

MARS 2024

L'essor rapide de l'IA d'entreprise et ses conséquences pour vous

Mike Leone, analyste principal

Résumé : L'intelligence artificielle de niveau entreprise, en particulier l'IA générative (GenAI), transforme la façon dont les organisations atteignent leurs objectifs stratégiques, de l'augmentation du chiffre d'affaires et des bénéfices à la création d'une expérience client de qualité supérieure. L'IA est devenue une priorité pour de nombreuses organisations, attirant l'attention et le soutien des cadres dirigeants, du conseil d'administration et des lignes métier. L'IA présente certains défis et obstacles à surmonter, mais les entreprises avancent à grands pas en raison du potentiel énorme de cette technologie en matière de changement, d'amélioration et d'innovation, en particulier au niveau du client. Voici pourquoi l'IA d'entreprise a suscité l'enthousiasme de tous et pourquoi les organisations ont besoin de l'aide de partenaires technologiques.

L'explosion de la demande de solutions d'IA d'entreprise

En très peu de temps, l'IA s'est placée en tête des priorités technologiques des entreprises. Elle surpasse même désormais le cloud computing en tant qu'initiative clé et prend de plus en plus d'importance, selon une recherche menée par le groupe Enterprise Strategy Group de TechTarget. Dans une récente enquête, 54 % des entreprises ont déclaré avoir l'intention d'adopter l'IA générative dans les 12 prochains mois.¹ Notez que le mot employé est bien *adopter*, pas *envisager* ou *évaluer*. L'IA générative est une mesure prioritaire, souvent prise par des acteurs des lignes métier qui veulent exploiter l'IA pour leurs objectifs commerciaux les plus critiques sur un large éventail d'applications et de cas d'utilisation. Ces initiatives reçoivent le soutien massif des cadres dirigeants et des membres du conseil d'administration, et beaucoup ne demandent pas *si* l'IA générative devrait devenir un élément clé du plan d'exploitation, mais *quand* et *comment* elle le devrait.

Informations sur le marché



54 % des entreprises prévoient d'adopter l'IA générative dans les 12 prochains mois.

L'IA est généralement considérée comme un multiplicateur de force hautement exploitable qui effectue les tâches essentielles plus rapidement et souvent plus précisément que les humains. Elle offre également une valeur ajoutée supplémentaire en transférant les tâches quotidiennes à des outils d'IA et des algorithmes d'apprentissage automatique. Le personnel peut ainsi consacrer plus de temps aux tâches plus innovantes et créatives, et ouvrir de nouvelles opportunités commerciales. Parallèlement, les organisations doivent reconnaître certains défis précoces liés à l'adoption et à l'utilisation de l'IA générative dans les environnements d'entreprise réels. Par exemple, seulement 36 % des entreprises ont déclaré avoir mis en place une politique qui limite l'utilisation de l'IA générative. Il est également important de noter que, selon la même enquête, seulement 46 % des organisations s'emploient à créer une politique d'utilisation acceptable pour l'IA.

¹ Source : rapport de recherche d'Enterprise Strategy Group, [Beyond the GenAI Hype: Real-world Investments, Use Cases, and Concerns](#), août 2023. Sauf indication contraire, toutes les références et tous les graphiques de recherche d'Enterprise Strategy Group contenus dans cette présentation sont issus de ce rapport.

Il existe d'autres défis importants, notamment combler les lacunes importantes et croissantes en matière de compétences en IA, identifier et surmonter les nouvelles considérations éthiques et juridiques, et assurer et fournir des données de qualité irréprochable. Bien évidemment, que ce soit dans le cloud, dans un environnement sur site ou dans un modèle hybride, les cas d'utilisation de l'IA générative doivent tenir compte des exigences de conformité réglementaire majeures, variées et en constante évolution, en particulier celles qui affectent la confidentialité des données.

Opportunités de l'IA générative en entreprise : les cas d'utilisation à forte valeur ajoutée en font un élément essentiel

Pour qu'une technologie puisse prospérer, des cas d'utilisation réels sont essentiels. Ces cas d'utilisation nécessitent des modèles d'infrastructure adaptés, ainsi que des logiciels puissants pouvant être déployés dans une grande variété d'environnements. Dans le monde en constante évolution de l'IA générative, les cas d'utilisation sont aujourd'hui légion. L'introduction de grands modèles de langage (LLM) pour l'intelligence artificielle a été un facteur clé de l'émergence de ces cas d'utilisation. La bonne nouvelle ? De nombreux autres LLM sont en projet, offrant encore plus de fonctionnalités, un apprentissage plus rapide et des moteurs d'inférence plus utiles. Et l'autre bonne nouvelle ? À mesure que ces modèles deviennent plus puissants, ils deviennent également plus précis et plus contextuels. Il en résulte des solutions pratiques et innovantes pour des avancées spectaculaires, et même inattendues.

Informations sur le marché



Aujourd'hui, le principal cas d'utilisation de l'IA générative, cité par près d'un tiers des entreprises (31 %), concerne les informations sur les données.

Lorsque les cas d'utilisation sont pris en compte, leur grande diversité encourage encore plus d'expérimentations, davantage de programmes pilotes et l'engagement organisationnel des principales ressources (budgets, personnel, infrastructure informatique et partenariats industriels). Le rapport de recherche d'Enterprise Strategy Group souligne que la catégorie principale de cas d'utilisation concerne les informations sur les données, c'est-à-dire la capacité à traiter d'énormes volumes de données brutes pour fournir une connaissance et une pertinence contextuelles

permettant de les transformer en informations exploitables. Près d'un tiers des entreprises (31 %) ont déclaré que les informations sur les données sont une priorité absolue pour leurs projets d'IA générative, ce qui en fait le principal domaine d'intérêt.

Quels autres cas d'utilisation attirent l'attention des entreprises ? L'analyse des menaces de cybersécurité est l'un des cas d'utilisation les plus importants, et il continuera assurément à prendre de l'importance. Suivre le rythme effréné des menaces émergentes est essentiel pour les entreprises qui peinent à combler le fossé croissant des compétences en cybersécurité. C'est un domaine dans lequel les organisations utilisent l'IA pour équilibrer le terrain face aux pirates. D'autres cas d'utilisation importants incluent l'automatisation des workflows, la recherche des menaces de cybersécurité, la base de connaissances du centre de contact, la réponse aux incidents et les discussions avec des chatbots. Mais la prédiction la plus révélatrice sur l'importance de l'IA générative dans les cas d'utilisation en matière de prise de décision a été notée par Enterprise Strategy Group : « Pour mieux évaluer l'IA générative, les entreprises vont se tourner vers... l'IA générative. »

La nécessité de prendre en charge les environnements d'IA hybrides avec une infrastructure basée sur le client

En fonction de la diversité des cas d'utilisation d'entreprise pour l'IA, notamment ceux mis en œuvre sur des appareils clients, une décision clé doit être prise quant à l'emplacement où ces cas d'utilisation doivent être déployés : dans le cloud, sur site dans le centre de données, sur site sur des PC hautes performances ou dans un

environnement hybride mélangeant plusieurs scénarios. En réalité, les entreprises ont indiqué à Enterprise Strategy Group que, bien qu'elles veuillent utiliser des environnements de cloud public ou hybride pour les déploiements de cas d'utilisation de l'IA générative, 40 % de ces cas d'utilisation seront regroupés dans des centres de données sur site, des sites en périphérie ou des hébergeurs.²

Ces décisions sont bien entendu très spécifiques aux priorités de chaque entreprise, qui englobent les exigences de conformité et de génération de rapports, les considérations de sécurité et de confidentialité, le temps de latence et les performances d'accès au réseau, les contraintes budgétaires ou encore la nécessité d'une résilience élevée. En outre, le volume des données augmentant de plus en plus, les organisations recherchent une approche complémentaire au cloud public, qui leur permettrait de mieux prévoir les coûts et de garantir la disponibilité des ressources en temps réel. Cela conduit inévitablement à la nécessité d'un déploiement et de modèles d'exploitation hybrides, utilisant le meilleur du cloud public et le meilleur de l'infrastructure sur site basée sur le client.

Bien que de nombreux déploiements initiaux de PC d'entreprise aient été effectués dans le cloud, les organisations adoptent désormais la sagesse et la flexibilité du déploiement de l'IA à l'aide d'un modèle de base axé sur le client. L'accent est ainsi mis sur l'établissement d'une infrastructure de calcul adaptée pour prendre en charge les cas d'utilisation susceptibles d'être exigeants en matière de performances et sensibles à la sécurité.

Au lieu de simplement « stationner » l'infrastructure d'IA dans le cloud, la stratégie intelligente sera d'utiliser des PC client afin de compléter le traitement de l'IA basé sur le cloud. L'IA sur le matériel client de nouvelle génération offre un contrôle, une flexibilité et une confidentialité améliorés. Les solutions basées sur les points de terminaison permettront aux entreprises de garder un contrôle absolu sur leurs données, ce qui sera particulièrement utile lorsqu'elles traiteront des informations personnelles ou des données professionnelles confidentielles. Comme le montre la Figure 1, la recherche menée par Enterprise Strategy Group a révélé plusieurs facteurs influençant l'emplacement des environnements d'IA.³

Figure 1. Facteurs influençant l'emplacement physique d'un environnement d'IA

Quels sont les principaux facteurs influençant le choix de votre entreprise en matière d'emplacement physique spécifique pour son environnement d'IA ? (Pourcentage des participants, N = 339, trois réponses acceptées)



Source : Enterprise Strategy Group, une division de TechTarget, Inc.

² Source : rapport de recherche d'Enterprise Strategy Group, [Navigating the Evolving AI Infrastructure Landscape](#), septembre 2023.

³ Ibid.

Comment AMD répond aux besoins des entreprises en matière d'infrastructure d'IA

Au cœur de cette infrastructure d'IA se trouvent les composants-processeurs qui permettent d'exécuter des solutions non seulement dans le cloud, mais également en périphérie et sur des points de terminaison à haute puissance. Ces composants sont les suivants :

- processeurs pour serveurs et points de terminaison ;
- accélérateurs pour les exigences telles que l'inférence d'IA et l'apprentissage de l'IA ;
- cartes graphiques pour prendre en charge les cas d'utilisation de l'IA visuels ;
- solutions de systèmes sur puce (SoC).

AMD est un partenaire technologique de longue date qui s'est engagé à fournir des solutions au niveau du processeur prenant en charge les cas d'utilisation de l'IA. L'engagement d'AMD vis-à-vis des solutions d'IA couvre tous les types de solutions mentionnés ci-dessus, ce qui permet à AMD de travailler avec un large éventail de partenaires OEM et d'entreprises clientes afin d'étendre les avantages de l'IA.

L'intérêt d'AMD pour l'IA se manifeste sous trois formes :

- un portefeuille de solutions étendu et riche ;
- un écosystème ouvert et vaste de partenaires matériels et logiciels ;
- des expériences utilisateur pour les cas d'utilisation de l'IA qui peuvent répondre aux exigences des entreprises et aux charges de travail à budget limité.

L'infrastructure d'IA d'AMD, conçue pour propulser les mises en œuvre réelles de l'IA générative et des LLM, comprend un certain nombre de solutions matérielles. L'un des éléments clés est la gamme de processeurs AMD Ryzen™ avec AMD Ryzen™ AI, qui exploite des moteurs d'IA natifs et permet d'offrir des capacités telles que le multitâche accéléré et l'amélioration de la productivité des utilisateurs.

- Avec d'autres solutions AMD comme les accélérateurs Instinct™ et Alveo™, les processeurs pour serveur EPYC™, les SoC adaptatifs Versal™ et les cartes graphiques Radeon™, les processeurs Ryzen permettent à AMD de travailler avec des partenaires matériels et logiciels de premier plan pour fournir des solutions d'IA sophistiquées aux entreprises.

Les solutions AMD aident les organisations à obtenir de hautes performances et même plus en matière de formation et d'inférence, en offrant une évolutivité supérieure pour gérer les LLM toujours plus complexes et exigeants à venir.

Les cas d'utilisation actuels et émergents de l'IA exigent un large éventail de fonctionnalités de l'infrastructure sous-jacente, notamment un faible temps de latence, un débit élevé, des fonctionnalités de confidentialité et de sécurité, et une rentabilité pour rendre l'IA exploitable et abordable au niveau des points de terminaison. AMD a conçu ses produits pour aider les créateurs de solutions d'IA à les mettre en œuvre sur l'ensemble de l'architecture, du point de terminaison local au centre de données, et de la périphérie au cloud.

Conclusion : placer l'IA générative au cœur de la stratégie commerciale de l'entreprise

L'IA générative est bien plus qu'une technologie passionnante à fort potentiel ; elle porte déjà ses fruits pour les entreprises de tous les secteurs et de toutes les tailles. L'IA générative est manifestement une ressource optimisable qui favorise l'innovation et offre une valeur commerciale tangible. Les organisations sont également en train d'apprendre que les initiatives d'IA basées sur le client sont essentielles pour leurs stratégies globales d'IA, car elles apportent contrôle, flexibilité et performances aux cas d'utilisation de l'IA basés sur PC qui complètent les programmes d'IA sur serveur et dans le cloud.

L'IA dans l'entreprise présente encore des défis à relever et à surmonter. Cependant, ces défis impliquent des décisions commerciales concernant les stratégies, les priorités, les types de garde-fous à créer pour bénéficier d'une utilisation appropriée et les options de déploiement. L'utilité et les performances de cette technologie sont indéniablement adaptées aux cas d'utilisation de transformation, soutenus par des avancées fascinantes dans l'infrastructure d'IA, du client au centre de données et de la périphérie au cloud. Une approche hybride de l'IA d'entreprise, associant des solutions d'IA locales à celles développées et déployées dans le cloud, représente une avancée importante pour faire de l'IA générative une technologie révolutionnaire véritablement utile et rentable pour l'entreprise.

Des leaders technologiques innovants comme AMD et d'autres posent les bases de nouvelles solutions d'IA basées sur le client. Par conséquent, les organisations disposent désormais de la flexibilité, du contrôle et des performances nécessaires pour utiliser l'IA générative de manière à améliorer la sécurité et à créer de nouvelles capacités.

Pour en savoir plus sur la façon dont AMD aide à activer l'IA d'entreprise au niveau du client et à la périphérie, rendez-vous sur <https://www.amd.com/fr/solutions/ai.html>.

©TechTarget, Inc. ou ses filiales. Tous droits réservés. TechTarget et le logo TechTarget sont des marques commerciales ou des marques déposées de TechTarget, Inc. et sont déposées dans d'autres pays. D'autres noms et logos de produits et de services, y compris pour BrightTALK, Xtelligent et Enterprise Strategy Group, peuvent être des marques commerciales de TechTarget ou de ses filiales. L'ensemble des autres marques commerciales, logos et noms de marque appartient à leurs propriétaires respectifs.

Les informations contenues dans cette publication ont été obtenues auprès de sources que TechTarget considère comme fiables, mais ne sont pas garanties par TechTarget. Cette publication peut contenir des opinions de TechTarget, qui sont susceptibles d'évoluer. Cette publication peut inclure des prévisions, des projections et d'autres déclarations prédictives qui représentent les hypothèses et les attentes de TechTarget à la lumière des informations actuellement disponibles. Ces prévisions sont fondées sur les tendances du secteur et comportent des variables et des incertitudes. Par conséquent, TechTarget ne donne aucune garantie quant à l'exactitude des prévisions, projections ou déclarations prédictives spécifiques contenues dans le présent document.

Toute reproduction ou redistribution de cette publication, en totalité ou en partie, que ce soit au format papier, électronique ou autre à des personnes non autorisées à la recevoir, sans le consentement exprès de TechTarget, constitue une violation de la loi américaine sur le droit d'auteur et fera l'objet d'une action en dommages-intérêts civils et, le cas échéant, de poursuites pénales. Pour toute question, veuillez contacter le service des relations client à l'adresse cr@esg-global.com.

À propos d'Enterprise Strategy Group

Le groupe Enterprise Strategy Group de TechTarget fournit des analyses du marché ciblées et exploitables, des recherches côté demande, des services de conseil et d'analyse, des orientations stratégiques GTM, des validations de solutions et des contenus personnalisés pour faciliter l'achat et la vente de technologies d'entreprise.

✉ contact@esg-global.com

🌐 www.esg-global.com