



L'IA en temps réel

verizon
business

L'IA en temps réel au service de décisions instantanées

Dans le monde de l'entreprise, l'IA en temps réel marque un tournant majeur. Fini l'approche purement réactive. Place aux opérations proactives. Concrètement, des systèmes d'IA traitent des flux de données issus de sources variées (capteurs, transactions, caméras...) pour prendre des décisions ou apporter des réponses quasi immédiates, en millisecondes.

Les grands modèles de langage (LLM) tels que ChatGPT sont réactifs par nature, car ils dépendent d'un prompt pour traiter les données par batch. À l'inverse, l'IA en temps réel ingère et exploite des informations en continu, sans aucune intervention humaine¹. Usines, salles d'opération, véhicules autonomes... Dans les environnements où la sécurité physique est en jeu et où le moindre retard peut avoir de graves conséquences, cette anticipation s'avère vitale.

Principaux ressorts de l'IA en temps réel :

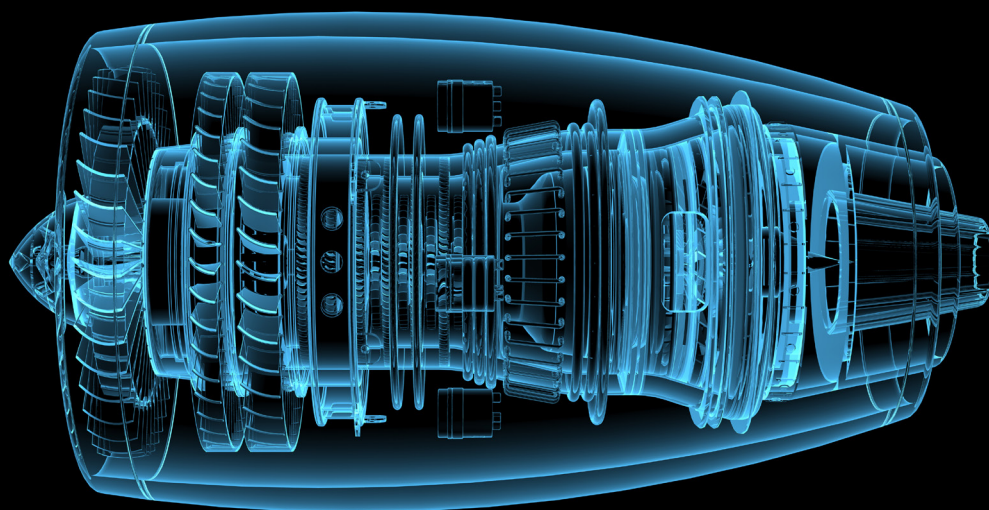
- Faible latence : traitement ultra-rapide, calculé en millisecondes
- Apprentissage continu : adaptation constante aux nouvelles données
- Automatisation : action sans aucune intervention humaine
- Traitement haute-capacité : analyse simultanée d'énormes volumes de données



Selon McKinsey, 60 à 70 % des workloads d'IA devraient passer à l'inférence en temps réel à l'horizon 2030. Ce basculement engendre des besoins urgents et bien supérieurs aux sollicitations actuelles sur trois plans : connectivité à faible latence, capacités de calcul et sécurité en périphérie du réseau².

¹Striim, [The Future of AI is Real-Time Data](#), 2026

²Verizon, [Verizon unveils AI strategy to power next-gen AI demands](#), 2025



Comment des entreprises leaders utilisent-elles l'IA en temps réel ?

L'analytique IA en temps réel livre aux entreprises les insights nécessaires pour orienter leurs décisions à chaque instant. Chaque jour, elle permet de résoudre des problèmes jusqu'alors jugés insurmontables, tout en générant des innovations à fort impact dans de multiples secteurs.

Dans le domaine de la finance, Barclays a révolutionné la détection antifraude en consolidant des systèmes disparates au sein d'Aerospike, une plateforme unifiée d'IA en temps réel³. En plus d'éliminer les silos de données, cette approche centralisée intègre les règles, modèles et profils clients pour harmoniser les décisions.

Grâce à l'accès single-hop aux données et à l'architecture mémoire hybride (HMA) d'Aerospike, Barclays a obtenu des résultats impressionnants. La latence a ainsi chuté de 80 %, avec des temps de réponse systématiquement inférieurs à 100 millisecondes, y compris en pics de charge. La plateforme traite désormais un volume de données quatre fois supérieur et absorbe une croissance multipliée par dix, dépassant les 30 téraoctets. Le tout, sans aucune refonte de l'infrastructure³.

Résultat : une solution à la fois performante, scalable et beaucoup plus précise dans la détection, comme en témoigne la forte réduction des faux positifs et faux négatifs.

Dans l'aéronautique, le programme Intelligent Engine de Rolls-Royce intègre des capteurs IA et IoT au cœur des moteurs d'avions. Dès lors, le moteur physique et sa réplique numérique deviennent indissociables.

Les jumeaux numériques permettent d'anticiper avec précision les besoins de maintenance, plusieurs semaines, voire plusieurs années à l'avance. L'objectif : maximiser le temps en service des moteurs (« time on wing ») et réduire les interruptions⁴. L'intégralité de ce dispositif numérique repose sur Blue Data Thread, une plateforme ouverte qui crée un écosystème connecté réunissant compagnies aériennes, prestataires de maintenance, réparation et révision (MRO) et constructeurs d'origine (OEM)⁵. Ce socle garantit un partage sécurisé des données, essentiel aux acteurs du secteur pour relever leurs défis actuels.

³ Aerospike, [Barclays delivers real-time fraud decisions with predictable sub-100ms performance](#), 2025

⁴ Ellioy, [Aircraft IT: Meet the Intelligent Engine taking predictive engine maintenance to the next level](#), 2021

⁵ Rolls-Royce, [Digital Platforms](#), 2026

Quelles sont les exigences techniques d'une opérationnalisation de l'IA en temps réel ?

La concrétisation de l'IA en temps réel dépend de la vitesse de transfert et de traitement des données. Or les architectures cloud classiques sont trop lentes pour nombre de ces cas d'usage, les allers-retours de données vers des data centers distants étant tout simplement inenvisageables.

Enjeux

L'opérationnalisation de l'IA en temps réel redéfinit en profondeur les exigences d'infrastructure, notamment en matière de latence et de capacité de traitement.

- Le système de recommandation en temps réel de Netflix illustre parfaitement la nécessité d'un traitement des données instantané et d'insights exploitables pour fidéliser les clients. À lui seul, il oriente plus de 80 % du temps de visionnage tout en réduisant significativement les risques d'abandon⁶.
- La latence, cependant, ne représente qu'une partie de l'équation. L'IA en temps réel, en particulier pour la vision par ordinateur et la supervision continue par capteurs, requiert des flux de données massifs et soutenus. Dans ce contexte, 59 % des entreprises citent les contraintes de bande passante comme un frein majeur à l'opérationnalisation de l'IA⁷.

Implications

L'adoption de l'IA en temps réel impose une remise à plat des capacités réseau et des interconnexions entre data centers.

- Face à l'explosion annoncée des volumes de données, les capacités DCI (Data Centre Interconnect) devraient être multipliées par 6 au cours des cinq prochaines années⁸.
- Même les modèles d'IA les plus rapides restent tributaires de la puissance du réseau. Lorsque l'infrastructure ne suit pas, les temps de réponse s'allongent et l'expérience utilisateur se dégrade, au point de menacer la viabilité même de l'IA en temps réel.

Conséquences

Les entreprises doivent accélérer la modernisation et la refonte de leurs infrastructures réseau pour répondre à ces nouvelles exigences.

- Il est impératif de privilégier des architectures réseau robustes et à faible latence qui s'étendent des data centers centraux aux nœuds distribués en périphérie. Le but : garantir des transferts de données fluides et rapides pour répondre aux exigences de l'IA en temps réel.
- Les investissements dans des interconnexions à très haut débit et des infrastructures edge résilientes deviennent indispensables pour éviter les goulets d'étranglement, préserver les performances de l'IA et exploiter ses capacités décisionnelles instantanées.

Remarques

La nature distribuée de l'IA en temps réel pose également de nouveaux défis majeurs en matière de cybersécurité.

- Le traitement des données à la périphérie (véhicules, sites industriels, espaces publics, etc.) engendre de nouvelles failles potentielles, ces nœuds étant physiquement plus exposés aux interférences malveillantes et plus difficiles à sécuriser.
- Les systèmes d'IA en temps réel constituent des cibles de choix. Exemple, l'injection de données erronées dans des systèmes autonomes peut avoir des répercussions catastrophiques. D'où la nécessité de déployer des dispositifs de sécurité avancés, en temps réel, à chaque point du réseau edge⁹.

⁶ Gomez-Urbe & Hunt, [Netflix Tech Blog: The Netflix Recommender System: Algorithms, Business Value, and Innovation](#), 2015

⁷ Wondrasek, [SoftwareSeri: How Bandwidth and Latency Constraints Are Killing AI Projects at Scale](#), 2025

⁸ Deslandes, [Tech Informed: Data centres brace for sixfold connectivity surge by 2030](#), 2025

⁹ Swabey and Glover, [Tech Monitor: The security challenges of edge computing](#), 2022

Vitesse et sécurité, les deux piliers de l'IA en temps réel

Des systèmes autonomes à l'automatisation des processus industriels, l'IA en temps réel présente un potentiel de transformation sans précédent. À condition d'opérer les refontes d'infrastructure nécessaires au niveau du réseau, de l'Edge Computing et de la cybersécurité.

Besoin de l'entreprise	Besoin de l'IA	Impact sur le réseau
Assurer des opérations instantanées et proactives	L'IA en temps réel traite des flux de données en continu, permettant des décisions et des réponses immédiates (en quelques millisecondes), sans intervention humaine.	Une connectivité réseau à très faible latence est indispensable. D'où le recours nécessaire à l'Edge Computing, les architectures cloud traditionnelles étant trop lentes pour prendre en charge des flux synchrones.
Gérer des flux de données massifs et ininterrompus	L'IA en temps réel exige des volumes de données considérables et soutenus pour alimenter ses mécanismes d'apprentissage et de décision en continu.	Cela nécessite une très grande bande passante et des DCI robustes. Les contraintes actuelles de bande passante constituent un frein majeur à l'opérationnalisation de l'IA.
Optimiser les performances grâce à un déploiement stratégique	La nécessité d'une latence ultra-faible et de réponses instantanées impose de rapprocher physiquement certains workloads d'IA en temps réel des sources de données.	Le réseau doit acheminer ces charges vers la périphérie, là où les données sont créées et où les décisions sont prises, ce qui requiert une connectivité solide entre le cœur et la périphérie.
Faciliter une prise de décision continue et adaptative	L'IA en temps réel ingère et exploite en permanence les données, s'adaptant constamment pour maintenir un haut niveau de réactivité.	Cela suppose un accès réseau fiable et rapide à des flux de données continus et à des ressources de traitement distribuées.
Garantir une cybersécurité robuste	Le traitement de données sensibles en périphérie fait des systèmes d'IA en temps réel une cible idéale. L'injection de données erronées peut causer des dégâts catastrophiques.	Les nœuds en périphérie étant plus difficiles à sécuriser, il est impératif d'établir des mécanismes de sécurité avancés, capables d'opérer en temps réel à chaque point du réseau edge.

Libérer la force de l'IA en temps réel à grande échelle



Les workloads d'IA avancés, en particulier ceux qui exigent une réponse en temps réel et des transferts massifs de données, requièrent des capacités réseau et de calcul sans précédent. Pourtant, pour une majorité de dirigeants, l'infrastructure demeure aujourd'hui le principal obstacle à une mise à l'échelle de l'IA.

Chez Verizon Business, nous sommes mieux placés que quiconque pour vous accompagner face à ces enjeux. En interne aussi, nous mobilisons l'IA sur plusieurs fronts : optimisation prédictive des réseaux, insights en temps réel pour le service client et prise de décision basée sur les données. Cette approche nous confère une expertise opérationnelle de l'IA, au-delà des seuls concepts théoriques.

En savoir plus

Aujourd'hui, il est impératif de comprendre comment monétiser l'IA en temps réel et bâtir le réseau sous-jacent. Partout dans le monde, Verizon aide les entreprises à exploiter le plein potentiel de

l'IA pour accélérer l'innovation, optimiser leurs services, générer des insights et renforcer leur business intelligence. Découvrez comment Verizon peut vous accompagner.

*Ce contenu a été créé à l'aide de l'intelligence artificielle et validé par des relecteurs humains.

**Les noms d'entreprise mentionnés le sont exclusivement à titre d'identification et n'impliquent aucun cautionnement.

