



Business

Intelligence artificielle.
Valeur réelle.

Le directeur des opérations de Lily estime que leur infrastructure numérique est prête à intégrer l'IA générative. Mais Lily en connaît les limites et sait comment y répondre. Découvrez ce qu'elle a compris, comme de nombreux clients d'Orange Business.

Les rapports IA d'Orange Business

**L'impact de l'IA
sur l'infrastructure
numérique**





Résumé exécutif	3
Déploiement et niveau de préparation	4
Recommandations	
Infrastructure/Cloud	7
Recommandations – Infrastructure	
Recommandations – Cloud	
Gestion et gouvernance des données	10
Recommandations	
Sécurité	11
Recommandations	
Création de valeur	12
Recommandations	



Résumé exécutif

La quasi-totalité des dirigeants d'entreprise aspire à être à l'avant-garde, et nombreux sont ceux qui portent leur attention sur l'intelligence artificielle générative (GenAI ou IA générative), déjà utilisée activement par 85 % des organisations dans au moins une fonction métier*. John F. Kennedy disait du leadership et de l'apprentissage qu'ils étaient indispensables l'un à l'autre. Il est donc essentiel de savoir comment vos pairs abordent cette technologie de rupture.

Être à l'écoute des autres est une clé de la sagesse, c'est pourquoi nous avons demandé à GlobalData d'interroger 400 entreprises dans le monde afin d'examiner comment elles déploient la GenAI et d'évaluer son impact sur leur infrastructure numérique. Il apparaît clairement qu'un écart important subsiste entre les PoC et les services IA opérationnels à l'échelle de l'entreprise.

Opérationnaliser l'IA

Faire appel à un hyperscaler pour lancer une preuve de concept (PoC) est relativement simple. Mais faire passer cette PoC à l'échelle de l'entreprise, ce que nous appelons l'opérationnalisation, est bien plus complexe. Un écart croissant se creuse entre les attentes des entreprises et leur capacité à concrétiser leurs ambitions.

Lorsqu'il s'agit de passer à plus grande échelle, notre étude montre que les entreprises ayant une certaine maturité dans l'IA ont une vision claire des défis à relever. Elles identifient le cloud, la sécurité, l'infrastructure et la gestion des données comme les fondations essentielles à tout service IA en production. Pourtant, elles rencontrent des difficultés majeures dans chacun de ces domaines.

Ainsi, 96 % des entreprises déclarent devoir réévaluer leur stratégie de cybersécurité en raison de la GenAI. 70 % affirment la même chose au sujet de leur stratégie cloud. Moins de la moitié disposent ou disposeront de l'infrastructure IT nécessaire pour soutenir leurs déploiements GenAI. Et enfin, 51 % citent la complexité de la gestion des données comme un point de difficulté majeur.



Ce sont là des obstacles substantiels que les organisations devront surmonter pour opérationnaliser leurs services d'IA. Mais, comme le dit le proverbe : « Un homme averti en vaut deux. » Nous espérons donc que ces enseignements vous aideront à mieux vous préparer pour la suite de votre parcours IA.

* Economist Impact report "Unlocking enterprise AI: opportunities and strategies"

Déploiement et niveau de préparation

La grande majorité des entreprises interrogées explore déjà la GenAI ou prévoit de le faire dans les 12 prochains mois. Cependant, peu d'entre elles ont franchi le cap d'un déploiement à grande échelle. De fortes préoccupations subsistent concernant les fondations d'une stratégie GenAI solide, en particulier vis-à-vis de la gestion et de la gouvernance des données, de l'intégration aux systèmes existants et des risques de dépendance aux fournisseurs (vendor lock-in). Dans plus d'un cas sur deux, le déploiement est piloté par la DSI, mais les directions métiers et les membres du comité exécutif (hors CEO) y ont également une responsabilité directe.

Le constat est clair : davantage de ressources sont nécessaires. La majorité des entreprises investissent dans :

- Du personnel spécialisé : un vrai défi sur un marché en tension, où les compétences en IA sont rares
- Des ressources de calcul : ce qui explique l'envolée des coûts liés au cloud à mesure que les entreprises cherchent à développer leurs projets IA.



Ce manque de ressources freine également la capacité à créer une solution unifiée d'observabilité, ce qui entrave aussi la mise en œuvre de l'IA.



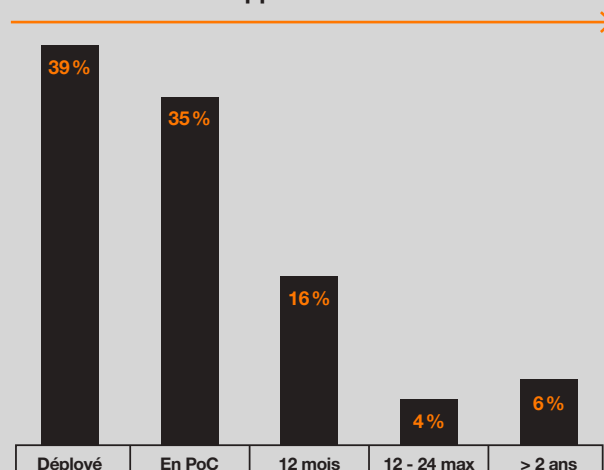
Peu entreprennent seuls leur parcours GenAI, une large minorité (42%) se tournant vers des tiers pour obtenir l'aide nécessaire : les fournisseurs de plateformes cloud/applications, les fournisseurs de services de communication et les prestataires de services informatiques sont les trois principaux partenaires vers lesquels les entreprises se tournent pour un soutien externe.

Déploiement

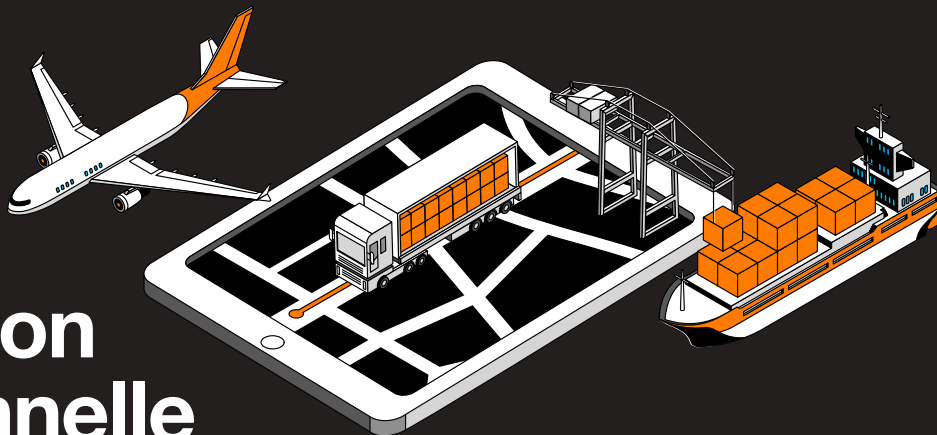
Les marchés de l'IA et de la GenAI ont enregistré une croissance rapide ces derniers mois. Au cours des deux prochaines années, 64 % des entreprises européennes prévoient de donner la priorité aux dépenses consacrées aux plateformes d'IA et 66 % prioriseront les dépenses consacrées aux plateformes d'IA générative. Cependant, si de nombreuses entreprises ont déjà lancé un premier projet, une grande partie d'entre elles en sont encore en phases d'essai ou de preuve de concept avec la GenAI.

La GenAI est évaluée dans l'ensemble de l'entreprise, avec des cas d'usage allant de l'écriture de code au support client.

Calendrier de développement de la GenAI



Niveau de préparation organisationnelle



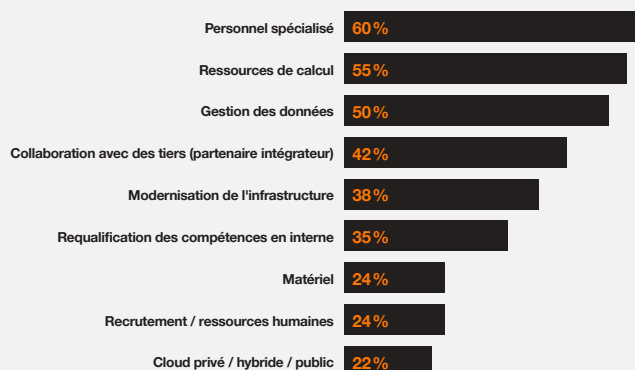
Depuis l'arrivée de ChatGPT, les entreprises s'efforcent de prendre la mesure du phénomène. Dans ce contexte, les CIO et DSI expriment des inquiétudes quant à la complexité de la GenAI. La dépendance vis-à-vis des fournisseurs est aussi un sujet d'alerte, alors que de nouvelles solutions émergent rapidement sur le marché. Pour éviter la dépendance technologique à un seul modèle fondationnel dans un marché en constante évolution, il est essentiel d'opter pour une solution qui garantit la portabilité des services IA.

Principales préoccupations liées à la mise en œuvre de la GenAI :



Pour surmonter ces obstacles et passer à l'échelle, les entreprises investissent à tous les niveaux afin de développer leurs projets GenAI dans des domaines tels que le développement des compétences, les ressources de calcul et la gestion des données.

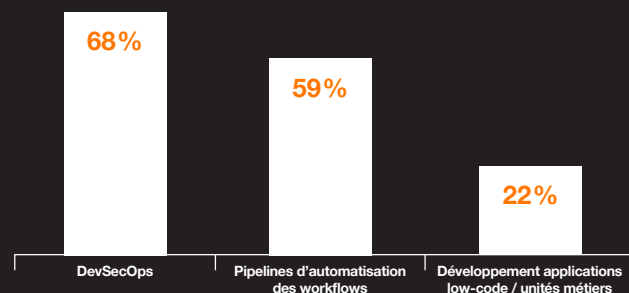
Investissements prévus pour accompagner la montée en puissance des projets GenAI :



DevOps et observabilité

Les entreprises estiment que leur visibilité sur les pratiques DevOps reste trop fragmentée, tant sur le plan technologique (sécurité, etc.) qu'opérationnel entre les équipes concernées. Cette fragmentation freine la collaboration et réduit l'efficacité des déploiements. Le manque d'expertise interne est à nouveau identifié, notamment dans le cadre de l'automatisation des workflows, le déploiement d'applications et l'observabilité.

Domaines nécessitant une meilleure collaboration



Recommandations

Vos déploiements GenAI seront évalués à l'aune de leur performance, de leur coût, de leur empreinte carbone et, surtout, de leur adéquation avec les tâches que vous devez leur confier. Le choix des modèles IA les plus efficaces et économiques doit reposer sur une évaluation rigoureuse, qui ne se limite pas aux effets de mode ou des marques. Dans certains cas, des modèles de langage de plus petite taille (small language model, SLM) adaptés à des usages spécifiques s'avéreront plus pertinents que de grands modèles généralistes.

La réussite dépendra également de votre capacité à acquérir et adapter des données d'entraînement de qualité. Cela implique d'avoir une stratégie d'intégration robuste – qui englobe le nettoyage, la transformation et la validation des données –, ainsi qu'une gouvernance adaptée.

L'entraînement des modèles nécessite d'importantes ressources. Pour maîtriser les coûts tout en assurant la fiabilité des résultats, vous pouvez :



- Opter pour du matériel éco-efficace et optimiser les algorithmes pour réduire les besoins en calcul (écoconception)
- Sélectionner des modèles réellement adaptés aux besoins pour éviter la surconsommation
- Utiliser des ressources matérielles à la demande (GPU) pour répondre aux pics lors de l'entraînement ou des inférences
- Déployer la puissance de calcul au plus près des données, afin de faciliter l'intégration de mécanismes de récupération
- Optimiser les ressources de calcul pour gérer efficacement l'intégration des processus de récupération et de génération.

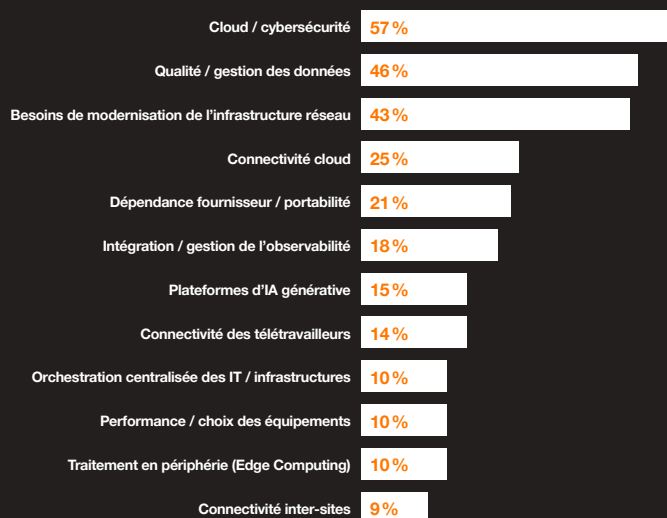
Infrastructure/Cloud

Comme indiqué précédemment, les répondants réévaluent leur stratégie en matière de cybersécurité et de cloud afin de soutenir leurs services GenAI, et modernisent leur infrastructure réseau pour les mêmes raisons.

Par ailleurs, 73 % déclarent devoir renforcer leurs capacités de stockage afin de pouvoir prendre en charge la génération augmentée de récupération (RAG). Le cloud, la cybersécurité, la qualité des données et leur gestion ressortent comme les deux préoccupations principales en matière d'infrastructure IT.

Lors de la mise en œuvre de la GenAI, les organisations expriment des inquiétudes quant à la sécurité, à la qualité et à la gestion des données, à l'infrastructure réseau ainsi qu'à la connectivité cloud. Même celles qui disposent d'une infrastructure adaptée doivent s'assurer qu'elles ont mis en place les systèmes nécessaires pour pouvoir évoluer à l'avenir.

Principaux sujets de préoccupation liés à l'infrastructure IT :



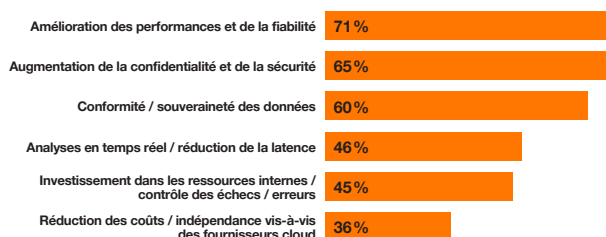
Stratégie de connectivité

Plus de 80 % des entreprises, en particulier celles qui utilisent des applications IA et GenAI pour surveiller la sécurité ou assurer le contrôle qualité, ont besoin d'une analyse de données en temps réel ou quasi temps réel. Puisque la GenAI s'appuie sur des données issues de l'ensemble du réseau d'entreprise, il est essentiel de disposer d'une connectivité fiable, sécurisée et rapide ; non seulement pour partager les données collectées, mais aussi pour pouvoir les analyser au bon endroit.

À mesure que les organisations progressent dans leur parcours IA, elles peuvent tirer parti de l'abandon des solutions basées sur le cloud au profit d'une évolution vers des solutions IA sur site ou à la périphérie du réseau. Le traitement IA en local peut permettre de répondre aux exigences de latence des applications en temps réel, de mieux gérer les risques en matière de sécurité et de conformité et de limiter les transferts de données sensibles par-delà des frontières géographiques.

L'étude révèle que plus de 65 % des entreprises ressentent un besoin fort d'exécuter la GenAI en local ou en périphérie, avec comme principaux objectifs l'amélioration des performances (70 %), le renforcement de la confidentialité (65 %) et la conformité réglementaire (60 %).

Bénéfices attendus du déploiement de la GenAI en local ou en périphérie via des modèles SLM :



Recommandations Infrastructure

Les données dont la GenAI a besoin sont générées partout et n'importe où, et les entreprises se concentrent de plus en plus sur l'analyse de ces données dans des sites décentralisés. Cela implique la mise en place d'une infrastructure réseau à haut débit et à faible latence pour faciliter le transfert rapide des données entre les sites périphériques, les sites régionaux, les installations centrales et les environnements cloud.

Vous pourriez intégrer des solutions de traitement en périphérie, qui permettent de traiter les données localement, de réduire la latence et d'activer des analyses en temps réel. Vous pourriez également envisager des approches telles que la génération augmentée de récupération (RAG), qui permet aux grands modèles de langage d'accéder à des informations provenant de sources de données externes en temps réel et de les intégrer sans devoir être entièrement réentraînés.

Le réseau doit naturellement être sécurisé, fiable et évolutif. Il convient donc de mettre en place des protocoles de sécurité robustes incluant le chiffrement et des contrôles d'accès sécurisés en périphérie, afin de protéger les données en transit. Une conception intégrant la redondance contribue à garantir la fiabilité du réseau, tandis que l'allocation de bande passante et la définition de niveaux de service minimaux permettent d'assurer la performance et l'évolutivité.

Pour aller plus loin en matière de flexibilité, de scalabilité et de résilience, l'adoption de stratégies multicloud constitue également un levier pertinent, tout en réduisant les risques de dépendance fournisseur.

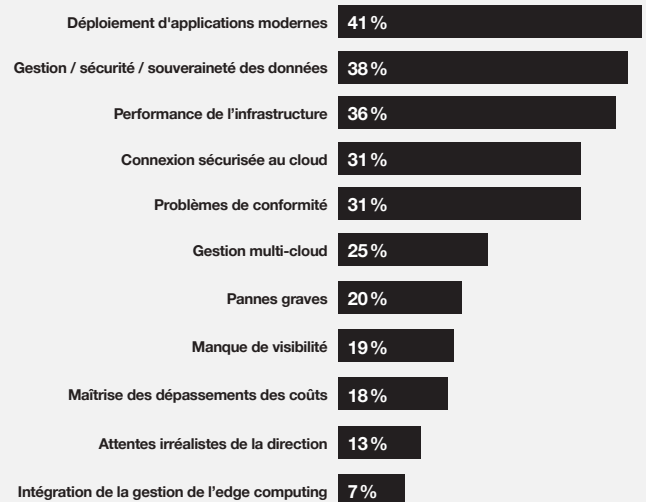


Stratégie cloud

De nombreuses entreprises débutent leur parcours IA, et en particulier GenAI, avec des solutions IA en cloud, qui exigent une connectivité robuste. Le cloud joue donc un rôle central dans le déploiement à l'échelle de l'IA. Pour soutenir des environnements d'IA multi-modèles, les déploiements cloud doivent être efficaces, sécurisés, correctement provisionnés et économiquement viables.

Comme indiqué précédemment, de nombreuses organisations rencontrent des difficultés lorsqu'il s'agit d'aligner leur stratégie cloud avec les déploiements GenAI qu'elles envisagent. Elles doivent notamment intégrer leurs systèmes existants, gérer les questions de sécurité, de confidentialité et de souveraineté des données, tout en maintenant un niveau de performance élevé de leur infrastructure. Nombre d'entre elles revoient actuellement leur stratégie cloud, pour répondre à ces problématiques mais aussi à l'explosion des coûts liée à la GenAI et aux exigences de performance des modèles LLM.

Difficultés liées à la gestion des données :



Stockage

Les besoins en stockage augmentent à mesure que les entreprises adoptent la RAG pour affiner les résultats produits par les LLM et qu'elles exploitent davantage de données non structurées. 73 % des répondants estiment devoir renforcer leur capacité de stockage pour pouvoir tirer pleinement parti de la RAG, qui s'appuie sur des bases de données vectorielles et des embeddings (intégrations). En parallèle, 65 % indiquent devoir procéder à la même mise à niveau, du fait de la capacité de la GenAI à faire émerger de nouveaux insights à partir de sources non structurées telles que des fichiers audio ou des images. Dans les deux cas, le constat est le même : un besoin d'investir dans l'évolution des capacités de stockage, avec les coûts qui y sont associés.



Recommandations – Cloud

La plupart des entreprises ont adopté une stratégie multi-cloud qui accroît la complexité, les risques de sécurité et les défis opérationnels. Les déploiements de la GenAI ne feront qu'accentuer ces problèmes. Il est donc crucial pour les entreprises de repenser leur stratégie cloud.

Comme pour les réseaux, les priorités sont la sécurité, la fiabilité et la scalabilité. Parmi les actions possibles :

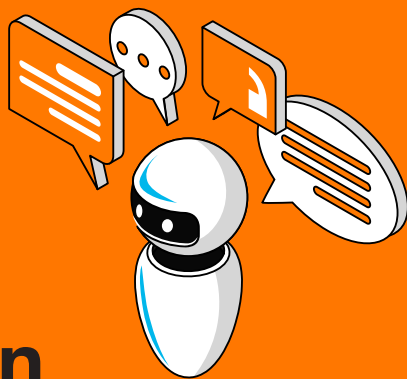
- Optimisation de l'usage des GPU pour maîtriser les coûts élevés et garantir leur disponibilité en s'assurant à l'avance l'accès au matériel spécialisé.
- Mise en place d'une sécurité multi-cloud robuste, de mesures de contrôle des accès, et des protocoles de cybersécurité avec une gestion de la sécurité multi-cloud.
- Utilisation de solutions de stockage évolutives offrant un accès rapide et préservant l'intégrité des données, avec des contrôles d'intégrité et des sauvegardes réguliers, adaptées à votre cloud privé, à votre périphérie ou à votre hyperscaler.
- Adoption d'architectures évolutives avec allocation dynamique des ressources pour garantir performance et efficacité lors des déploiements à grande échelle.

Pour que cette nouvelle stratégie soit efficace, il est essentiel de mettre en place une infrastructure réseau optimisée, capable de faire circuler les applications entre différents environnements cloud. Cela suppose une architecture de bout en bout et des opérations unifiées pour la gestion des réseaux dans plusieurs environnements cloud et au sein même des clouds.

Afin de mieux maîtriser les coûts, il peut être pertinent d'envisager l'externalisation de la gestion et du contrôle du réseau et de la connectivité des données vers une infrastructure cloud dédiée.

Les enjeux de conformité et de gouvernance sont centraux dans le domaine du cloud. Il est nécessaire de mettre en œuvre des processus rigoureux de nettoyage des données et des techniques de réduction des biais pour garantir la fiabilité des résultats produits par les modèles et mettre en place une gouvernance des données et des documents. Le respect des réglementations en matière de souveraineté des données doit être assuré, ce qui peut passer par le déploiement de l'IA dans différents emplacements de données.

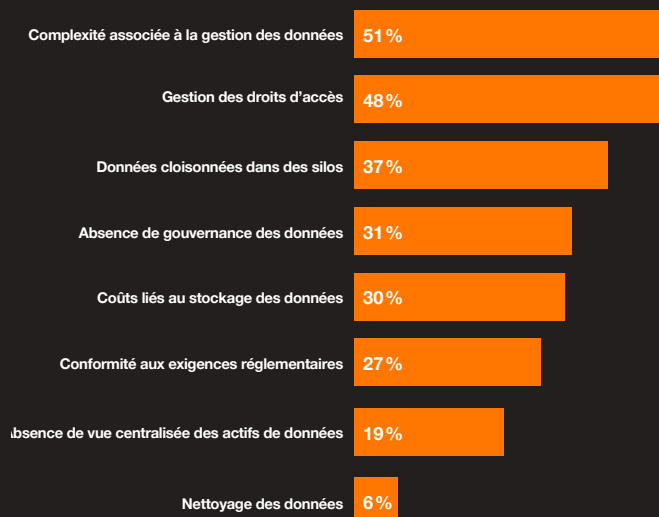
Gestion et gouvernance des données



La qualité des données n'est pas qu'un atout : c'est une condition essentielle. Elle a été citée comme étant le facteur de satisfaction le plus déterminant dans un projet GenAI (51 %) et comme la troisième cause d'insatisfaction en cas de projet mal implémenté (38 %). La qualité des données est donc le principal facteur de réussite, ou d'échec, d'un projet GenAI. Ce n'est pas un enjeu à traiter a posteriori : l'expérience des clients d'Orange Business montre qu'il est au moins 20 % plus coûteux de corriger des problèmes de qualité de données après l'entraînement d'un modèle que de partir sur des données propres dès le début.

La gouvernance des données, en particulier sur les questions de propriété intellectuelle et de confidentialité, constitue à la fois une préoccupation sociétale majeure et un défi opérationnel pour les entreprises. L'émergence de la GenAI a mis en lumière l'importance d'une stratégie de gestion et de gouvernance des données capable de répondre aux nouvelles exigences de contrôle et de complexité introduites par l'IA.

Principales difficultés liées à la gestion des données :



Recommandations

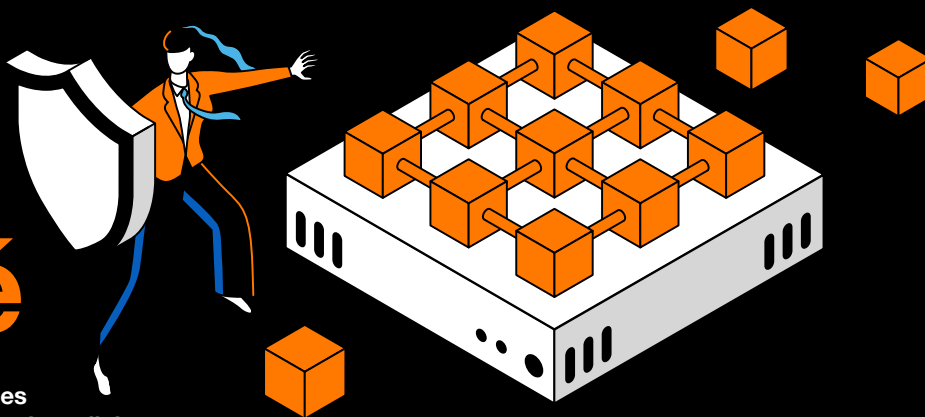


Une visibilité de bout en bout est indispensable pour garantir un accès aux données tout en assurant le plus haut niveau de sécurité. La gouvernance des données est donc essentielle. Dans le cadre d'une IA performante, il est crucial de définir des règles claires de responsabilité et de traitement des données afin d'assurer la conformité aux réglementations en vigueur. Intégrer cette stratégie data au cœur du plan d'entreprise permet de garantir la qualité, l'accessibilité et la sécurité des données.

Ces principes s'appliquent également aux données non structurées (les documents par exemple), qui sont essentielles pour l'affinage des modèles ou la réussite de la réalisation d'une RAG. Il convient de mettre en œuvre des mécanismes de contrôle d'accès, mais aussi une gouvernance documentaire qui définisse des critères de qualité des documents. On verra sans doute émerger des rôles dédiés tels que « document engineer » ou « document scientist ».

Le renforcement de politiques sur la sensibilité des données et leur cycle de vie permet de réduire les risques d'erreurs ou d'atteinte à la confidentialité. Au global, réussir sa transformation vers une organisation pilotée par les données suppose d'adopter une culture centrée sur la donnée, ce qui inclut la formation des collaborateurs à l'usage de la GenAI, et de les guider sur quand et comment l'intégrer dans leurs pratiques professionnelles.

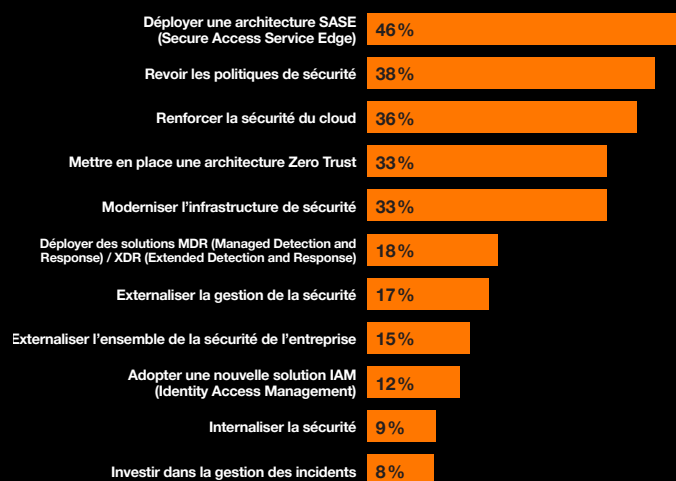
Sécurité



La protection des données d'IA lorsqu'elles circulent sur le réseau de l'entreprise est primordiale. Il est nécessaire de l'intégrer dès la conception, ce ne peut se faire a posteriori. Protéger séparément les données et l'infrastructure, puis tenter de combiner ces éléments dans un service IA, ne permettra pas d'atteindre le niveau de sécurité dont vous avez besoin. Et il est toujours moins coûteux de sécuriser dès le départ que de corriger les conséquences d'une faille.

Sans surprise, les architectures SASE (Secure Access Service Edge) et Zero Trust, qui permettent de renforcer la sécurité des services IA via un contrôle strict des accès et une supervision de l'activité des utilisateurs sur l'ensemble du réseau, quels que soient le terminal ou la localisation, sont les approches les plus fréquemment adoptées par les répondants souhaitant mettre à jour leur stratégie de sécurité.

La majorité des entreprises interrogées revoient leur stratégie cybersécurité en lien avec l'usage actuel ou projeté de la GenAI. Celles qui s'appuient sur des solutions open source expriment une inquiétude particulière. Un manque d'expertise, des contraintes budgétaires et des difficultés de gestion des données sont autant de défis à relever.



Recommandations

L'un des principaux atouts de la GenAI est sa rapidité d'action. Mais cette vélocité induit aussi une propagation potentielle plus rapide, et à grande échelle, des problèmes dans l'environnement interconnecté de l'entreprise. C'est pourquoi, dans la révision de votre stratégie de cybersécurité, vous devez inclure la mise en place de contrôles d'accès sécurisés aux données et aux déploiements GenAI, pour protéger les informations sensibles. Une posture de sécurité rigoureuse et de strictes procédures de gouvernance sont également essentielles pour limiter les risques de fuite de données et d'autres menaces.

Ces protections doivent être déployées de manière à ne pas restreindre les bénéfices potentiels de la GenAI. Trouver cet équilibre suppose :

- Une architecture Zero Trust, fondée sur des technologies rationalisées et des règles utilisateurs cohérentes
- Des politiques de sécurité centralisées à l'échelle du réseau, pour simplifier leur mise en œuvre, leur mise à jour et leur pilotage, et ainsi garantir une posture de sécurité homogène dans toute l'organisation
- Un réseau sécurisé, distribué et flexible, s'appuyant sur une solution Secure Access Service Edge (SASE) combinant une connectivité cloud à la demande et le Security Service Edge (SSE). Le SSE est un sous-ensemble du cadre SASE que Gartner définit comme « un ensemble de services de sécurité cloud protégeant l'accès aux applications, aux services cloud et au web ».
- Utiliser l'approche cloud-native de SASE pour évoluer rapidement et globalement.

Il convient également de mettre en place une supervision continue et des protocoles de réponse rapide en cas de violations de la sécurité. Il est essentiel de s'appuyer sur des expertises en cybersécurité et des solutions de défense avancées pour prévenir la corruption des données et limiter les comportements imprévus ou les biais susceptibles d'affecter les modèles.

Création de valeur



Le déploiement de la GenAI représente un investissement stratégique, mais c'est bien le rapport coût/bénéfice qui reste, pour la majorité des entreprises, l'un des principaux indicateurs de réussite. Si une majorité des répondants estiment que leurs projets répondent à cet objectif, seuls 21 % sont convaincus que leurs investissements dans l'IA génèrent un véritable retour sur investissement. Par ailleurs, si plus d'un tiers (37 %) peinent à percevoir concrètement les bénéfices de leurs investissements technologiques, les répondants restent toutefois optimistes quant à leur capacité à mesurer efficacement le retour sur investissement de leurs déploiements à l'avenir.

Plus de la moitié des entreprises se déclarent satisfaites de leurs projets GenAI, mais les échecs restent fréquents. Ceux-ci s'expliquent souvent par un manque d'expertise en interne ou par des difficultés à intégrer les nouveaux services d'IA dans les environnements IT existants. La mise en place d'une stratégie de gestion des données robuste est une condition essentielle à la réussite, tout comme le déploiement d'une stratégie cloud bien structurée qui intègre une sécurité solide.



Motifs d'insatisfaction :



Recommandations



Pour optimiser le retour sur investissement des projets GenAI, il est crucial de veiller à ce que ses solutions soient exploitées pleinement. Comme pour tout autre investissement technologique, la réussite de la GenAI repose sur un fort taux d'adoption par les utilisateurs.

Il est donc essentiel de développer des programmes de formation spécifiques pour renforcer les compétences internes, afin de permettre aux utilisateurs de tirer pleinement parti des outils GenAI. Il est également essentiel de s'assurer que les équipements existants sont adaptés : rien ne freine davantage l'adoption que des dispositifs informatiques incompatibles avec les solutions GenAI.

Les coûts restent un sujet de préoccupation important. Pour améliorer la rentabilité, il peut être judicieux de s'éloigner des modèles d'abonnement classiques pour l'accès à la connectivité cloud. Par ailleurs, la mise en œuvre de technologies permettant l'orchestration de modèles d'IA sur différentes plateformes cloud peut favoriser le suivi des coûts et renforcer l'interopérabilité.