



Documenttabbladen

☰ **Tabblad 1** ⋮

[Le boom de l'IA est un men...](#)

Arguments clés de l'en...

• GPU inutilisés ou sous-...

• Nvidia continue de livrer...

• Centres de données «...»

Contexte tiré des trava...

Contrepoints et nua...

Implications et cas par...



Le boom de l'IA est un mensonge : centres de données factices et GPU inutilisés

La vidéo YouTube « Le boom de l'IA est un mensonge : centres de données factices et GPU inutilisés » (mise en ligne fin mars 2026 sur la chaîne The Tech Report) présente une interview concise entre Isaac Pound et Ed Zitron, commentateur technologique de longue date, spécialiste des relations publiques, auteur de la newsletter « Where's Your Ed At » et animateur du podcast « Better Offline ». Zitron s'est imposé comme l'un des sceptiques les plus virulents et constants du phénomène de surmédiatisation de l'IA générative, arguant que la plupart des déploiements d'infrastructures, des prévisions de revenus et des promesses économiques reposent sur des hypothèses spéculatives, un financement circulaire et une demande exagérée, plutôt que sur une création de valeur durable.

Arguments clés de l'entretien

La thèse principale de Zitron dans ce segment d'environ 13 minutes s'inscrit dans la continuité de l'ensemble de

ses travaux : le déploiement massif de l'IA – notamment les investissements considérables dans les centres de données, les GPU et les infrastructures connexes – n'est pas aussi important ni aussi axé sur l'utilisation que le laissent entendre les gros titres. Parmi les points clés qu'il

souligne :

****GPU inutilisés ou sous-déployés****

: Nvidia continue de livrer d'énormes volumes d'accélérateurs haut de gamme (H100, Blackwell, etc.), mais leur installation physique accuse un retard considérable. La construction de centres de données se heurte à des difficultés concrètes : disponibilité de l'énergie, transformateurs, refroidissement, interconnexions au réseau, main-d'œuvre et autorisations. Zitron souligne qu'il faut parfois six mois, voire plus, pour installer

seulement un trimestre de GPU Nvidia, ce qui crée un stock important de matériel inutilisé dans des entrepôts ou des installations partiellement alimentées. De ce fait, les GPU perdent de la valeur plus rapidement que prévu en raison d'une obsolescence accélérée (les architectures plus récentes comme Blackwell ou Rubin rendent les générations précédentes moins compétitives) et de l'amortissement.

- ****Centres de données « fictifs » ou utopiques****

: Les annonces de projets de plusieurs gigawatts devançant souvent les capacités réellement mises en chantier ou opérationnelles. Seule une fraction de la capacité annoncée (par exemple, les estimations autour de 5 GW en construction dans le monde, sur des chiffres annoncés bien plus importants) se concrétise dans les délais. De nombreux accords et projections supposent une mise à l'échelle sans faille, ignorant les réalités de la chaîne d'approvisionnement en équipements électriques et en importations étrangères.

****Problèmes liés à la dette adossée aux GPU, notamment pour CoreWeave**** : Ce point est crucial. ****CoreWeave****, fournisseur de « néocloud » spécialisé dans le calcul pour l'IA, a emprunté massivement des milliards de dollars (entre 14 et plus de 21 milliards de dollars de dettes d'ici fin 2025/début 2026) en utilisant son parc de GPU Nvidia comme garantie. Zitron affirme que ce modèle est précaire : si le taux d'utilisation est inférieur aux prévisions, que les GPU se déprécient rapidement (potentiellement en six ans, voire moins, compte tenu des cycles d'innovation) et que les revenus de la location de

puissance de calcul ne couvrent ni les intérêts ni le capital, l'entreprise risque un défaut de paiement technique ou une vente forcée à prix bradés. CoreWeave a fait l'objet d'un examen minutieux concernant d'importantes faiblesses de ses contrôles internes, des

violations de clauses contractuelles (dont des dispenses ont été accordées par des prêteurs comme Blackstone) et une forte dépendance à l'égard de quelques grands clients (OpenAI, Meta, Microsoft et même Nvidia elle-même en tant qu'acheteur de dernier recours).

Zitron présente cela comme faisant partie d'une dynamique plus large de « crise de l'IA à risque » — un financement spéculatif soutenu par le battage médiatique, où les signaux de la demande (carnets de commandes, annonces de dépenses d'investissement) masquent une économie sous-jacente faible.

Contexte tiré des travaux plus généraux de Zitron

Cette interview ne constitue pas une opinion isolée et spontanée, mais s'inscrit dans la continuité des épisodes de *Better Offline* et des articles de la newsletter de Zitron. Il a consacré plusieurs segments à :

Flux monétaires circulaires : Nvidia vend des GPU à CoreWeave, Lambda, etc., qui les utilisent comme garantie pour obtenir des prêts afin d'acquérir davantage de GPU. Nvidia rachète ensuite de la capacité de calcul auprès de ces fournisseurs (par exemple, un accord de 6,3 milliards de dollars avec CoreWeave) pour maintenir artificiellement la demande. Les hyperscalers et les

laboratoires d'IA (OpenAI, Anthropic) investissent

massivement dans l'inférence et l'entraînement, avec un retour sur investissement limité et démontré pour les clients finaux.

- Absence de revenus réels en dehors de la phase de surmédiation : la demande réelle des entreprises pour l'IA générative (au-delà des expérimentations ou des projets pilotes subventionnés) reste modeste. Les gains de productivité sont faibles ou difficiles à mesurer ; de nombreuses entreprises ayant adopté l'IA font état de résultats surestimés.

- Comparaison avec les bulles spéculatives du passé : Il compare certains aspects à ceux de WeWork (croissance à tout prix surévaluée avec une faible rentabilité unitaire) ou à l'essor des entreprises Internet/télécoms, mais soutient que l'IA pourrait être pire en raison des énormes besoins en énergie et des coûts environnementaux sans retombées sociétales proportionnelles.

Zitron oppose souvent cela aux succès technologiques historiques (par exemple, AWS a mis des années à devenir rentable avec des investissements en capital bien moins importants, alimentés par le battage médiatique, par rapport aux rendements).

Contrepoints et nuances : une vision équilibrée

Si le scepticisme de Zitron met en lumière des risques réels, la question de l'infrastructure d'IA présente de multiples facettes. Les partisans et les documents déposés par les entreprises dressent un tableau différent :

- ****Excellents résultats publiés par CoreWeave et Nvidia**** : CoreWeave a enregistré une hypercroissance

de son chiffre d'affaires (par exemple, 1,36 milliard de dollars au cours d'un récent trimestre, avec des projections pour l'exercice 2025 dépassant les 5 milliards de dollars et des prévisions de 12 à 13 milliards de dollars pour 2026). L'entreprise revendique un carnet de commandes colossal (50 à 66 milliards de dollars et plus), un taux d'utilisation des GPU élevé (plus de 50 % selon certains rapports, contre moins pour les clouds traditionnels) et une capacité opérationnelle importante (plusieurs centaines de mégawatts en ligne, avec des objectifs à l'échelle du gigawatt d'ici 2030). Nvidia continue d'afficher un chiffre d'affaires explosif pour ses centres de données (plusieurs dizaines de milliards de dollars par trimestre) et a investi directement dans CoreWeave (plus de 2 milliards de dollars). Son carnet de commandes comprend des contrats avec Meta, OpenAI et d'autres entreprises.

****Signaux de la demande vs. délais de construction**** : Il existe un décalage reconnu entre les livraisons de GPU et les capacités opérationnelles. Les contraintes liées à l'énergie et aux infrastructures sont bien réelles (les transformateurs ont de longs délais de livraison ; les réseaux électriques sont saturés). Cependant, les investissements collectifs des hyperscalers devraient se chiffrer en centaines de milliards de dollars par an, la part dédiée à l'IA étant prépondérante. Le PDG de Nvidia a réfuté les rumeurs de « bulle énergétique », évoquant une demande soutenue, voire croissante, et des gains d'efficacité dans les nouvelles puces, susceptibles d'allonger la durée de vie du matériel pour les charges de travail d'inférence. Certaines analyses suggèrent que le goulot d'étranglement s'est déplacé des puces vers l'énergie, mais le taux d'utilisation pourrait s'améliorer à mesure que les déploiements rattrapent leur retard.

****Risques et mesures d'atténuation liés au financement**** :

L'utilisation des GPU comme garantie est innovante mais risquée. L'innovation rapide accélère l'amortissement, et les charges d'intérêts (de l'ordre de 10 à 20 %) pèsent sur les marges. CoreWeave a obtenu d'importants financements (par exemple, un prêt à tirage différé de 8,5 milliards de dollars en 2026), le soutien de Nvidia et des contrats clients qui lui assurent une certaine visibilité. Les critiques soulignent les risques d'« obstacles à l'échéance » (importants remboursements de capital à partir de 2026), un flux de trésorerie disponible négatif et

la dépendance à un engouement continu ou à des renflouements par le biais d'investissements. Les partisans, quant à eux, font valoir que le carnet de commandes offre une marge de manœuvre de plusieurs années et que le potentiel à long terme de l'IA (automatisation, nouvelles applications) justifie un investissement initial important, à l'instar des premiers déploiements du cloud ou d'Internet. Dans certains cas particuliers, si un client majeur (par exemple, OpenAI) rencontre des difficultés de financement ou réoriente ses activités en interne, ou si le retour sur investissement pour les entreprises ne se concrétise pas, le taux d'utilisation pourrait chuter, rendant les actifs non rentables. À l'inverse, si des avancées de type AGI ou une IA agentive généralisée émergent, la demande pourrait exploser.

- **Implications plus larges** :

- **Aspects économiques** : Les dépenses prévues de plusieurs milliers de milliards de dollars liées à l'IA pourraient mettre à rude épreuve les marchés de l'énergie, faire grimper les coûts de l'électricité et créer des actifs dépréciés en cas de surproduction. Ce phénomène concentre les risques entre les mains de quelques acteurs (Nvidia détient environ 90 % des parts de marché des GPU dédiés à l'IA).

- **Marché** : La valorisation de Nvidia est fortement liée à ce discours ; tout ralentissement des ventes de GPU pourrait entraîner des corrections.

- ****Aspects sociétaux/environnementaux**** : La forte consommation d'énergie soulève des questions sur la durabilité, la modernisation du réseau et les coûts d'opportunité (par exemple, par rapport à d'autres infrastructures).
- ****Innovation contre effet de mode**** : Les sceptiques, comme Zitron, affirment que les formes actuelles d'IA générative (chatbots, génération d'images) offrent une fiabilité médiocre et une productivité transformatrice limitée à l'échelle de l'entreprise, avec des coûts d'inférence élevés. Les optimistes, quant à eux, mettent en avant l'amélioration des modèles, les applications spécialisées (assistants de programmation, simulation scientifique) et les effets potentiellement déflationnistes liés à l'efficacité.

Les autres éléments à prendre en compte incluent le contrôle réglementaire (lutte antitrust contre les investissements des géants de la tech dans l'IA), la concentration des talents et la nature « subjective » de nombreuses évaluations (les annonces influencent davantage les mouvements boursiers que les résultats audités).

Implications et cas particuliers

L'analyse de Zitron met en garde contre une possible correction, voire un effondrement comparable à celui des

subprimes, en cas de raréfaction des financements, de faible utilisation ou de demande accrue de rentabilité par les clients. Ceci pourrait entraîner des ventes massives de GPU dépréciés, des pertes pour les prêteurs et des répercussions négatives sur Nvidia et les hyperscalers.

Un scénario alternatif plus optimiste envisage une accélération de la construction grâce à des mesures politiques favorables (par exemple, des réformes des permis énergétiques), une normalisation de l'utilisation à mesure que les logiciels et les outils gagnent en maturité, et la validation des investissements par de nouvelles sources de revenus (services d'IA pour entreprises, IA souveraine).

En réalité, la vérité se situe probablement dans une zone intermédiaire complexe : une demande significative, mais inégale, existe en matière de calcul pour l'IA, concentrée entre les mains des hyperscalers et d'une poignée de laboratoires. Cependant, l'ampleur des investissements comporte des risques classiques de bulle spéculative : un optimisme excessif quant aux délais, une sous-estimation des contraintes matérielles et des incitations mal alignées (les fournisseurs profitent de l'engouement suscité par l'IA). L'histoire montre que les booms des infrastructures technologiques (chemins de fer, fibre optique, débuts du cloud) entraînent certes du gaspillage et une consolidation, mais peuvent générer des gains à long terme. La question qui reste ouverte est de savoir si

l'utilité marginale de l'IA générative justifie la trajectoire actuelle.

Cet entretien offre une synthèse pertinente et accessible de la critique de Zitron – à voir absolument pour sa franchise – mais il est enrichi par des recoupements avec les documents déposés par l'entreprise, des analyses indépendantes des investissements et des données d'utilisation. Le débat sur le déploiement de l'IA reste très polarisé, les retards dans l'obtention des données et les projections prospectives rendant toute décision définitive difficile. Des cas particuliers, comme des restrictions géopolitiques sur les puces, des pénuries d'énergie ou des avancées soudaines, pourraient rapidement faire basculer la situation dans un sens ou dans l'autre.

passer la situation dans un sens ou dans l'autre.