



Boostez l'efficacité et optimisez les performances avec des solutions d'IA d'entreprise clés en main

AUTEURS

Steven Dickens

Conseiller en chef en technologie | The Futurum Group

Ron Westfall

Directeur de la recherche | The Futurum Group

AOÛT 2024

EN PARTENARIAT AVEC





Synthèse

L'IA est devenue une priorité absolue pour les entreprises et les organisations, mais son adoption rapide est entravée par diverses difficultés. Malgré la popularité des applications d'IA générative publiques, les services informatiques dans les entreprises rencontrent des problèmes : manque de compétences spécifiques à un domaine, préparation limitée de l'entreprise, problèmes de sécurité et une expertise insuffisante. Ces difficultés sont encore aggravées par des solutions de gestion des données inadéquates, des réglementations à respecter et des coûts élevés occasionnés par le Cloud et la consommation d'énergie. En outre, le marché fragmenté des éditeurs de logiciels indépendants pour les applications d'IA basées sur des cas d'utilisation complexifie encore les choses et ralentit potentiellement les progrès.

L'IA hybride apparaît comme une solution stratégique aux défis relatifs à l'adoption de l'IA. Cette approche combine le Machine Learning, le Deep Learning et les réseaux neuronaux à l'expertise humaine pour créer des modèles d'IA très précis et spécifiques à un domaine. En utilisant un mélange de ressources d'IA privées et publiques, les entreprises peuvent tirer parti des avantages inhérents aux deux : le contrôle et la personnalisation de l'infrastructure privée et l'évolutivité et les capacités de l'IA dans le Cloud. Ce modèle hybride permet aux organisations d'optimiser leur stratégie d'IA en fonction de leurs besoins spécifiques, de leur budget et de leurs exigences technologiques.

Pour les clients préoccupés par les questions de confidentialité et de sécurité des données, ou qui souhaitent gérer leurs propres systèmes d'IA, l'infrastructure d'IA privée est la solution à privilégier. En hébergeant les modèles d'IA et le traitement sur leur propre infrastructure sécurisée, les entreprises gardent une maîtrise et une visibilité complètes sur l'utilisation de leurs données et systèmes d'IA. Ce point est d'autant plus important pour les secteurs pour lesquels les réglementations en matière de données sont strictes, les questions de propriété intellectuelle sont sensibles ou pour les applications critiques. L'infrastructure privée dédiée à l'IA permet également plus de personnalisation et d'intégration avec les systèmes et les flux de travail existants d'une entreprise.

Pour les décideurs informatiques tels que les DSI, les directeurs de la technologie et les responsables de l'architecture, l'IA privée permet d'adapter les modèles d'IA aux besoins spécifiques de leur organisation avec une meilleure maîtrise de l'ensemble de leur Stack d'IA, y compris les modèles, les applications, l'infrastructure et les cadres d'exploitation. Ce contrôle amélioré permet aux entreprises de répondre à des exigences spécifiques en matière de personnalisation, de flexibilité, de respect de la vie privée, d'inférence en temps réel, de respect des normes et de sécurité. Ainsi, l'IA hybride avec des modèles privés devient un outil puissant pour les organisations qui cherchent à exploiter le potentiel de l'IA tout en s'attaquant aux difficultés relatives à la mise en œuvre et à la gestion des données.

Ce document explique comment les organisations peuvent planifier et créer des solutions d'IA privées et en quoi l'IA privée répond à la demande croissante d'un plus grand contrôle des solutions d'IA, y compris en termes de protection et de transparence. La stratégie Lenovo Smarter AI for All donne la priorité à l'IA d'entreprise, en exploitant sa puissance pour entraîner une transformation intelligente dans tous les secteurs et permettre aux organisations d'innover rapidement et de garder une avance technologique sur leurs concurrents.

Découvrez pourquoi la création de solutions d'IA privées réussies nécessite un écosystème de partenaires, une expertise en IA, un groupe soigneusement trié sur le volet d'éditeurs de logiciels indépendants et leurs applications d'IA, une infrastructure de taille adaptée pour des charges de travail ciblées et l'assistance des développeurs pour garantir une adoption de l'IA et une innovation à l'échelle de l'organisation.



Section 1 : le paysage de l'IA privée

À mesure que l'adoption de l'IA s'accélère, la protection des données sensibles est devenue primordiale. Les techniques d'IA privées, telles que la confidentialité différentielle et l'informatique chiffrée, offrent une solution puissante qui permet aux organisations d'assurer un entraînement réussi des modèles d'IA tout en maintenant une sécurité robuste.

Dans un contexte où l'IA privée prend de l'ampleur, nous avons identifié les secteurs qui, aujourd'hui, semblent prêts à en tirer le meilleur parti :

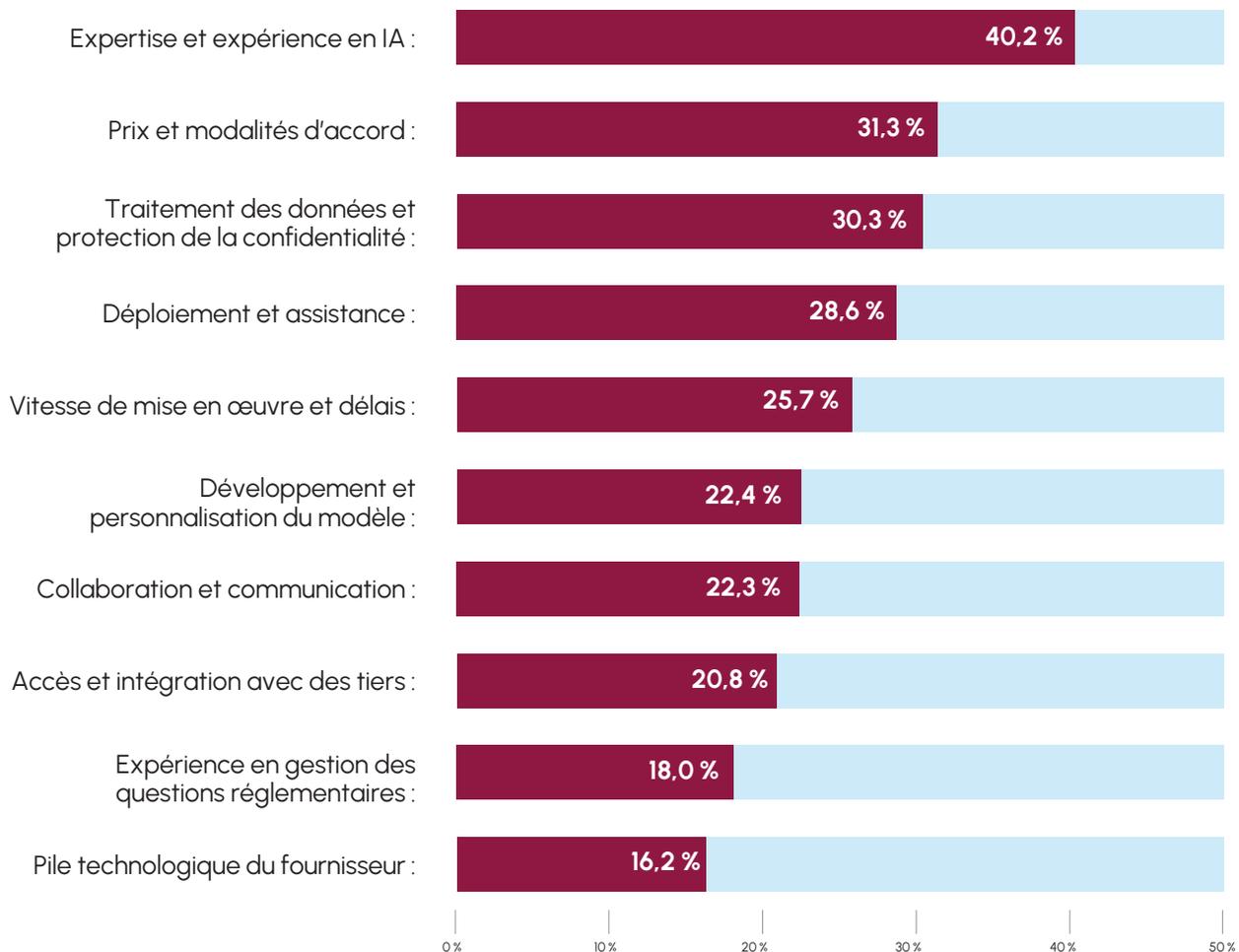
- Les secteurs hautement réglementés, tels que les services financiers, le secteur de la santé, le gouvernement, les services publics, ainsi que le pétrole et le gaz.
- Les secteurs distribués très axés sur l'Edge et avec de nombreux sites, comme le retail, les hôpitaux, les aéroports et les stades.
- Les organisations qui disposent déjà de beaucoup de données sur site et de centres de données prenant en charge de grands environnements privés.
- Les secteurs qui dépendent d'assistants IA hautement qualifiés/optimisés et spécifiques à un domaine, tels que les organisations juridiques, d'ingénierie et médicales.

Dans l'actuel paysage dynamique des entreprises, les sociétés (et les développeurs) adoptent une approche sophistiquée de la mise en œuvre de l'IA. Ils combinent stratégiquement divers éléments pour créer des solutions d'IA privée optimisées : petits et grands modèles, divers ensembles de données, CPU, GPU, etc. Cette personnalisation leur permet d'équilibrer des facteurs cruciaux tels que le coût, la sécurité, la conformité, les performances et la latence, ce qui rend l'IA privée idéale pour leurs besoins spécifiques.

De plus, les organisations qui explorent l'IA générative adoptent un modèle hybride. Cette approche intègre des modèles d'IA générative publics et privés et tire parti des atouts de chacun. Les modèles publics sont facilement accessibles et leurs fonctions sont disponibles immédiatement pour la génération de texte, d'images, de vidéos et de code. En revanche, les modèles privés, qu'ils soient auto-hébergés ou basés sur le Cloud, offrent une personnalisation plus poussée. Ces modèles privés sont souvent basés sur des modèles de base publics auxquels ont été ajoutées des applications personnalisées qui permettent des résultats spécialisés, un meilleur contrôle des données et des mesures de sécurité renforcées. Cette stratégie hybride permet aux entreprises d'exploiter la puissance de l'IA générative tout en gardant la flexibilité nécessaire pour répondre à leurs besoins et à leurs limites uniques.

Choisir le bon fournisseur

Dans le contexte actuel en matière d'IA, les critères de sélection de fournisseurs des organisations évoluent. Les entreprises accordent désormais autant d'importance aux capacités d'IA privée et à l'expertise générale dans le domaine lorsqu'elles choisissent leurs partenaires en IA. Plus précisément, elles accordent la priorité aux fournisseurs qui excellent dans le traitement sécurisé des données et mettent en œuvre des mesures robustes de protection de la confidentialité, tout en démontrant une vaste expérience et une grande compétence technique en développement et en déploiement d'IA. Ce changement reflète une prise de conscience grandissante que les solutions d'IA efficaces doivent non seulement être innovantes, mais aussi fiables et compatibles avec les normes de protection des données :



Source : The Futurum Group

Les organisations sont en train d'adopter rapidement des solutions d'IA privée et donnent la priorité aux fournisseurs qui répondent à leurs critères. Critères eux-mêmes en constante évolution. Ce changement est motivé par les principaux avantages inhérents à l'IA privée :

1. Mesures de sécurité améliorées
2. Inférence de l'IA rentable à long terme
3. Respect permanent des règlements
4. Capacités de personnalisation
5. Cadres de gouvernance améliorés

L'IA privée est très utile pour les organisations qui doivent se conformer à des lois strictes sur la protection des données, comme le RGPD européen ou le CCPA californien. En garantissant que les données sont gérées de manière sécurisée et dans le respect de ces règlements, l'IA privée aide les entreprises à évoluer au sein de paysages juridiques complexes tout en tirant parti du potentiel de l'IA.

Les réglementations spécifiques à l'IA sont en train d'entrer en jeu puisque la loi sur l'IA de l'UE établit une structure de gouvernance aux niveaux européen et national qui exige un audit de conformité avant de déployer des systèmes d'IA à haut risque et met en place des mécanismes d'application. De leur côté, les États-Unis, le Royaume-Uni et la Chine sont en train de développer des réglementations similaires. Une organisation aura besoin d'un modèle privé si :

- **La confidentialité des données est impérative** : un acteur du domaine de la santé pourrait utiliser un modèle d'IA générative pour développer un système de diagnostic de patient entraîné à partir des données des patients. Il pourrait ainsi offrir aux patients des diagnostics privés plus précis et personnalisés.
- **Des informations transactionnelles à jour sont nécessaires** : un détaillant pourrait utiliser un modèle d'IA générative pour développer un système de tarification dynamique mis à jour avec les dernières données de vente. Il pourrait ainsi optimiser les prix et maximiser ses bénéfices.
- **Il est possible d'exploiter les données propriétaires et la propriété intellectuelle** : un fournisseur de services financiers pourrait utiliser un modèle d'IA générative pour développer une stratégie de trading propriétaire entraînée sur ses données et ses recherches.

Capacités de personnalisation : l'IA générative privée

L'IA privée est devenue cruciale pour la priorité absolue des entreprises : l'adoption réussie de la technologie d'IA générative. Ce changement souligne l'importance croissante d'exploiter efficacement les modèles GPT (Generative Pretrained Transformer) privés. Ce faisant, les entreprises sont en mesure de créer leurs propres systèmes sécurisés d'IA générative et de moteurs décisionnels privés basés sur l'IA. Pour nous, les organisations qui souhaitent créer un Cloud privé pour l'IA devraient se focaliser sur les facteurs clés suivants :

- Des plateformes intégrées de qualité, capables de gérer la gouvernance, les données et l'apprentissage automatique, généralement pour les entreprises du Fortune 500.
- Des piles d'IA et de données délibérément pensées pour offrir un choix d'applications, d'agents et d'assistants IA et d'expertise, et complétées par un écosystème d'éditeurs de logiciels indépendants et de partenaires validés.

- Explorer et opter pour un mélange de plateformes open source et fermées, comme NVIDIA AI Enterprise et d'autres propriétaires de piles de plateformes de données.
- Donner la priorité aux plateformes open source qui peuvent être personnalisées pour les applications d'IA.
- L'Edge doit satisfaire aux exigences relatives aux données et à la latence, fondamentales pour créer des solutions de pointe à l'Edge avec les éditeurs de logiciels indépendants.
- Opter pour des capacités RAG (Retrieval-Augmented Generation) doit s'accompagner d'une base de données bien définie pour améliorer la définition du domaine ou du contexte spécifique et donc les réponses des modèles de langage.

L'écosystème de l'IA : opportunités et difficultés

Le paysage de l'IA évolue rapidement, et de nouveaux éditeurs de logiciels indépendants et applications d'IA émergent à mesure que les modèles open source gagnent en puissance et deviennent plus abordables. Les plateformes comme Hugging Face Hub, qui héberge plus de 350 000 modèles, 75 000 ensembles de données et 150 000 applications de démonstration, illustrent cette tendance. Cet écosystème en pleine expansion permet aux entreprises d'adapter les solutions d'IA privées à leurs besoins spécifiques.

Elles sont toutefois confrontées à des difficultés majeures pour naviguer dans ce paysage complexe et se tenir au courant des dernières avancées. Un partenariat avec des experts fiables peut aider les organisations :

1. À valider et à intégrer divers éléments de l'écosystème.
2. À accéder à des compétences critiques en IA et en données.
3. À choisir l'infrastructure d'IA optimale pour un équilibre prix, performances et sécurité.

Développement responsable de l'IA

Le développement éthique de l'IA repose sur une gestion et une protection responsables des données. L'objectif est de garantir que les modèles d'IA sont entraînés à partir de données représentatives, impartiales et précises. Les entreprises donnent la priorité à la flexibilité au moment de choisir des solutions d'IA privées qui protègent et gèrent au mieux leurs charges de travail.

Les facteurs clés du succès pour une IA à destination des entreprises comprennent :

1. Le respect des principes d'IA responsable.
2. L'implémentation de pratiques MLOps robustes.
3. Le développement de stratégies complètes de gestion des données pour tous les sites de l'organisation.

Ces éléments sont cruciaux pour renforcer la confiance dans les applications et les modèles d'IA, et ainsi les rendre véritablement prêts pour les entreprises.

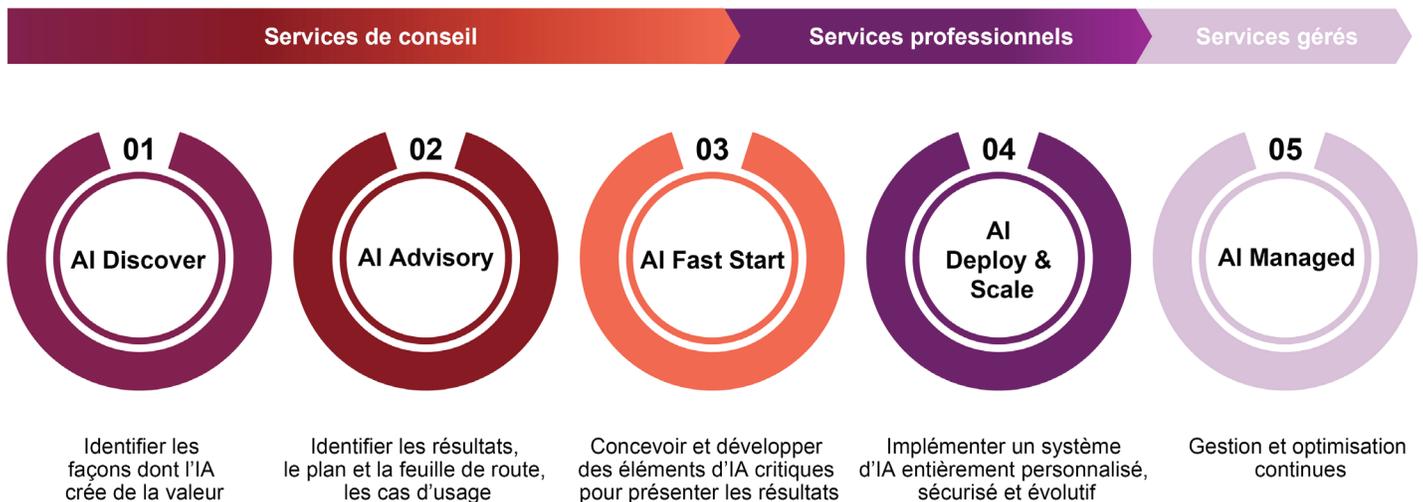
Section 2 : solutions d'IA d'entreprise de Lenovo

Lenovo Smarter AI for All propose une large gamme de solutions d'IA, des appareils mobiles à l'infrastructure d'entreprise. Les solutions d'IA d'entreprise de Lenovo sont sécurisées, efficaces et évolutives, ce qui aide les entreprises à utiliser l'IA de manière responsable et réussie.

Nous savons que Lenovo propose une suite complète de solutions d'IA d'entreprise qui donnent la priorité à la protection et à la gestion des données. Ces solutions offrent un équilibre entre flexibilité, sécurité et efficacité, ce qui permet aux organisations d'exploiter le potentiel de l'IA tout en protégeant les informations sensibles. L'approche de Lenovo comprend des capacités d'IA sur appareil et sur site, ce qui garantit que les données restent isolées des réseaux publics. Cette stratégie permet aux entreprises de mettre en œuvre l'IA de manière sûre et responsable, que ce soit sur des appareils individuels ou sur une infrastructure de serveur sur site. En mettant l'accent sur ces aspects critiques, Lenovo aide les entreprises à tirer parti de technologies d'IA avancées tout en maintenant une sécurité des données et une conformité réglementaire strictes.

Les organisations qui envisagent de créer une IA privée doivent commencer par comprendre leurs objectifs commerciaux et leurs capacités concernant les données. Ce processus important comporte de multiples facettes et dépend des capacités de l'organisation et de son objectif, mais les six étapes suivantes peuvent aider à réussir :

1. Identifier le problème commercial et le résultat souhaité.
2. Rassembler et analyser les données
3. Réaliser une analyse de rentabilité
4. Planifier le modèle opérationnel, l'infrastructure, la conception du modèle et le déploiement
5. Créer/entraîner ou régler le modèle
6. Déployer, surveiller les performances et affiner au besoin

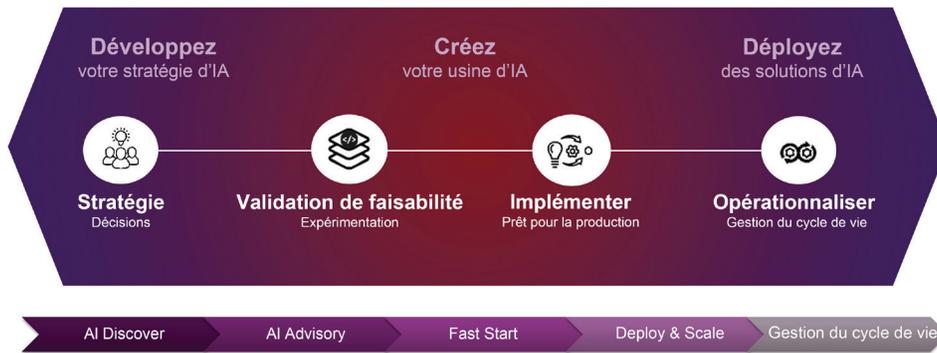


Source : Lenovo

Les équipes de Lenovo accélèrent le processus grâce aux programmes AI CoE et AI Innovators avec des partenaires et des éditeurs de logiciels indépendants. Les points clés de la solution comprennent :

- Des PC et les stations de travail avec IA à destination des employés, des data scientists et des développeurs
- Des solutions de gestion et de protection des données
- Des services et une expertise en IA capables d'accélérer le processus de planification, de stratégie et d'exécution
- Des solutions AI Fast Start qui permettent des cas d'utilisation et des projets pilotes industriels basés sur les données des organisations
- Des solutions d'infrastructure d'IA écoénergétiques pour les besoins des centres de données et à l'Edge, offrant une taille de traitement (CPU, GPU) adaptée, du stockage, des innovations en matière de refroidissement liquide et du choix

Commencez votre aventure IA où que vous soyez et évoluez avec Lenovo et NVIDIA



Source : Lenovo

Lenovo développe une architecture modulaire qui offre la flexibilité nécessaire pour s'adapter aux besoins uniques des clients tout en établissant des standards à l'aide de composants communs. Tout commence par des cas d'utilisation et des besoins métier, et se termine au niveau du serveur et du stockage, et cette approche se retrouve dans l'architecture de solution de Lenovo pour les clients qui souhaitent créer leur usine ou leur centre de compétence d'IA.



Source : Lenovo

Pour nous, le portefeuille d'IA d'entreprise de Lenovo et l'alliance avec NVIDIA se démarquent. Le tout nouveau design de référence de Lenovo pour l'IA générative est basé sur les LLM et utilise la plateforme logicielle NVIDIA AI Enterprise (comprenant le framework NVIDIA NeMo) pour montrer aux entreprises comment déployer et commercialiser des outils et des modèles d'IA générative de base, en tirant parti d'une solution pré-validée, entièrement intégrée et optimisée pour les performances. Grâce à son partenariat avec NVIDIA pour le co-développement de solutions, Lenovo est en mesure d'aider les entreprises à améliorer leurs flux de travail LLM et à maximiser les performances, la sécurité et le choix d'une taille appropriée.

Le paysage des infrastructures d'IA est complexe, et l'incertitude règne quant aux composants nécessaires aux architectures RAG et aux grands modèles de langage. Lenovo relève ce défi en proposant des solutions d'infrastructure modulaires et évolutives, allant des options standards aux options hautes performances, afin de permettre aux clients de prendre des décisions éclairées en fonction de leurs besoins spécifiques.

Offres principales :

1. Serveur NVIDIA OVX avec 4 ou 8 GPU pour l'inférence et les architectures RAG. Celui-ci utilise des GPU NVIDIA L40S et les technologies réseau NVIDIA Connect-X et BlueField et peut être mis à niveau vers NVIDIA H100NVL ou HGX à 8 GPU, pour des performances améliorées.
2. Réseau IA haute performance avec les plates-formes NVIDIA Spectrum Ethernet ou NVIDIA Quantum InfiniBand.
3. Solutions de stockage flexibles qui prennent en charge une infrastructure définie par logiciel. Options Lenovo DM/NetApp. Stockage haute performance de Weka ou de DDN.

Cette approche permet aux organisations d'adapter leur infrastructure d'IA en équilibrant exigences de performances, restrictions budgétaires et besoins d'évolutivité futurs.

Lenovo adopte une approche similaire pour les topologies et les architectures réseau, jusqu'à la stack logicielle, allant des offres open source aux offres d'entreprise entièrement prises en charge (par exemple, NVIDIA AI Enterprise, NVIDIA NIM, OpenShift AI, etc.). Le cas échéant, l'entreprise fait appel aux partenaires Lenovo AI Innovator pour fournir la couche applicative déterminée par le cas d'utilisation. De plus, Lenovo a développé des services professionnels pour accompagner les clients tout au long de leur parcours d'adoption, de la découverte au déploiement et à la gestion.

À l'ère où les données sont le nouveau « pétrole » et où l'IA est la raffinerie, Lenovo a développé une stratégie complète pour maximiser la valeur pour ses clients. Cette approche englobe l'infrastructure et les outils logiciels, ce qui garantit que les données des organisations sont prêtes pour l'IA, correctement gouvernées et gérées efficacement.

Les bases de la gestion des données de Lenovo reposent sur un portefeuille de stockage diversifié, notamment :

- Solutions de stockage traditionnelles
- Stockage défini par logiciel
- Offres hautes performances de partenaires tels que Weka et DDN

Cette flexibilité s'étend au traitement des données, via la prise en charge des lacs de données open source et des solutions d'entreprise telles que Cloudera. Pour les environnements hybrides, les services professionnels de Lenovo facilitent l'intégration fluide d'offres sur site et dans le Cloud, telles que Databricks.

En s'appuyant sur ces fondations robustes en matière de données, Lenovo a réalisé d'importants progrès dans le développement de solutions d'IA d'entreprise. Ces offres sont conçues pour être fiables, écoénergétiques et de taille adaptée. Elles sont adaptées à diverses applications d'IA, notamment l'entraînement, la modélisation et l'inférence.

Les principales caractéristiques de l'infrastructure d'IA de Lenovo sont :

- Des options « as-a-service » (aaS) de TruScale pour la gestion des données, l'informatique haute performance et les GPU
- Les GPU en tant que service via des serveurs ThinkSystem (SR680a V3, SR685a et SR780a V3) prenant en charge jusqu'à huit GPU hautes performances, adaptés à l'IA d'entreprise, à l'informatique haute performance et aux charges de travail graphiques

Cette approche intégrée garantit que les clients Lenovo peuvent gérer efficacement leurs données et déployer de manière transparente des solutions d'IA puissantes, le tout au sein de solutions d'infrastructure flexibles et évolutives.

Lenovo intègre l'IA dans ses offres pour guider les clients dans leur parcours d'adoption de l'IA. L'entreprise a élargi ses services d'IA via l'ajout de nouvelles solutions telles qu'AI Discover et AI Fast Start. Celles-ci peuvent être mises en œuvre en tant que service via Lenovo TruScale, pour aider les entreprises à déployer et à faire évoluer efficacement les solutions d'IA et d'IA générative.

Au niveau de la gestion de l'infrastructure, Lenovo propose XClarity One : une plateforme Cloud hybride avec une interface conviviale pour la gestion des serveurs et des appareils à l'Edge. Cette solution intègre une assistance intelligente basée sur l'IA, qui tire parti de l'apprentissage automatique pour la maintenance prédictive et l'automatisation des tâches de routine. De plus, XClarity One dispose d'un hub de gestion sécurisé qui utilise l'IA générative pour adapter et améliorer en permanence la cybersécurité.

Pour répondre aux préoccupations environnementales, Lenovo a présenté LISSA (Lenovo Intelligent Sustainability Solutions Advisor) : un outil basé sur l'IA qui permet aux entreprises de réduire leur impact sur l'environnement dans le domaine de l'informatique. LISSA représente une nouvelle approche de la gestion du matériel informatique tout au long de son cycle de vie, en fournissant des informations exploitables qui aident les entreprises à comprendre et à gérer plus efficacement leur empreinte carbone.

Pour nous, la solution de refroidissement liquide Neptune® de 6e génération de Lenovo est un choix éprouvé et leader du marché pour répondre aux exigences uniques des charges de travail d'IA. Ce procédé avancé de refroidissement liquide surpasse de loin le refroidissement par air traditionnel, notamment en matière de dissipation thermique. De quoi garantir des performances supérieures tout en améliorant considérablement l'efficacité énergétique. En outre, les services d'alimentation et de refroidissement de Lenovo sont pensés pour aider les centres de données à gérer plus efficacement leur consommation énergétique en gérant de manière proactive les systèmes d'alimentation et de refroidissement afin de garantir un fonctionnement fiable et de réduire les risques d'arrêt.

XClarity Energy Manager de Lenovo est piloté par l'IA et optimise l'efficacité énergétique en identifiant les serveurs à faible utilisation, y compris les composants éteints ou inutilisés, et en évaluant la capacité des serveurs à prendre en charge de nouvelles charges de travail en fonction des ressources disponibles. De plus, la plateforme est adaptée à l'utilisation de l'analyse prédictive des défaillances par l'IA via l'exploitation des capacités d'assistance intelligente basées sur l'IA incluses dans Lenovo XClarity One.

Lenovo Smarter AI for All sert de fondation à l'offre de la société en matière d'IA d'entreprise, portée par un investissement d'un milliard de dollars sur trois ans, destiné à accélérer le déploiement de l'IA dans plusieurs secteurs. Cet investissement soutient la création de plus de 150 solutions d'IA en partenariat avec des éditeurs de logiciels indépendants. Lenovo propose un ensemble diversifié de plus de 90 plateformes d'IA hybrides adaptées à un large éventail de besoins dans les secteurs public, privé et personnel.

De plus, les solutions d'IA Lenovo ont remporté de nombreux prix, dont le CRN 2024 AI 100 et le CRN 2023 Product of the Year.

Le personnel adapté	Plus de 400 employés en IA dans le monde	4 centres d'IA mondiaux	182 marchés desservis	Plus de 10 consortiums sur les politiques pour guider l'avenir de l'IA	<p>Les membres de notre équipe participent à OECD AI, NAIAC, AAAI, AAAS, ACM, MLCommons, Linux Foundation et bien d'autres</p>	Plus de 1,2 milliard de dollars d'investissement promis en IA Plus de 2 milliards de dollars de revenus avec l'IA Logiciels, matériels et solutions d'IA primés depuis 2018
Les partenaires et solutions adaptés	Plus de 60 fournisseurs de solutions d'IA	Plus de 165 solutions d'IA d'entreprise	<p>NetApp AIPOD™ NVIDIA ANACONDA Data Science WS NUTANIX GPT-in-a-Box™ VMware Plateforme d'AI privée</p>	<p>Red Hat OpenShift AI Copilot Microsoft Copilot+</p>	30 000 canaux et partenaires de livraison	
L'infrastructure et les plateformes adaptées	Plus de 80 plateformes prêtes pour l'IA	N°1 en fiabilité de l'infrastructure	N°1 en fourniture de superordinateurs dans le monde	Top 3 en fourniture d'infrastructures d'IA dans le monde	N°1 en fourniture de PC, avec un portefeuille diversifié de PC avec IA	

**Le portefeuille de marque unique le plus diversifié.
Leader de l'informatique durable avec Lenovo Neptune®**

Source : Lenovo

La stratégie ODM+ de Lenovo renforce encore son offre dans le domaine de l'IA d'entreprise. Cette approche implique que Lenovo est propriétaire de ses processus de développement et de fabrication, et tire parti de sa chaîne d'approvisionnement de 20 milliards de dollars pour réduire les coûts et fournir des services de bout en bout. Le programme Lenovo AI Innovators est un écosystème de partenaires logiciels qui travaillent avec Lenovo pour fournir aux clients des solutions d'IA prêtes à être déployées. Le programme met l'accent sur la réduction des coûts des développeurs de logiciels d'IA et donne accès à des laboratoires d'IA partout dans le monde pour les tests et les bancs d'essai. Le programme aide les éditeurs de logiciels indépendants et leurs clients à déployer l'IA de façon évolutive et sécurisée.





Section 3 : conclusions et recommandations

Nous pensons que la proposition Lenovo Smarter AI for All garantit la mise en œuvre de solutions d'IA d'entreprise prêtes pour le marché. Lenovo propose le portefeuille et les services essentiels pour encourager la collaboration entre les partenaires de l'écosystème, l'expertise en IA et les éditeurs de logiciels indépendants, tous fondamentaux pour le succès de l'IA d'entreprise à l'échelle de l'organisation.

L'approche de Lenovo garantit que les clients peuvent maîtriser leurs systèmes d'IA, leur infrastructure de données et leurs environnements de déploiement, ainsi que la sécurité et la confidentialité des données, quel que soit leur emplacement géographique. Compte tenu de ces facteurs, nous partageons les recommandations suivantes à destination des organisations et des décideurs informatiques concernant l'aptitude de Lenovo Smarter AI for All à répondre aux besoins de leurs entreprises en matière d'IA.

Un portefeuille complet. Lenovo Smarter AI for All propose un large éventail de solutions d'IA, des appareils d'IA personnels à l'infrastructure d'IA d'entreprise. Les solutions d'IA d'entreprise de Lenovo sont sécurisées, efficaces et évolutives, ce qui aide les entreprises à utiliser l'IA de manière responsable.

Renforcer les avantages de l'IA d'entreprise. Les équipes de Lenovo renforcent le processus grâce au programme AI Innovators en partenariat avec des éditeurs de logiciels indépendants, afin d'obtenir des résultats commerciaux avec des applications d'IA qui intègrent des solutions de gestion et de protection des données. Elles travaillent avec des partenaires, des services et une expertise en IA capables d'accélérer le processus de planification, de stratégie et d'exécution, avec les solutions IA Fast Start qui offrent des cas d'utilisation et des projets pilotes industriels basés sur les données des organisations, ainsi qu'avec des solutions d'infrastructure d'IA écoénergétiques offrant la taille et le choix appropriés.

Avantages du partenariat avec NVIDIA. Découvrez comment l'utilisation de la plateforme logicielle NVIDIA AI Enterprise (comprenant le framework NVIDIA NeMo) et du tout nouveau design de référence de Lenovo pour l'IA générative (basé sur les LLM) permet de déployer et de commercialiser des outils et des modèles de base d'IA générative en tirant parti d'une solution pré-validée, entièrement intégrée et optimisée pour les performances. Grâce à son partenariat avec NVIDIA pour le co-développement de solutions, Lenovo est en mesure d'aider les entreprises à améliorer leurs flux de travail LLM et à maximiser les performances, la sécurité et le choix d'une taille appropriée.

Informations importantes à propos de ce rapport

CONTRIBUTEURS

Steven Dickens

Conseiller en chef en technologie | The Futurum Group

Ron Westfall

Directeur de la recherche | The Futurum Group

ÉDITEUR

Daniel Newman

PDG | The Futurum Group

DEMANDES DE RENSEIGNEMENTS

Contactez-nous si vous souhaitez discuter de ce rapport et The Futurum Group vous répondra sous les meilleurs délais.

CITATIONS

Cet article peut être cité par la presse et les analystes accrédités, mais doit être cité dans son contexte, en indiquant le nom de l'auteur, son titre et « The Futurum Group ». Les citations dans tout autre contexte nécessitent l'obtention d'une autorisation écrite préalable de The Futurum Group

LICENCES

Ce document et toute documentation à l'appui sont la propriété de The Futurum Group. Cette publication ne peut être reproduite, distribuée ou partagée sous quelque forme que ce soit sans l'autorisation écrite préalable de The Futurum Group.

DIVULGATIONS

The Futurum Group fournit des services de recherche, d'analyse, de conseil et d'assistance à de nombreuses entreprises de haute technologie, y compris celles mentionnées dans cet article. Aucun employé de l'entreprise ne détient de positions au capital d'une société citée dans ce document.



À PROPOS DE LENOVO ET NVIDIA

Lenovo ouvre la nouvelle ère de l'innovation basée sur l'IA à tous. Notre portefeuille complet offre des solutions d'IA puissantes, flexibles et responsables pour transformer les industries et donner aux individus les moyens de leurs ambitions. Nous créons un avenir d'IA plus intelligente pour tous. Chez Lenovo, nous pensons que l'avenir de l'IA implique la coexistence de l'IA publique et d'entreprise. Lenovo rend l'IA accessible à vous et à vos données.

En partenariat avec NVIDIA, les solutions d'IA hybrides sont élaborées grâce à une collaboration en ingénierie pour rendre l'IA accessible aux données des clients où et quand les utilisateurs en ont le plus besoin. De quoi contribuer à la vision de Lenovo de rendre l'IA accessible à tous et d'aider à la commercialisation de technologies et d'architectures révolutionnaires pour la prochaine génération d'IA générative. Les solutions hybrides Lenovo sont déjà optimisées pour exécuter le logiciel NVIDIA AI Enterprise : de quoi garantir une IA de production sécurisée, prise en charge et stable. En outre, elles permettent aux développeurs d'accéder aux microservices NVIDIA, y compris NVIDIA NIM et NeMo Retriever.



À PROPOS : THE FUTURUM GROUP

[TheFuturum Group](#) est une société indépendante de recherche, d'analyse et de conseil axée sur l'innovation numérique et les technologies et tendances qui révolutionnent le marché. Chaque jour, nos analystes, nos chercheurs et nos conseillers aident les dirigeants d'entreprise du monde entier à anticiper les changements majeurs dans leurs secteurs et à tirer parti des innovations disruptives pour obtenir ou maintenir un avantage concurrentiel sur leurs marchés.



COORDONNÉES

The Futurum Group LLC | futurumgroup.com | (833) 722-5337 |