



Hyperscalabilité IA

Maîtriser une trajectoire de progression rapide

www.keyrus.com

La trajectoire vers une IA véritablement scalable

L'intelligence artificielle n'est plus une simple tendance : elle constitue aujourd'hui un levier incontournable de performance et d'innovation pour les entreprises. Pourtant, les outils tels que Copilot, bien qu'impressionnants, ne représentent qu'une étape initiale dans la trajectoire vers une IA véritablement scalable et industrialisée. L'enjeu ne réside pas uniquement dans l'adoption de ces technologies, mais dans leur intégration stratégique au sein d'écosystèmes plus larges.

Chez Keyrus, nous croyons fermement que la scalabilité de l'IA nécessite de penser grand dès le départ, en structurant rigoureusement les cas d'usage et en adoptant des plateformes hyperscalers comme Microsoft Azure. Ces technologies permettent la création de parcours de capacité adaptés aux besoins métiers, garantissant une vision globale dès les premières initiatives et renforçant ainsi l'efficacité et la portée des solutions déployées.

Notre ambition est de guider nos clients au-delà des limitations des POC, en les accompagnant vers des déploiements robustes et différenciants.

Cet eBook est une invitation à explorer ces concepts et à découvrir comment une stratégie bien pensée peut transformer les promesses de l'IA en réalité tangible, en permettant aux entreprises de réduire les coûts, d'améliorer l'efficacité des processus et de créer des occasions de croissance durable, et bien réelles.

Bonne lecture !

Amram Azoulay
Directeur conseil data & IA, Keyrus



Hyperscalabilité IA

Maîtriser une trajectoire de progression rapide

Les avancées rapides de l'intelligence artificielle transforment en profondeur les processus opérationnels et stratégiques des organisations. Les progrès récents de l'intelligence artificielle montrent que plus de la moitié des entreprises utilisant l'IA observent une amélioration significative de leurs processus décisionnels, illustrant ainsi son impact concret sur la performance organisationnelle.

Parmi ces innovations, des outils comme Copilot promettent d'automatiser des tâches complexes et d'améliorer la productivité des collaborateurs. Cependant, ces solutions, bien qu'efficaces pour des besoins ponctuels, peinent à évoluer à grande échelle en raison d'une absence d'intégration cohérente aux systèmes existants et d'un manque de vision stratégique globale.

L'industrialisation de l'IA nécessite bien plus qu'une simple adoption technologique. Elle implique une structuration des cas d'usage, l'intégration de solutions modulaires – telles que des composants technologiques interopérables permettant une personnalisation rapide et une maintenance simplifiée – et une stratégie globale orientée vers la création de valeur.

Ainsi, ce contexte soulève des enjeux majeurs : comment dépasser les limites des outils initiaux pour déployer des solutions IA robustes et déployables à l'échelle (scalables) ? Quels rôles jouent les hyperscalers dans cette transformation, et comment tirer parti de leur écosystème ?

Ces questions sont au cœur des réflexions actuelles sur l'avenir de l'IA en entreprise. Questions auxquelles cet ebook tenter d'apporter quelques éléments de réponses, pour éclairer la réflexion.



Images générées par Midjourney avec la consigne :
« Hyper realistic portrait of a woman in a modern office environment, empowering her tasks with an artificial intelligence assistance. --chaos 10 --ar 3:4 --style raw --weird 300 --v 6.7. »

Un point de départ prometteur mais insuffisant

Microsoft Copilot s'impose comme un objet d'étude exemplaire pour analyser l'impact de l'IA dans les environnements professionnels. Bien qu'ancré dans l'écosystème Microsoft, les principes et observations qui en découlent s'appliquent largement à tout assistant ou brique IA capable de simplifier des processus clés, et parfois même d'automatiser des tâches complexes. Dans des environnements bureautiques, ces outils permettent d'améliorer la productivité des collaborateurs en intégrant des fonctionnalités d'assistance directement dans les outils du quotidien ; nous avons par exemple tous entendu parler de telles ou telles fonctions soudainement « assistées par l'IA », de Pierre au marketing, à Sarah à la comptabilité en passant par Manuel et Ingrid à la logistique et aux achats. Cette approche offre des gains immédiats en termes d'efficacité opérationnelle (moins de temps pour plus de précision-complexité), notamment dans des tâches telles que la rédaction automatisée, la gestion des courriels ou encore la création de tableaux de bord interactifs, ... la liste est longue.

Les limites des solutions initiales

Malgré tout son potentiel, Copilot n'est pas exempt de limitations. Les Proofs of Concept (POC), bien qu'essentiels pour valider la faisabilité technique d'un projet, et malgré les trésors de louanges des plus enthousiastes, peinent souvent à être transformés en déploiements opérationnels à grande échelle.



Les causes principales incluent sont multiples, et incluent la plus part du temps une fragmentation des outils, une dépendance aux silos de données et une absence de stratégie intégrée pour connecter Copilot aux processus métiers complexes. Ces obstacles freinent l'industrialisation des cas d'usage et limitent leur portée dans un cadre organisationnel plus large.

Les enjeux d'une stratégie à grande échelle

Mais il n'y a point de fatalité. Pour dépasser ces limitations, il est possible de développer une approche stratégique, fondée sur une vision globale des objectifs et des besoins métiers. Cela implique de structurer les cas d'usage autour de modules technologiques modulaires et scalables.

Une telle approche permet non seulement de répondre aux besoins immédiats, mais aussi de préparer les organisations à des évolutions futures, en posant les bases d'une transition stratégique durable.

En exploitant des plateformes hyperscalers telles que Microsoft Azure, Google Cloud ou AWS, pour n'en citer que trois, les entreprises peuvent concevoir des solutions robustes, capables de s'adapter à des contextes variés tout en optimisant les coûts et les ressources.

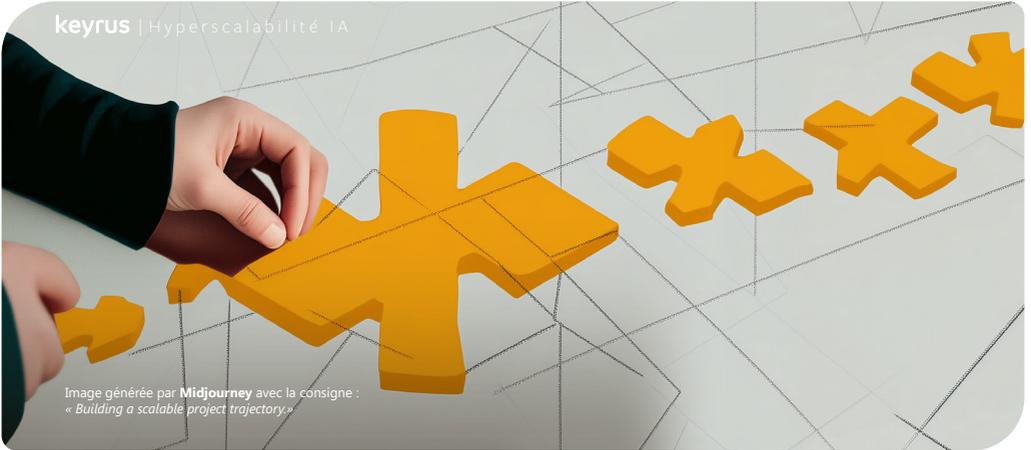


Image générée par Midjourney avec la consigne :
« Building a scalable project trajectory. »

Construire une trajectoire hyperscalable

Dans le domaine de l'intelligence artificielle, les hyperscalers (nouveau buzzword !) offrent des infrastructures robustes et adaptables qui intègrent des capacités avancées telles que les Large Language Models (LLM) et des outils d'analyse en temps réel, facilitant la création de solutions IA évolutives. Par exemple, Microsoft se distingue grâce à son partenariat stratégique avec OpenAI, renforçant ses capacités en IA avancée. Google, de son côté, propose une offre complémentaire axée sur l'optimisation des flux de données et la performance des algorithmes.

Ces plateformes permettent aux entreprises de concevoir des solutions IA à grande échelle en s'appuyant sur des ressources hautement disponibles et sécurisées.

La distinction entre ces acteurs permet aux organisations de choisir l'approche la plus adaptée à leurs besoins métiers spécifiques, qu'il s'agisse d'intégration rapide, de capacités analytiques avancées ou d'optimisation des coûts. En capitalisant sur ces capacités, les organisations peuvent accélérer le passage de concepts initiaux à des déploiements opérationnels.

On vous a perdu avec ce jargon théorique ? En fait, ce n'est pas si compliqué. Accrochez-vous, vous allez bientôt tout comprendre !

Définir la trajectoire

Une trajectoire hyperscalable repose sur le concept de Capability Pathway – une méthodologie qui structure les cas d'usage par familles de capacités modulaires. Ce cadre permet de connecter des modules technologiques complémentaires pour répondre à des besoins spécifiques.

Par exemple, un cheminement dans la gestion logistique pourrait inclure la prédiction des stocks, l'optimisation des itinéraires de livraison et l'automatisation des commandes. Chaque module technologique agit comme un bloc interconnecté dans une architecture globale, facilitant ainsi l'adaptabilité et l'évolutivité des solutions.

Les Business Apps Accélérateurs constituent un autre élément clé de l'hyperscalabilité. Ces solutions pré-packagées, construites sur des plateformes comme Microsoft Power Apps ou Azure Logic Apps, permettent de répondre rapidement à des besoins métiers spécifiques tout en garantissant une intégration fluide avec les systèmes existants.

Par exemple, une application de recommandations personnalisées pour un site e-commerce peut être déployée rapidement en combinant des modèles d'IA existants avec des connecteurs de données prêts à l'emploi. Cette approche réduit les temps de développement tout en maximisant l'efficacité opérationnelle.

En s'appuyant sur ces stratégies et outils, les entreprises peuvent construire des trajectoires hyperscalables, capables de transformer les initiatives IA en moteurs de croissance durable, à l'échelle.

“ Les **Business Apps Accélérateurs** constituent un autre élément clé de l’hyperscalabilité. Ces solutions pré-packagées, construites sur des plateformes comme Microsoft Power Apps ou Azure Logic Apps, permettent de **répondre rapidement à des besoins métiers spécifiques** tout en garantissant une intégration fluide avec les systèmes existants. ”

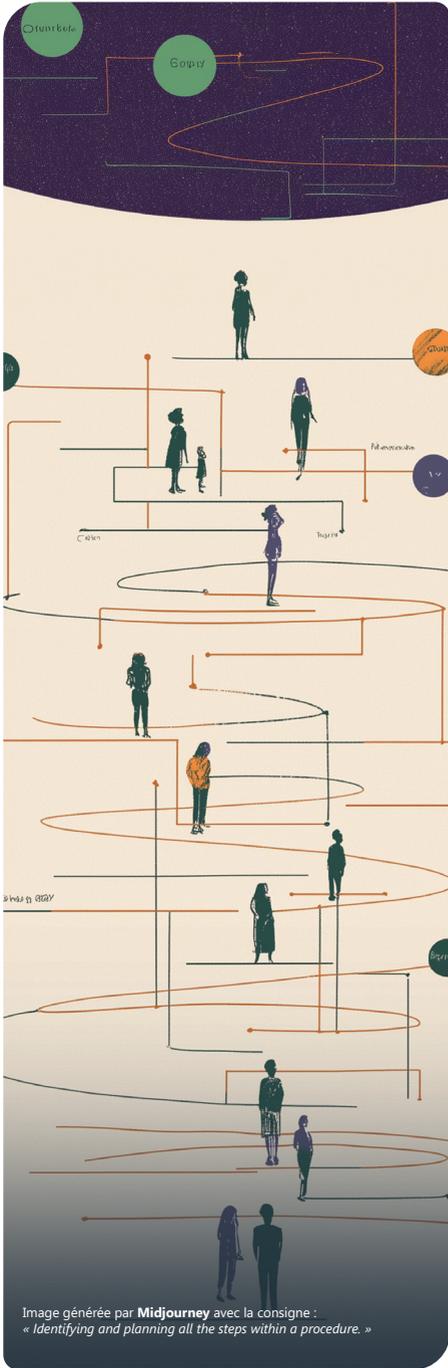


Image générée par Midjourney avec la consigne :
« Identifying and planning all the steps within a procedure. »

Identifier les étapes

De plus, passer d'un POC à un déploiement à grande échelle nécessite une analyse approfondie des cas d'usage afin d'identifier les étapes nécessaires pour les convertir en solutions opérationnelles.

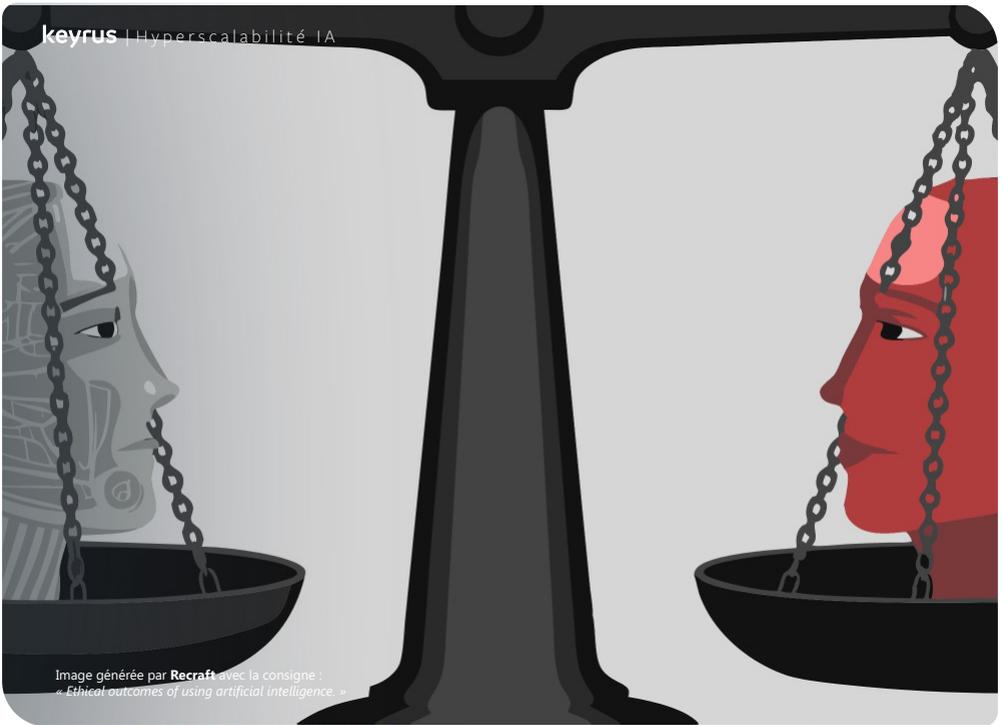
Par exemple, un cas d'usage en marketing personnalisé pourrait commencer par l'implémentation d'un moteur de recommandations dans une région pilote, suivi d'un déploiement progressif à travers d'autres régions ou segments de clientèle. Chaque étape s'appuie sur des retours d'expérience précis pour affiner et ajuster la solution.

Un autre cas d'usage pertinent se situe dans la gestion des ressources humaines. L'automatisation de la présélection des candidatures peut initialement être testée sur un périmètre restreint, comme un service spécifique, pour valider les algorithmes de scoring et garantir l'équité dans les recommandations. Une fois le modèle ajusté, il peut être étendu à l'ensemble des recrutements, réduisant le temps nécessaire au tri des candidatures tout en améliorant la diversité et la pertinence des profils sélectionnés.

Dans le domaine de la logistique, nous l'avons évoqué, un cas d'usage pourrait consister à optimiser les itinéraires de livraison à l'aide de l'IA. Un pilote peut être lancé sur une zone urbaine à forte densité pour identifier les points d'amélioration liés aux délais et aux coûts. Les données collectées permettent ensuite d'ajuster les modèles avant d'étendre le système à une couverture nationale ou internationale, améliorant ainsi significativement l'efficacité des opérations logistiques.

Optimiser les coûts et les ressources

L'une des clés du succès dans une stratégie hyperscalable est l'optimisation des ressources financières et technologiques. Cela implique une gestion rigoureuse des coûts à chaque étape du projet, de la phase initiale de prototypage jusqu'au déploiement global. Les entreprises doivent également évaluer les besoins en infrastructure et en personnel, en s'assurant que les ressources disponibles sont alignées avec les objectifs stratégiques.



Par exemple, l'utilisation de solutions cloud flexibles comme Azure peut réduire considérablement les investissements initiaux en infrastructure tout en offrant une scalabilité adaptée aux besoins croissants des entreprises.

De même, le recours à des plateformes de gestion centralisée comme Snowflake facilite la consolidation des données en temps réel, réduisant les coûts liés aux silos de données et améliorant la prise de décision stratégique.

Dans le domaine de la maintenance prédictive, des outils comme IBM Maximo permettent d'identifier et de résoudre les problèmes d'équipement avant qu'ils ne surviennent, réduisant ainsi les temps d'arrêt et les coûts opérationnels.

Par ailleurs, l'automatisation des flux de travail, avec des solutions comme UiPath, libère les équipes des tâches répétitives et chronophages, leur permettant de se concentrer sur des activités à forte valeur ajoutée.

Enfin, des plateformes comme Tableau ou Power BI fournissent des capacités de visualisation avancées, rendant les analyses complexes accessibles aux équipes non techniques et améliorant ainsi l'agilité organisationnelle.

La transition d'un POC à un déploiement à grande échelle exige donc plus qu'une simple mise en œuvre technique. Elle repose sur une planification stratégique approfondie, une maîtrise des coûts, et une capacité à aligner les ressources avec des objectifs clairs et mesurables.

Cette transition repose également sur l'utilisation judicieuse des plateformes technologiques, des outils d'automatisation et des solutions de visualisation. En structurant méthodiquement chaque étape, les organisations peuvent transformer des idées prometteuses en leviers d'innovation pérennes, posant ainsi les bases d'une transformation à grande échelle.

“ La transition d’un POC à un déploiement à grande échelle exige donc plus qu’une simple mise en œuvre technique. Elle repose sur **une planification stratégique** approfondie, **une maîtrise des coûts**, et **une capacité à aligner les ressources avec des objectifs** clairs et mesurables. ”

Au-delà de la scalabilité

L'évolution rapide des technologies d'intelligence artificielle ouvre des opportunités sans précédent, mais impose également de nouvelles attentes. Les entreprises doivent anticiper l'émergence de nouveaux paradigmes technologiques, comme l'intégration croissante de l'IA générative ou l'expansion des capacités d'automatisation dans des domaines encore sous-exploités, tels que la personnalisation avancée dans la santé ou l'optimisation en temps réel dans les chaînes logistiques.

Parallèlement, des innovations comme les modèles fondés sur l'IA explicable (XAI) et l'interopérabilité étendue des systèmes promettent de réduire les risques tout en augmentant la confiance dans les décisions automatisées. Une stratégie durable repose sur la capacité à identifier les futurs besoins métiers et à ajuster les trajectoires IA en conséquence, tout en assurant une flexibilité pour s'adapter aux disruptions imprévues et aux exigences réglementaires évolutives.

La montée en puissance de l'intelligence artificielle pose également des questions fondamentales sur ses impacts sociaux et éthiques.

Les systèmes de reconnaissance faciale, par exemple, soulèvent des préoccupations majeures en matière de vie privée et de surveillance de masse, et les entreprises doivent désormais s'engager dans des pratiques responsables, en intégrant des audits éthiques et en garantissant une gouvernance claire pour éviter les dérives potentielles.

Désormais, il faut aborder des sujets tels que la transparence des algorithmes, la protection des données personnelles, et l'équité dans les prises de décision automatisées. Adopter une approche responsable signifie intégrer ces considérations dès la conception des solutions IA, en impliquant des parties prenantes variées pour assurer un équilibre entre innovation technologique et acceptabilité sociétale.



Image générée par Lexica Aperture v5 avec la consigne :
« Harmonising strategy and technical innovation. »

Harmoniser vision stratégique et innovation technologique

L'intelligence artificielle s'impose comme une technologie révolutionnaire, mais sa véritable valeur réside dans sa capacité à s'intégrer de manière stratégique et scalable dans les processus organisationnels. Les défis liés à l'industrialisation de l'IA, ne sauraient se résumer à la nécessité de dépasser les POC pour atteindre des déploiements robustes et efficaces ; ce n'est qu'une première étape. Étape que viendront compléter outils et méthodologies permettant de structurer des trajectoires hyperscalables, tout en anticipant les évolutions futures et les enjeux éthiques.



Image générée par **Midjourney** avec la consigne :
« Hyper realistic portrait of a robot copilot, in the
attire of an airplane or race car copilot. --chaos
10 --ar 3:4 --style raw --weird 300 --v 6.1 »

Chez Keyrus, nous sommes convaincus que la réussite des projets d'IA repose sur l'alliance entre une innovation technologique audacieuse et une approche stratégique sur mesure. En tant que partenaire de confiance, nous accompagnons nos clients dans cette transformation en proposant des solutions adaptées à leurs besoins spécifiques, tout en assurant une vision à long terme. De la structuration des cas d'usage à leur mise en œuvre opérationnelle, nous apportons une expertise pointue pour garantir des résultats concrets et durables.

En collaborant avec Keyrus, les organisations peuvent tirer pleinement parti des occasions offertes par l'intelligence artificielle, tout en relevant les défis complexes de la scalabilité et de l'éthique.

**Article co-écrit par keyrus,
ChatGPT4^o et Mistral, Claude, Copilot,
Perplexity et Gemini**

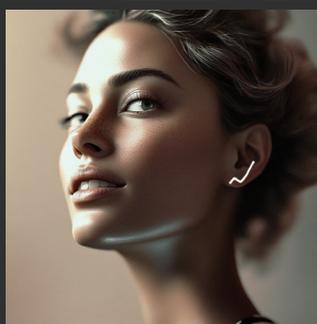
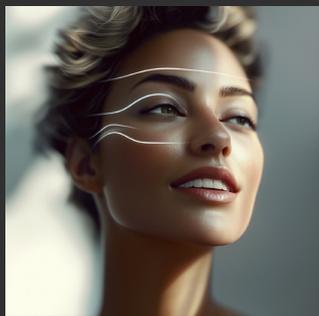
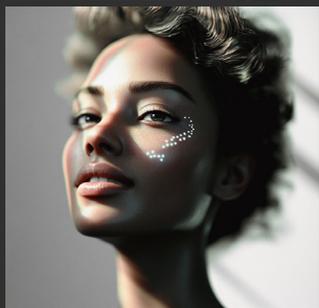
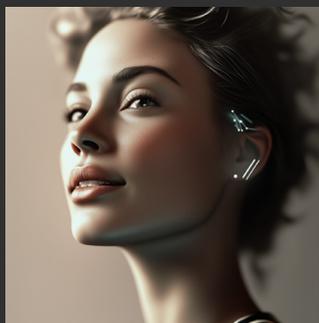
Les unes auxquelles vous avez échappé...

Lors de la création de nos ebooks, chaque image naît d'un processus itératif avec le moteur d'IA. Des allers-retours répétés permettent d'affiner chaque visuel, jusqu'à atteindre l'expression la plus fidèle de l'idée souhaitée. Mais le travail ne s'arrête pas là : parmi les images générées, il faut ensuite opérer des choix. Ce processus de sélection repose sur un effort d'objectivation, où il s'agit de mettre de côté sa propre subjectivité pour respecter au mieux le « cahier des charges ». Une démarche délicate, tant les images peuvent inspirer, surprendre, voire fasciner, mais essentielle pour rester fidèle à l'intention initiale.

L'atelier du prompteur...

Images générées par
MidJourney avec la

consigne : « *Hyper realistic portrait of a woman in a modern office environment, empowering her tasks with an artificial intelligence assistance. --chaos 10 --ar 3:4 --style raw --weird 300 --v 6.1* »



Vous avez trouvé cette lecture utile ?

Vous aimerez sûrement aussi :

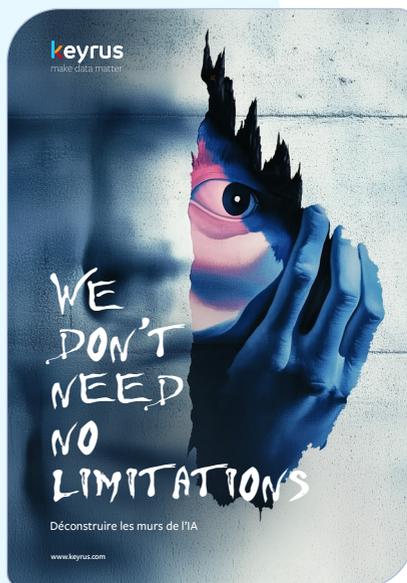
We don't need no limitations

Déconstruire les murs de l'IA

We Don't Need No Limitations explore les barrières qui freinent le développement et l'intégration de l'intelligence artificielle, à travers la métaphore des murs. S'appuyant sur la théorie des cinq murs de l'IA de Bertrand Braunschweig (confiance, consommation énergétique, sécurité des systèmes, interactions homme-machine, inhumanité des machines), il ajoute un sixième obstacle : la limitation des données d'apprentissage. Le document analyse ces barrières sous l'angle actif/passif : certaines sont intentionnellement érigées pour protéger, d'autres émergent des limites techniques. LeBook invite à une réflexion sur la démolition consciente de ces murs pour intégrer l'IA de manière harmonieuse, équilibrée et éthique.

Quelles sont les 3 principales idées ?

1. Les murs de l'IA comme obstacles : les six murs (confiance, consommation énergétique, sécurité, interactions homme-machine, inhumanité, limitation des données) freinent le développement et l'intégration de l'IA.
2. Distinction actif/passif : certains murs sont consciemment érigés pour protéger, tandis que d'autres émergent des limites techniques ou contextuelles.
3. Démolition consciente : surmonter ces barrières nécessite innovation, régulation et éthique pour intégrer l'IA de manière équilibrée et au service de l'humain.





Tech Advisory Board

Tech Advisory Board TAB

Conçu pour éclairer les orientations technologiques de Keyrus en s'appuyant sur des convictions fortes et des offres de pointe, notre TAB incarne une vision ambitieuse et innovante au service de l'excellence technologique.

Le TAB, c'est qui ?

Il est aujourd'hui composé de 9 experts, qui couvrent de concert toute la chaîne de valeur de la donnée : de l'architecture à l'ingénierie, l'analyse, la gouvernance, la gestion, jusqu'à la stratégie des données. Et évidemment notre ancrage dans la data science avec nos experts en Intelligence Artificielle.

Les membres du TAB : **Van Tuan Dang, Cao Tri Do, IkmeI Hamdi, Nassim Nalla, Gilles Dansou, Benoit Domas, Jonathan Meklat, Patricio Touchard** et **Ali Hadji**.

Le TAB, c'est quoi ?

C'est LE board réunissant les leaders Tech de Keyrus France ! Il incarne l'expertise collective de nos Tech Leads en France pour positionner Keyrus comme un acteur de référence en termes de technologie et d'innovation.

Le TAB, pourquoi ?

- Participer à la co-construction de nos offres innovantes
- Transformer les technologies disruptives en cas d'usage à valeur
- Développer l'excellence de réalisation de nos équipes
- Challenger nos clients dans leurs orientations et choix technologiques

Le TAB, pour qui ?

Pour nos clients et partenaires dans l'écosystème Keyrus France ! Parce que notre objectif est de toujours vous offrir une expérience technologique unique, le TAB est à vos côtés pour vous aider à naviguer avec succès dans un paysage technologique en constante évolution.

keyrus

make data matter

Acteur international du conseil et des technologies, Keyrus a pour mission de donner du sens aux données, en révélant toute leur portée, notamment sous un angle humain.

Parce que ce ne sont pas tant les données elles-mêmes qui importent, mais les opportunités que nous pouvons développer en les apprivoisant vraiment, nous nous efforçons constamment de comprendre les objectifs que nos clients souhaitent atteindre. Nous explorons et mesurons les comportements, nous les comprenons et les traduisons en un résultat concret. Nous donnons un sens aux réalités que les données portent afin d'aider nos clients à prendre des décisions plus efficaces.

Les données, qu'elles soient grandes, petites, humaines, complexes, historiques ou prospectives, n'ont de sens que lorsqu'elles sont utilisées pour développer les expériences, affiner la compréhension du quotidien et prendre les meilleures décisions.

Notre proposition de valeur est fondée sur cinq grands groupes de services, chacun comprenant des offres multiples :

- **Automatisation et intelligence artificielle** : nous fournissons à nos clients les moyens d'améliorer leur productivité et leur précision sur l'ensemble de leurs processus, afin de se concentrer sur le travail à plus forte valeur ajoutée.
- **Expérience numérique centrée sur l'humain** : la relation avec les clients et l'engagement des collaborateurs constituent deux des plus grands contributeurs au succès global des entreprises. Nous aidons les entreprises à imaginer et à créer des expériences numériques multimodales et fluides pour atteindre leurs objectifs.
- **Mise en œuvre des données et des analyses** : les données sont une clé incontestable du succès pour les entreprises. Lorsqu'elles sont utilisées intelligemment, elles ouvrent des opportunités uniques pour faire face aux défis actuels et futurs. Nous permettons aux organisations de déployer tout le potentiel de leurs données : nous mettons la science des données au profit du développement de l'entreprise.
- **Cloud et sécurité** : le Cloud et les plateformes numériques ont le potentiel de révolutionner la façon dont les données sont transformées en valeur, tout en portant l'extensibilité et la flexibilité à un niveau supérieur. Nous sécurisons l'ensemble de vos données et veillons à ce qu'elles soient protégées et confidentielles.
- **Transformation et innovation** : pour prospérer dans l'écosystème actuel, chaque entreprise doit non seulement accélérer sa transformation numérique, mais aussi acquérir des compétences pour stimuler son adaptabilité, sa résilience et sa compétitivité. Nous aidons nos clients à se transformer avec succès pour développer un meilleur futur.

S'appuyant sur l'expérience cumulée de plus de 3 500 collaborateurs et présent dans 27 pays sur 4 continents, Keyrus est l'un des principaux experts internationaux en matière de données, de conseil et de technologie.

Pour en savoir plus : www.keyrus.fr

Jean-Philippe CLAIR
Directeur Marketing, Communication & Expérience client
jean-philippe.clair@keyrus.com