

L'incroyable saga de Nvidia, nouveau roi des puces électroniques

Le spécialiste américain des puces graphiques affiche des résultats stratosphériques, portés par l'essor de cette nouvelle technologie.



RÉCIT

LUCAS MEDIAVILLA (« LE FIGARO »)

Ce n'est pas encore un lieu de pèlerinage. Mais en périphérie de San José, en Californie, le Denny's du 2484 Berryessa Road est en bonne voie pour entrer dans la légende de la Silicon Valley, aux côtés des « garage's companies » fondées par Google, Amazon, Hewlett-Packard, Microsoft et autres Dell. C'est dans cette modeste chaîne de « diner » américain, il y a maintenant une trentaine d'années, que Jensen Huang, Chris Malachowsky et Curtis Priem ont posé les fondements d'une société, Nvidia, valorisée aujourd'hui plus de 1.200 milliards de dollars (1.110 milliards d'euros) et dont les résultats publiés mercredi dernier attestent de sa stature de géant de l'intelligence artificielle (IA).

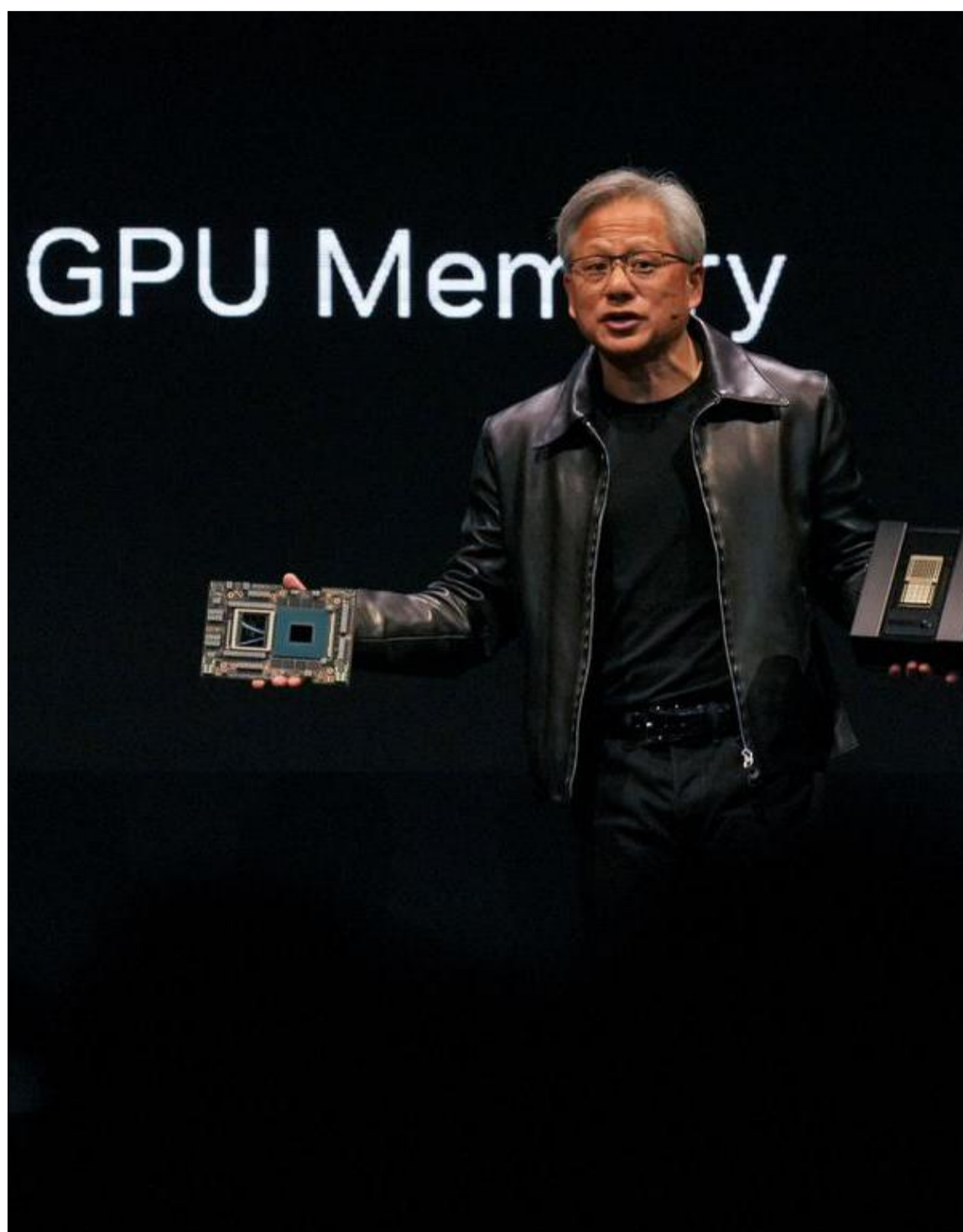
En trente ans d'existence, son nom n'a longtemps été connu qu'auprès de la communauté des « gamers ». Dès qu'ils en apercevaient le logo au lancement de leurs jeux vidéo préférés sur PC ou console, ils pouvaient être presque certains que la qualité et le réalisme des graphismes seraient au rendez-vous. Aujourd'hui, c'est à l'essor de l'IA que Nvidia doit ses lettres de noblesse. Et, comme pour les jeux vidéo, c'est en raison de la puissance de ses cartes graphiques (GPU), dont l'entreprise américaine revendique la paternité.

Pour le dire simplement, ces « cerceaux » de silicium permettent une puissance de calcul phénoménale, dédiée aux traitements graphiques. Ils sont dès lors rendus indispensables pour l'apprentissage automatique des ordinateurs (« machine learning ») l'apprentissage profond (« deep learning ») et tous les usages de l'IA, comme la reconnaissance faciale, mais aussi pour les logiciels de génération de texte comme GPT-4, d'images comme Midjourney, qui font l'actualité aujourd'hui.

Ruée vers l'or

Dans ce nouvel âge d'or de l'intelligence artificielle, Nvidia, entreprise sans usines, délègue la construction des cartes qu'il dessine au géant taïwanais TSMC. Sa part de marché atteindrait 90 % sur les puces IA, selon les analystes. « Pour les usages liés à l'IA, notamment générative, les processeurs GPU sont ce qui se fait de mieux sur le marché à l'heure actuelle », indique Tristan Gerra, analyste semi-conducteur chez Baird.

Dès lors, toutes les entreprises se les arrachent. « La demande est si forte que nous connaissons aujourd'hui une pénurie aiguë de GPU Nvidia », confirme Glenn O'Donnell, analyste pour le cabinet Forrester. « Cela pousse les prix du marché à dépasser, de loin, le prix catalogue et permet au groupe d'obtenir une marge plus élevée grâce à cette demande. » La puce H100 par exemple, spécifiquement conçue pour l'IA générative, s'échange aujourd'hui à 30.000 dollars (27.760 euros) l'unité. A ce prix-là, évidemment, peu d'entre-



Jensen Huang, le PDG de Nvidia Corp, au forum Computex de Taipei, à Taïwan, le 29 mai dernier.

© ZUMA PRESS.

dépendant des capacités de production de ce dernier pour livrer ses clients. « Nous nous attendons à ce que notre approvisionnement augmente dans les prochains mois », assure Colette Kress, patronne du groupe. Selon Tristan Gerra, TSMC pourrait « multiplier par 2, voire 2,5, ses capacités l'année prochaine », ce qui devrait à ses yeux « propulser les ventes de Nvidia ». De quoi soutenir un marché en forte croissance. Selon Gartner, le marché des puces IA, qui devrait atterrir autour de 53,4 milliards de dollars cette année, pourrait doubler pour atteindre 119 milliards de dollars (110 milliards d'euros) à l'horizon 2027.

La concurrence s'active

Quelle part de ce gâteau s'arrogera Nvidia ? L'explosion de l'IA aiguise les appétits et la concurrence fourbit ses armes depuis plusieurs mois. Les principaux concurrents, comme AMD et Intel, disent avoir progressé dans la conception de puces pour mordre les mollets du groupe californien. S'ils ont les moyens de payer leurs GPU, Google, Amazon et Microsoft travaillent également sur des alternatives avec des puces maison pour leurs activités cloud. Apple également, pour l'instant pour ses propres besoins.

Compte tenu de leur poids dans les ventes de Nvidia – au moins 60 % sur le pôle data center, le plus en croissance –, cela pourrait constituer une menace. Mais pas dans l'immédiat. « Leur seule chance », remarque Pierre Ferragu, « c'est que Nvidia s'endorme, ce qui est peu probable vu l'historique de la société. Je ne vois personne s'imposer sur les dix prochaines années, notamment en raison du rattrapage qui serait nécessaire sur le software ». En effet, la richesse de Nvidia repose aussi bien sur son hardware que son interface Cuda. « C'est une plateforme logicielle qui permet aux développeurs d'utiliser les GPU pour des applications non graphiques, comme le calcul de haute performance », souligne John Lorenz, spécialiste de ce secteur pour le cabinet Yole Group. « Ils travaillent depuis quinze ans sur ce sujet, avec une expérience énorme dans la résolution de problèmes. » Pas moins de 30 milliards de dollars auraient été investis depuis la création de Cuda, en 2007. « Personne n'a autant investi sur ce sujet que Nvidia », poursuit Tristan Gerra.

Le découplage technologique avec la Chine, l'un des premiers marchés du groupe, n'inquiète pas plus Jensen Huang. Il faut dire que son lobbying intense auprès de Washington pour réduire l'ampleur des restrictions à l'export vers Pékin a payé : seules les puces les plus avancées sont visées par les interdictions, et les Chinois se ruent sur ses modèles plus anciens. Même sans la Chine, la croissance des marchés promet de faire tourner les usines de TSMC à plein. Et, dans son sillage, la « cash machine » de Nvidia.

De quoi doper encore un peu plus la capitalisation boursière d'un groupe déjà valorisé par les marchés plus de 40 fois son résultat d'exploitation ? Dans le concert des analystes, certains restent encore sceptiques. « Le battage médiatique gonfle le prix de son action à un niveau dangereusement élevé », estime Glenn O'Donnell. « Il s'agit d'une excellente société, dotée d'un excellent leadership, mais elle ne vaut pas 1.000 milliards en Bourse. »

Surcoté ou non, Nvidia prouve avec ses résultats que la bulle IA n'est pas prête à éclater.

prises ont les moyens de s'aligner, d'autant qu'il faut un certain nombre de puces dans un data center pour entraîner efficacement les modèles d'IA.

Peu surprenant, dès lors, de trouver parmi les premiers clients de Nvidia les géants Amazon, Google, Microsoft ou encore Meta. Les quatre mastodontes américains constituent les stocks pour entraîner leurs intelligences artificielles et proposer, dans le cas des trois premiers, des services basés sur l'IA à leurs clients, via leurs filiales cloud.

Les géants du net chinois, qui ne peuvent se fournir sur les modèles les plus récents, comme la H100, à cause des restrictions américaines à l'export, n'en hésitent pas moins à se ruier sur les modèles précédents. Ils en auraient récemment commandé pour cinq milliards de dollars. Les acteurs du Golfe, enfin, aux poches bien pleines, constituent les stocks dans l'espoir de devenir demain des acteurs incontournables dans l'IA.

Il y a quelques semaines, Elon Musk disait que ces puces étaient plus difficiles à trouver que de la drogue.

Doublement du chiffre d'affaires

Voilà pour les signaux, pas si faibles. Nvidia devait néanmoins prouver aux marchés qu'il avait les épaules solides pour faire face à l'explosion de la demande. Analystes, commentateurs et investisseurs s'étaient tous donné rendez-vous le 23 août pour les résultats du second trimestre publiés par le groupe de Jensen Huang. Pour mémoire, c'est lors de la présentation de chiffres déjà exceptionnels au début de l'été que le groupe amé-

Les investisseurs reconnaissent que la demande en IA est aussi réelle que n'importe quelle autre tendance technologique des trente dernières années

Daniel Ives
Spécialiste des valeurs tech



Avril 1993

Création de la société Nvidia en Californie, par Jensen Huang, Chris Malachowsky et Curtis Priem, salariés dans l'informatique.

Octobre 1999

Nvidia invente le premier processeur graphique (GPU) au monde, et entre au Nasdaq cette année-là.

Juin 2007

Nvidia lance Cuda, une interface qui permet aux développeurs de déployer leurs logiciels sur les GPU de Nvidia.

2020-2022

Avec ses gammes Ampere puis Hopper, Nvidia lance une catégorie de produits dédiés aux data centers pour les usages d'IA. L.M.A