux compatibles.



En 2008, NVIDIA racheta Ageia et intégra la technologie PhysX directement dans ses cartes graphiques. Grâce à cette acquisition, PhysX devint une solution logicielle optimisée pour les GPU NVIDIA, éliminant le besoin d'un PPU dédié. Cette stratégie permit à NVIDIA de proposer des effets physiques avancés dans de nombreux jeux, mais créa aussi une certaine controverse, car PhysX fonctionnait de manière limitée ou était désactivé sur les cartes graphiques concurrentes d'AMD.

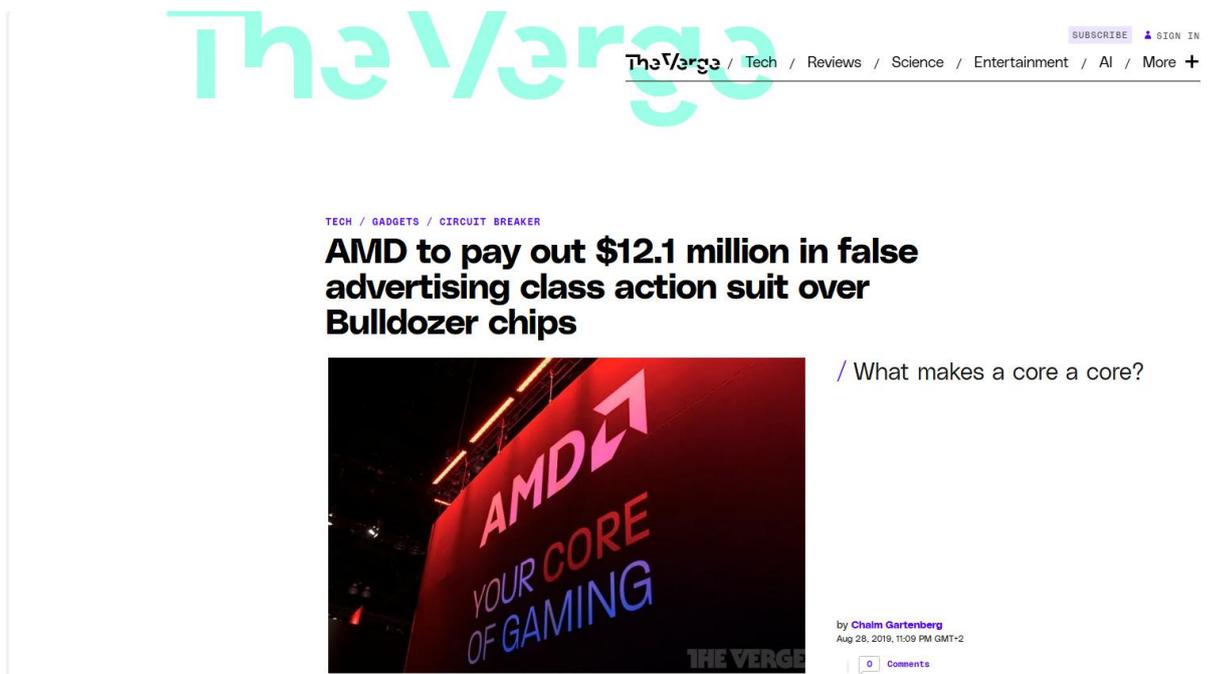
Aujourd'hui, bien que PhysX ait perdu en importance face aux évolutions des moteurs physiques intégrés aux moteurs de jeu comme Unreal Engine ou Unity, il reste utilisé dans certaines applications et continue d'être mis à jour par NVIDIA pour ses solutions de simulation et d'intelligence artificielle, bien que son support de la version 32 bits a été supprimé de la série RTX 5000 récemment, causant des pertes de performance.



## 2014 : La potentielle banqueroute d'AMD et son renouveau

C'est en 2014 que AMD commençait à voir les **prémices d'une banqueroute** suite à une période difficile et un positionnement inférieur à son concurrent, ses parts de marché continuant à chuter, **certaines analystes spéculaient sur une éventuelle faillite ou un rachat par une autre société**. Le manque d'innovation et de compétitivité face à Intel, combiné à une difficulté à s'imposer sur le marché des processeurs mobiles et serveurs, rendaient l'avenir d'AMD incertain.

**C'est dans ce contexte que Lisa Su** (ex employée IBM) fut nommée PDG d'AMD en octobre 2014. Dès son arrivée, elle entreprit une profonde restructuration de l'entreprise. Plutôt que de chercher à concurrencer Intel sur son propre terrain en poursuivant le développement de l'architecture Bulldozer, qui s'était avérée inefficace (en plus de valoir un procès pour "faux coeurs"), **AMD décida de repartir sur de nouvelles bases avec l'architecture Zen**.



The screenshot shows a webpage from The Verge. At the top, the logo "The Verge" is displayed in a light blue font. To the right of the logo, there is a navigation menu with links for "Tech", "Reviews", "Science", "Entertainment", "AI", and "More". Below the logo, the article title "AMD to pay out \$12.1 million in false advertising class action suit over Bulldozer chips" is prominently displayed in bold black text. Underneath the title, there is a sub-header "TECH / GADGETS / CIRCUIT BREAKER". To the right of the title, there is a link "/ What makes a core a core?". Below the title, there is a photograph of a red AMD Bulldozer chip with the text "AMD YOUR CORE OF GAMING" on it. At the bottom right of the article, there is a byline "by Chalm Gartenberg" and a date "Aug 28, 2019, 11:09 PM GMT-2".

Source: ([AMD to pay out \\$12.1 million in false advertising class action suit over Bulldozer chips | The Verge](#))

Ce pari stratégique porta ses fruits avec l'**arrivée des processeurs Ryzen en 2017**. Grâce à cette nouvelle génération de processeurs, AMD réussit à revenir dans la course face à Intel en proposant des CPU compétitifs en termes de performances, d'efficacité énergétique et surtout d'un excellent rapport qualité/prix. Les Ryzen marquèrent un tournant dans l'industrie et ont permis à AMD de regagner des parts de marché significatives, tant dans le domaine des ordinateurs grand public que dans celui des serveurs avec la gamme EPYC.

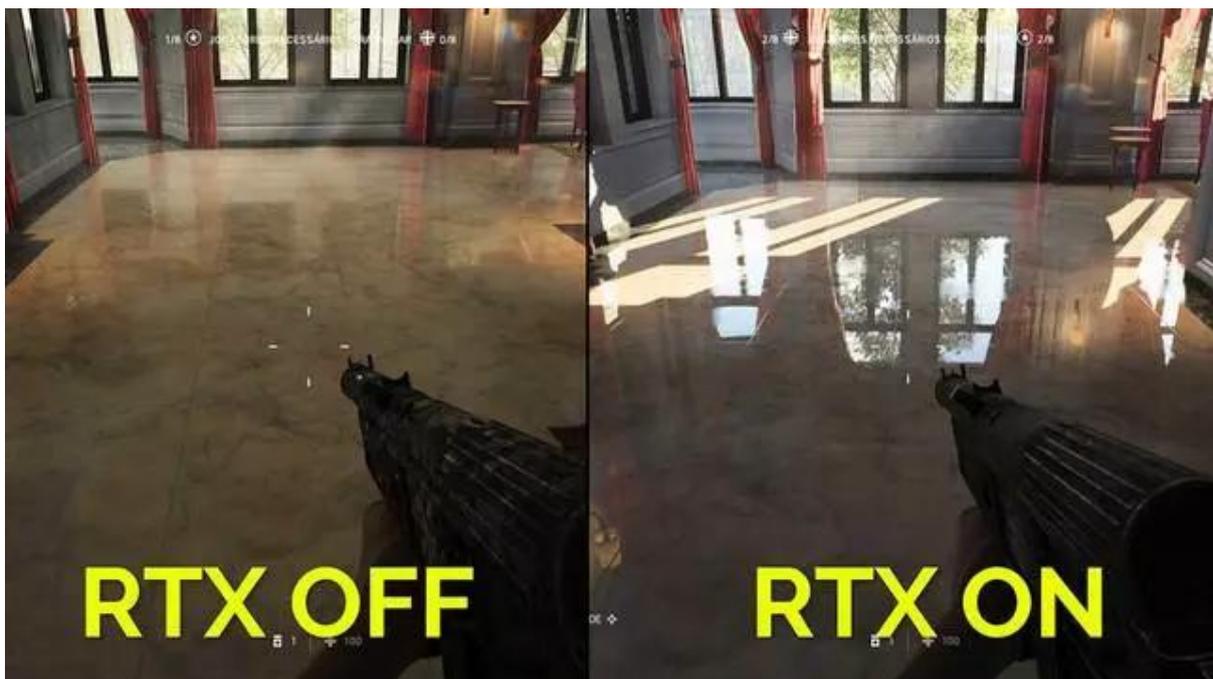


L'arrivée de Lisa Su et la stratégie adoptée ont donc permis un **véritable renouveau pour AMD**. De société en difficulté, elle est devenue un acteur incontournable du marché des semi-conducteurs, rivalisant à nouveau avec Intel et s'imposant même sur certains segments du marché.



## 2018 : Ray Tracing avec les Nvidia RTX 2000

En 2018, NVIDIA révolutionna l'industrie du jeu vidéo en lançant sa gamme de cartes graphiques **GeForce RTX 2000**, introduisant pour la première fois le **ray tracing en temps réel** sur des GPU grand public. Cette technologie, auparavant réservée aux studios d'animation et aux effets spéciaux de films, permet une gestion plus réaliste de la lumière, des ombres et des reflets en simulant le comportement physique des rayons lumineux.



Pour y parvenir, NVIDIA développa une architecture spécifique, nommée **Turing**, qui intégrait des **RT Cores** (cœurs dédiés au ray tracing) et des **Tensor Cores** (accélérateurs d'intelligence artificielle, notamment pour le **DLSS**, une technologie d'upscaling). Ces

innovations permettaient aux jeux compatibles de proposer des graphismes plus immersifs, bien que les premières implémentations du ray tracing entraînaient une forte baisse des performances.

Malgré un accueil mitigé à son lancement en raison du coût élevé des cartes et du nombre limité de jeux compatibles, la gamme RTX 2000 a posé les bases de l'adoption du ray tracing dans l'industrie. NVIDIA a ensuite optimisé cette technologie avec les générations suivantes, rendant le ray tracing plus accessible et performant.

## 2020 : Arrivée de l'IA avec DLSS

En 2020, avec le lancement des cartes graphiques **GeForce RTX 3000**, NVIDIA démocratisa l'utilisation de l'intelligence artificielle dans le domaine du jeu vidéo grâce à l'amélioration du **DLSS (Deep Learning Super Sampling)**. Introduit initialement avec la génération RTX 2000, le **DLSS 2.0**, sorti en 2020, apporta des améliorations majeures en matière de qualité d'image et de performances.

Le **DLSS** repose sur les **Tensor Cores**, des unités de calcul dédiées à l'IA présentes dans les GPU RTX. Son principe est d'utiliser un réseau de neurones entraîné par NVIDIA pour reconstruire une image en haute définition à partir d'une résolution plus basse. Cela permet d'améliorer considérablement les performances en réduisant la charge de calcul, tout en conservant une qualité visuelle proche du rendu natif, voire meilleure dans certains cas.



Cette avancée permet aux jeux exigeants, notamment ceux intégrant le **ray tracing**, d'atteindre des fréquences d'images élevées tout en maintenant une excellente qualité graphique. Le **DLSS** s'imposa rapidement comme une technologie clé pour NVIDIA, et son adoption par les développeurs ne cessa de croître, influençant d'autres solutions concurrentes comme **AMD FSR** et **Intel XeSS**.

## 2020 : Pénurie des cartes graphiques

L'année 2020 fut marquée par une **pénurie mondiale des cartes graphiques**, résultant d'une combinaison de facteurs économiques, technologiques et sanitaires. Tout d'abord, la **pandémie de COVID-19** perturba profondément la chaîne d'approvisionnement mondiale. Les usines de semi-conducteurs, principalement situées à **Taiwan**, tournèrent au ralenti en raison des restrictions sanitaires, tandis que la demande en matériel informatique explosa avec la généralisation du **télétravail** et du **jeu vidéo** à domicile.

Parallèlement, le **boom des cryptomonnaies**, notamment du **Bitcoin et de l'Ethereum**, incita de nombreux mineurs à acheter massivement des GPU, aggravant encore la rareté des cartes graphiques sur le marché. Cette demande excessive provoqua une **inflation des prix** et la revente spéculative par des **scalpers**, rendant les nouvelles cartes NVIDIA RTX 3000 et AMD Radeon RX 6000 quasiment introuvables à leur prix d'origine.



Un "mining rig" massif de 70 Mégawatts.

Enfin, la situation géopolitique autour de **Taïwan**, principal fournisseur mondial de semi-conducteurs via **TSMC**, ajouta une pression supplémentaire. La **crise des composants électroniques** affecta non seulement le marché des GPU, mais aussi celui des processeurs, des consoles de jeu et de nombreux appareils électroniques.

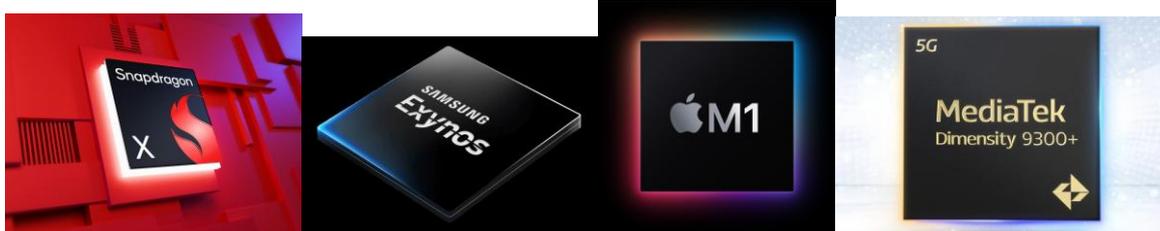


Cette pénurie dura plusieurs années, impactant les consommateurs et les industriels, avant qu'un retour progressif à la normale ne s'amorce en **2022**, avec l'effondrement des cryptomonnaies et l'augmentation de la production de semi-conducteurs.

## 2022 : La potentielle acquisition d'ARM par Nvidia

L'acquisition d'**ARM** par **NVIDIA** fut l'un des plus grands projets de rachat de l'industrie des semi-conducteurs. Annoncée en **2020** pour un montant de **40 milliards de dollars**, cette acquisition visait à faire de **NVIDIA** un acteur dominant non seulement dans le domaine des GPU, mais aussi dans celui des architectures de processeurs, où **ARM** joue un rôle clé.

Fondée en 1990, **ARM** est une entreprise britannique spécialisée dans la conception d'architectures de processeurs à faible consommation, largement utilisées dans les **smartphones, tablettes, objets connectés** et même certains **serveurs**. Contrairement à Intel ou AMD, ARM ne fabrique pas de puces, mais **licencie son architecture** à des entreprises comme **Apple, Qualcomm ou Samsung**.



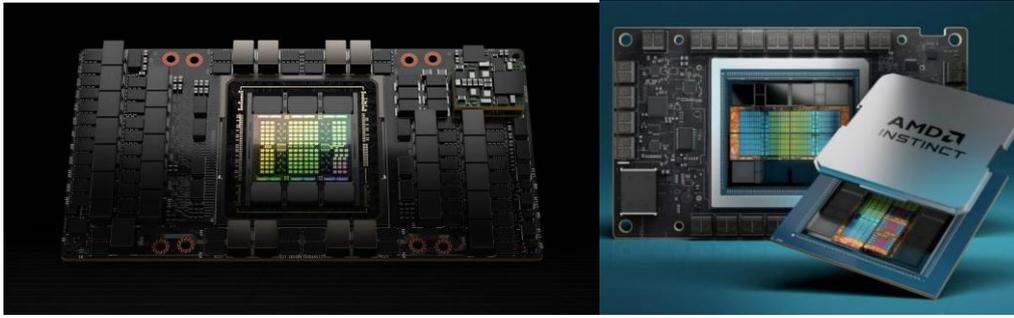
Cependant, ce projet de rachat par **NVIDIA** se heurta rapidement à de nombreux obstacles. Les **autorités de régulation** aux États-Unis, en Europe et en Chine exprimèrent des craintes quant à un **abus de position dominante**, estimant que **NVIDIA** pourrait limiter l'accès aux licences ARM pour ses concurrents. De plus, plusieurs entreprises du secteur, dont **Qualcomm et Microsoft**, s'opposèrent publiquement à l'accord.

Face à cette pression et aux **blocages réglementaires**, **NVIDIA** finit par **abandonner l'acquisition en février 2022**. ARM resta alors sous le contrôle de son propriétaire **SoftBank**, qui annonça par la suite son intention d'introduire ARM en bourse en 2023.

## 2022 : Boom de l'IA

L'année **2022** marqua un tournant majeur dans le domaine de l'**intelligence artificielle**, avec l'explosion des modèles de **génération de texte et d'images**. Le lancement de **ChatGPT** par OpenAI en novembre 2022 démocratisa les IA conversationnelles, montrant au grand public le potentiel des **grands modèles de langage (LLM)**. Face à ce succès, des concurrents comme **Microsoft Copilot, Google Bard (puis Gemini)** ou encore **DeepSeek** se sont rapidement positionnés pour offrir des alternatives basées sur des architectures similaires.

Cette révolution fut rendue possible grâce à des infrastructures massivement **accélérées par GPU**, en particulier les cartes **NVIDIA H100 et A100**, optimisées pour l'IA et l'apprentissage automatique. **NVIDIA** domina largement le marché du **hardware IA**, tandis qu'**AMD** tenta timidement de se positionner avec sa gamme **Instinct MI**, qui peinait encore à rivaliser avec l'écosystème CUDA et TensorRT de **NVIDIA**.



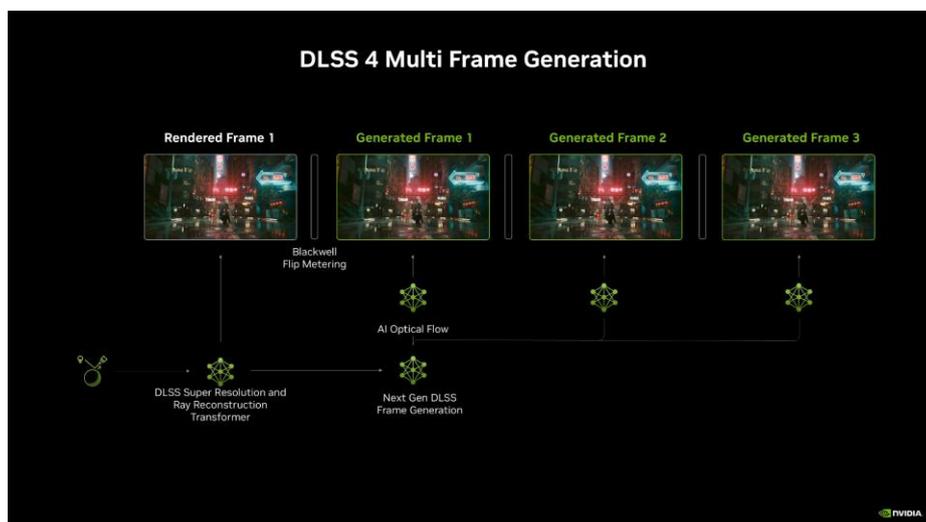
Parallèlement, l'industrie amorça l'intégration de **NPU (Neural Processing Units)** directement dans les processeurs grand public. **Intel**, avec sa gamme **Meteor Lake**, et **AMD**, avec ses processeurs **Ryzen 7000 mobile (Ryzen AI)**, commencèrent à embarquer des unités spécialisées pour accélérer les tâches d'IA locales, comme la reconnaissance vocale ou l'optimisation des performances énergétiques. **Qualcomm**, avec ses puces **Snapdragon X Elite**, se positionna également comme un acteur clé dans cette transition.

Ce boom de l'IA entraîna une **demande massive en puissance de calcul**, générant une course aux **semi-conducteurs IA** et posant de nouveaux défis en matière d'approvisionnement et d'efficacité énergétique.

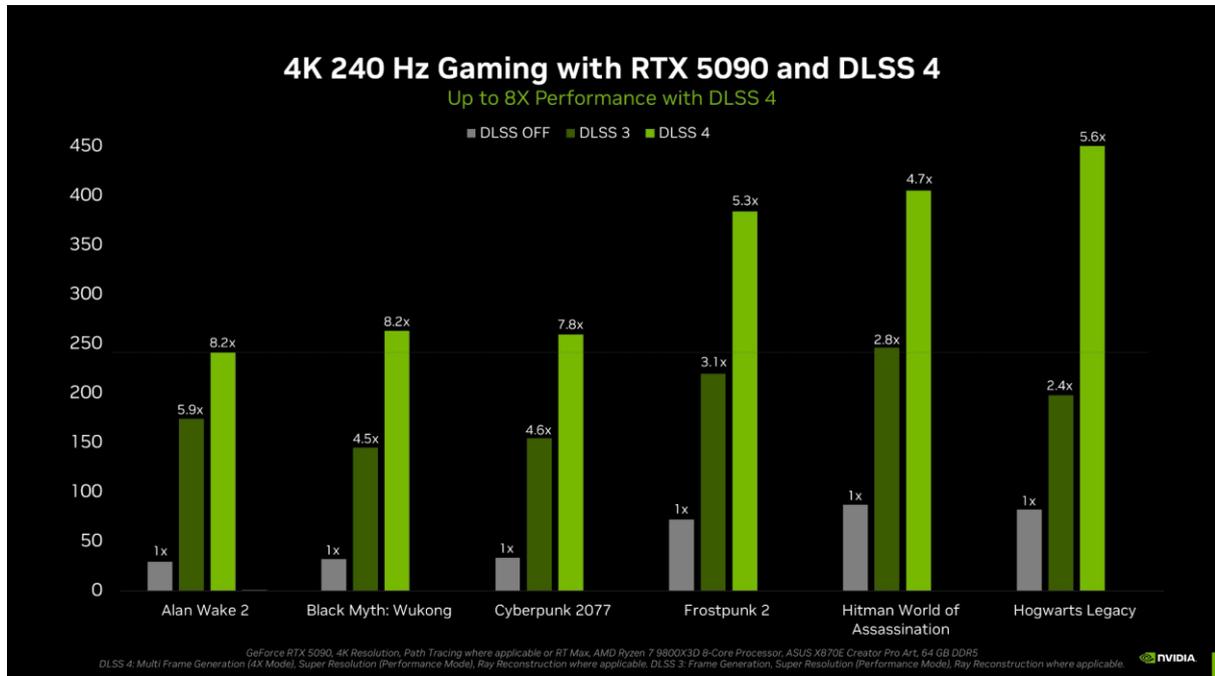
## 2025 : Génération d'images en temps réel avec DLSS 4

En **2025**, NVIDIA franchit une nouvelle étape dans l'utilisation de l'**intelligence artificielle** pour le rendu graphique avec le lancement de **DLSS 4**. Successeur du **DLSS 3**, cette version améliore encore davantage la génération d'images en temps réel en combinant plusieurs avancées technologiques, rendant les jeux plus fluides et détaillés que jamais.

L'une des innovations majeures de **DLSS 4** repose sur une **génération d'images entièrement AI-driven**, où l'algorithme ne se contente plus d'interpoler des images intermédiaires, mais **prédit et reconstruit des scènes complètes** avec un niveau de fidélité inégalé. Grâce aux nouveaux **Tensor Cores** des cartes **RTX 5000**, l'IA analyse les mouvements et les détails de l'environnement en temps réel pour offrir un rendu ultra-réactif, minimisant les artefacts et améliorant la netteté des textures.



Cette évolution permet également de **réduire drastiquement la charge de calcul** sur le GPU, offrant des performances jusqu'alors impossibles sur les jeux les plus exigeants, notamment ceux utilisant le **ray tracing intégral**. NVIDIA pousse ainsi encore plus loin son avantage sur le marché, tandis qu'**AMD (avec FSR 4)** et **Intel (avec XeSS 2.0)** tentent de rivaliser avec leurs propres solutions d'upscaling et de génération d'images.



Avec DLSS 4, l'IA ne se contente plus d'améliorer la fluidité, elle **transforme la manière dont les images sont produites**, rendant possible un futur où le rendu des jeux repose autant sur le **machine learning** que sur la puissance brute des cartes graphiques.

Des solutions alternatives par la communauté ont également émergé avec la sortie de **Lossless scaling**, une application distribuée par Steam visant à générer des images, bien que le résultat reste loin de la solution très performante d'Nvidia, elle reste tout de même intéressante, surtout pour ceux n'ayant pas le matériel nécessaire.

## 2) Le lien entre matériel et logiciel: Les pilotes et les API graphiques

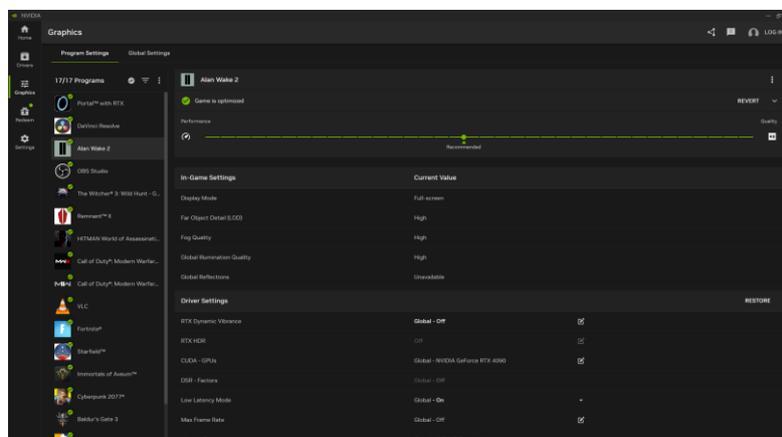
Lorsqu'on joue à un jeu sur PC, il ne suffit pas d'avoir une bonne carte graphique : il faut aussi un **logiciel** pour lui dire comment afficher les images à l'écran. Ce rôle est assuré par les **pilotes graphiques** et les **API graphiques**.

### Les pilotes graphiques : le traducteur entre le jeu et la carte graphique

Un **pilote graphique** (ou driver) est un logiciel fourni par **NVIDIA, AMD ou Intel**, qui permet au système d'exploitation et aux jeux de **communiquer avec la carte graphique**. Sans lui, un jeu ne saurait pas comment exploiter la puissance de la carte.

Les fabricants mettent régulièrement à jour ces pilotes pour **améliorer les performances, corriger des bugs, ajouter des fonctionnalités** et **optimiser les nouveaux jeux**. C'est pourquoi on conseille souvent de garder ses pilotes graphiques à jour.

Il est souvent installé avec un panneau de configuration des pilotes, permettant d'affiner des paramètres du GPU et d'activer / désactiver certaines fonctionnalités.



## Les API graphiques : les outils pour afficher les jeux

Une **API graphique (Application Programming Interface)** est un ensemble de règles et d'outils qui permettent aux développeurs de jeux de dire à la carte graphique **comment dessiner les images**. Il existe plusieurs API graphiques, chacune avec ses avantages :

-  **DirectX** : Développée par **Microsoft**, c'est l'API la plus utilisée sur **Windows**. DirectX 12 permet un meilleur contrôle du matériel pour optimiser les performances.
-  **OpenGL** : Une API ouverte plus ancienne, utilisée sur plusieurs systèmes (**Windows, Linux, macOS**), mais aujourd'hui remplacée par Vulkan dans les jeux récents.
-  **Vulkan** : Une API ouverte plus moderne et performante, qui offre plus de contrôle aux développeurs et permet de mieux exploiter les cartes graphiques multicœurs. Elle est souvent utilisée dans les jeux récents et sur Linux.
-  **Metal** : Développée par **Apple**, cette API est optimisée pour **macOS, iOS et iPadOS**. Elle permet de mieux exploiter les performances des appareils Apple, mais est moins répandue dans les jeux que DirectX ou Vulkan car réservée aux appareils Apple.

### Comment tout cela fonctionne ensemble ?

Quand on lance un jeu, celui-ci utilise une **API graphique** (comme DirectX, Vulkan ou Metal) pour envoyer des instructions à la carte graphique, qui va ensuite **afficher les images sur l'écran grâce aux pilotes**. Plus l'API est efficace, plus le jeu est fluide et beau, car elle permet d'optimiser la manière dont la carte graphique traite les informations.

En résumé, **les pilotes sont le lien entre le matériel et le logiciel**, et **les API sont les outils que les jeux utilisent pour parler à la carte graphique**. Une bonne combinaison des deux permet d'avoir de **meilleurs graphismes et de meilleures performances**.

Vulkan se présente de plus en plus comme **concurrent solide à DirectX par ses capacités**, il permet d'ailleurs d'**exécuter des jeux Windows sur Linux et macOS** grâce à des projets comme que DXVK, il permet même d'améliorer les performances de certains jeux Windows (sur Windows) en passant par DXVK, comme c'est le cas avec GTA 4 (DX9) qui est connu pour être un port d'une médiocrité remarquable sur PC.

Il existe également, mais plus rarement, des mods permettant de changer complètement le moteur de rendu d'un jeu, par exemple pour Minecraft, un mod open-source permet de **passer d'OpenGL à Vulkan**, améliorant les performances:



*Fun fact: La Xbox porte son nom en référence au X de DirectX.*

### 3) Futures évolutions

L'industrie des cartes graphiques continue d'évoluer rapidement, portée par des avancées en **intelligence artificielle, en efficacité énergétique et en nouvelles architectures**. Plusieurs tendances se dessinent pour les années à venir, avec des impacts majeurs sur le jeu vidéo, la création de contenu et l'informatique en général.

#### L'intelligence artificielle et la génération d'images

Les dernières avancées, comme **DLSS 4** de NVIDIA ou les prochaines versions de **FSR d'AMD**, montrent que l'IA jouera un rôle toujours plus important dans le rendu des images. L'objectif est de **réduire la charge de calcul** en générant des images prédictives tout en améliorant la fluidité et la qualité graphique. À terme, l'IA pourrait même être capable de **reconstruire des scènes entières en temps réel** avec des détails que le jeu ne calculerait pas normalement.

#### Le quantum computing et son impact potentiel

Même si l'**informatique quantique** est encore loin d'une application grand public, elle pourrait bouleverser le secteur en permettant des **calculs ultra-rapides** pour la simulation physique, l'IA et l'optimisation des algorithmes de rendu. NVIDIA et d'autres acteurs investissent déjà dans cette technologie, bien que son intégration aux GPU classiques reste encore hypothétique à court terme.

#### L'efficacité énergétique : un enjeu crucial et géopolitique

Avec la demande croissante en puissance de calcul, l'optimisation de la consommation énergétique devient une priorité. NVIDIA, AMD et Intel travaillent sur des architectures plus économes en énergie, notamment avec des **techniques de gestion dynamique de la puissance** et l'intégration de **NPU (Neural Processing Units)** pour alléger la charge du GPU pour certains calculs. L'objectif est d'offrir des performances maximales tout en limitant la chauffe et la consommation électrique, un point critique pour les enjeux géopolitiques actuels avec les questions de **protection de l'environnement** et de **réchauffement climatique**.

#### Le cloud gaming : l'avenir du jeu vidéo ?

Avec l'essor du **cloud gaming** (comme **GeForce Now, Xbox Cloud Gaming et Shadow**), une partie des calculs graphiques est désormais effectuée à distance sur des serveurs équipés de puissants GPU, évitant aux joueurs d'investir dans du matériel coûteux. À l'avenir, l'amélioration des **connexions Internet et des infrastructures de serveurs** pourrait rendre cette technologie plus accessible, permettant de jouer en **4K et à haute fréquence d'images** depuis n'importe quel appareil, même un simple smartphone ou une TV connectée.

#### La simulation scientifique et industrielle

Les cartes graphiques sont devenues un outil essentiel dans de nombreux domaines scientifiques et industriels. Grâce à leur puissance de calcul parallèle, elles sont utilisées

pour des **modélisations complexes nécessitant des milliards d'opérations simultanées**.

- **La recherche médicale** : les GPU accélèrent la conception de nouveaux médicaments et la simulation de protéines, comme on l'a vu avec **Folding@Home**, un projet exploitant la puissance des cartes graphiques pour étudier les maladies.
- **La finance et la physique des particules** : les institutions financières et les laboratoires de recherche utilisent des GPU pour des analyses massives de données, permettant de **modéliser des phénomènes complexes** comme la prévision des marchés ou les interactions subatomiques.
- **La modélisation climatique** : pour prévoir avec précision l'évolution du climat et anticiper les catastrophes naturelles, les GPU sont utilisés pour effectuer des simulations à grande échelle.

### L'automobile et la robotique

Dans l'industrie automobile et la robotique, les cartes graphiques sont un élément clé pour le traitement en temps réel des informations issues des capteurs.

- **Les voitures autonomes** : NVIDIA, avec sa gamme **Drive**, propose des GPU dédiés aux véhicules autonomes. Ces cartes permettent d'analyser en **quelques millisecondes** les images des caméras et les données des capteurs LIDAR pour détecter les obstacles, prendre des décisions et améliorer la sécurité des trajets.
- **La robotique avancée** : les robots modernes, qu'ils soient industriels ou destinés à l'assistance humaine, nécessitent une forte puissance de calcul pour **analyser leur environnement, planifier leurs actions et interagir avec les humains**. Des GPU spécialisés permettent d'améliorer leur autonomie et leur réactivité.

## 4) Impact environnemental

L'évolution des cartes graphiques a permis des avancées majeures dans de nombreux domaines, mais elle s'accompagne également d'un **coût environnemental croissant**. De leur fabrication à leur utilisation, les GPU consomment des ressources et de l'énergie en quantités importantes, soulevant des **problématiques écologiques** que les fabricants cherchent aujourd'hui à atténuer.

### Une consommation énergétique en forte hausse

Les cartes graphiques modernes sont de plus en plus puissantes, mais cela s'accompagne d'une **hausse significative de leur consommation électrique**. Les GPU haut de gamme comme les RTX 4090 de NVIDIA ou les Radeon RX 7900 XTX d'AMD peuvent **dépasser les 400 W de consommation**, un chiffre bien supérieur à celui des générations précédentes. Cette tendance a un impact direct sur :

- **La consommation des datacenters** : les serveurs utilisant des GPU pour l'intelligence artificielle, le cloud gaming et la simulation scientifique nécessitent une quantité massive d'énergie.
- **L'empreinte carbone des joueurs et professionnels** : une utilisation prolongée de GPU puissants entraîne une consommation électrique accrue, impactant l'empreinte carbone individuelle et collective.

### Fabrication et extraction des matières premières

La production des cartes graphiques repose sur des composants nécessitant des **métaux rares** comme le cobalt, le tantale ou le lithium. L'extraction de ces ressources pose plusieurs problèmes :

- **Pollution des sols et des eaux** : l'exploitation minière, notamment dans des pays comme la Chine ou la République Démocratique du Congo, génère des déchets toxiques et pollue les écosystèmes environnants.
- **Conditions de travail difficiles** : certaines mines fonctionnent dans des conditions précaires, avec une exploitation de la main-d'œuvre, y compris des enfants.

De plus, la fabrication des GPU repose sur un processus complexe nécessitant **des usines de semi-conducteurs**, comme celles de TSMC ou Samsung, qui consomment d'énormes quantités d'eau et d'énergie.

### L'obsolescence et la gestion des déchets électroniques

Le rythme de renouvellement des cartes graphiques est rapide, avec **une nouvelle génération tous les deux à trois ans**. Cela entraîne :

- **Une obsolescence rapide du matériel** : même si certaines cartes restent performantes plusieurs années, les nouvelles technologies (DLSS, Ray Tracing...) poussent les utilisateurs à acheter du nouveau matériel, augmentant ainsi les déchets électroniques.
- **Un recyclage insuffisant** : bien que certaines initiatives existent, le recyclage des composants électroniques reste limité. Beaucoup de GPU usagés finissent en **décharge ou sont mal recyclés**, contribuant à la pollution électronique mondiale.

### Quelles solutions pour réduire l'impact environnemental ?

Face à ces enjeux, plusieurs initiatives émergent pour rendre les cartes graphiques plus respectueuses de l'environnement :

- **Optimisation de l'efficacité énergétique** : NVIDIA, AMD et Intel développent des architectures plus efficaces, réduisant la consommation tout en augmentant les performances. Le recours à des **NPU et à l'IA** permet aussi d'optimiser les calculs et d'éviter le gaspillage énergétique.
- **Développement du cloud gaming et du partage des ressources** : en déportant la puissance de calcul vers des serveurs optimisés, le cloud gaming pourrait réduire le besoin de cartes graphiques ultra-puissantes chez les utilisateurs.
- **Recyclage et économie circulaire** : des entreprises et organisations encouragent le **recyclage des GPU**, la réutilisation des composants et la fabrication de cartes plus durables.
- **Production plus responsable** : des initiatives visent à limiter l'utilisation de **métaux rares**, à favoriser des circuits de production plus éthiques et à réduire la consommation d'eau et d'énergie dans la fabrication des semi-conducteurs.

De manière générale, on retrouve des efforts de plus en plus concrets de la part des entreprises et des Big Tech pour réduire leur impact environnemental, avec l'utilisation croissante de matériaux recyclés, Apple étant l'un des acteurs principaux dans cette initiative.

Avec l'arrivée récente de l'IA **DeepSeek**, on voit qu'il est possible d'avoir une IA aux mêmes capacités de ses concurrents ChatGPT, Copilot ou Gemini mais **avec 97% de ressources nécessaires en moins**, offrant alors une alternative beaucoup plus respectueuse des impacts environnementaux.



Le data center derrière Grok AI aux États-Unis, utilisant environ 200 Mégawatts.