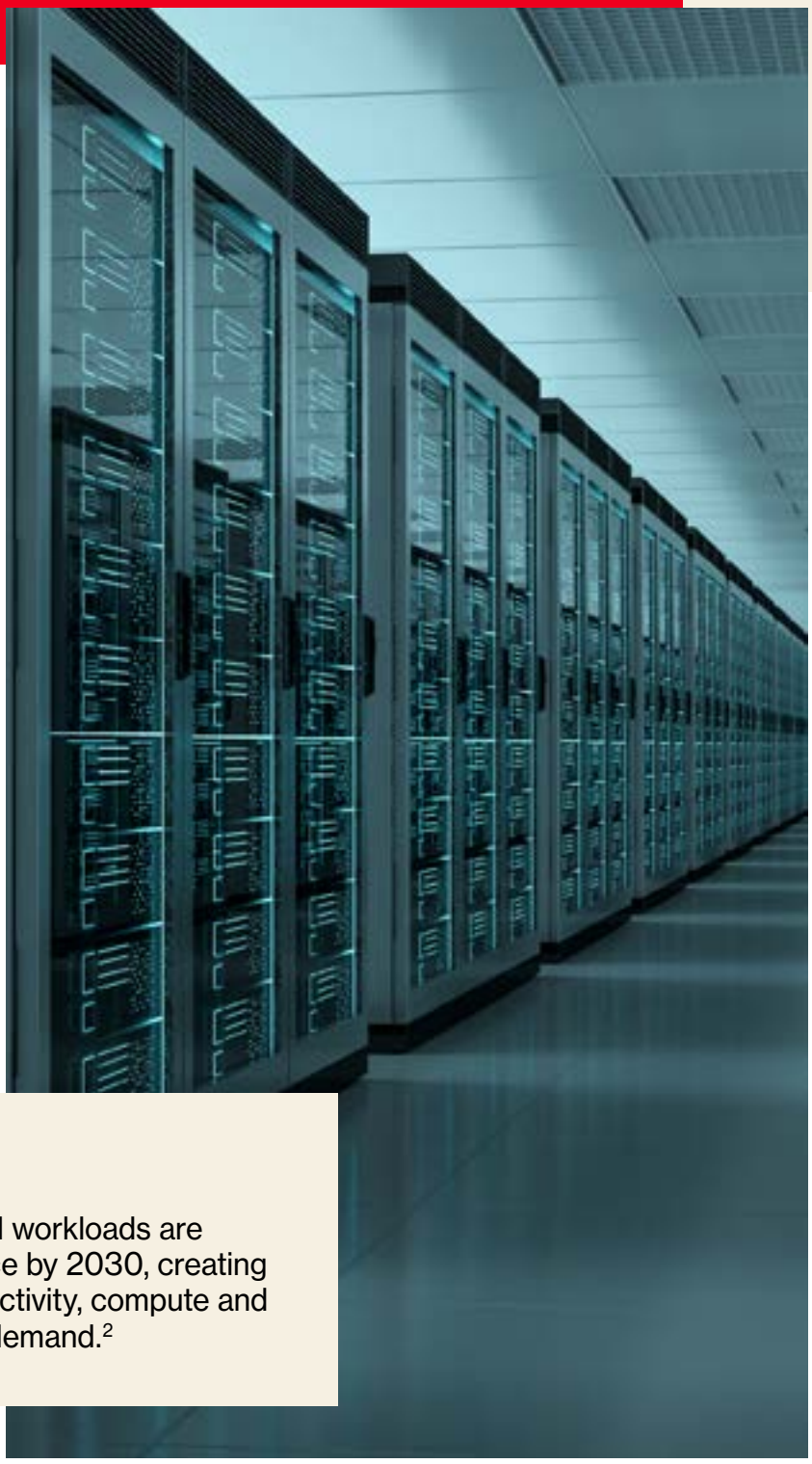
Real-time AI

verizon
business

Real-time AI means instant business decisions



For businesses, real-time AI signifies a fundamental shift from reactive to proactive operations. It refers to AI systems that process live data streams from diverse sources—like sensors, transactions or cameras—and make decisions or generate responses almost instantaneously, typically within milliseconds.

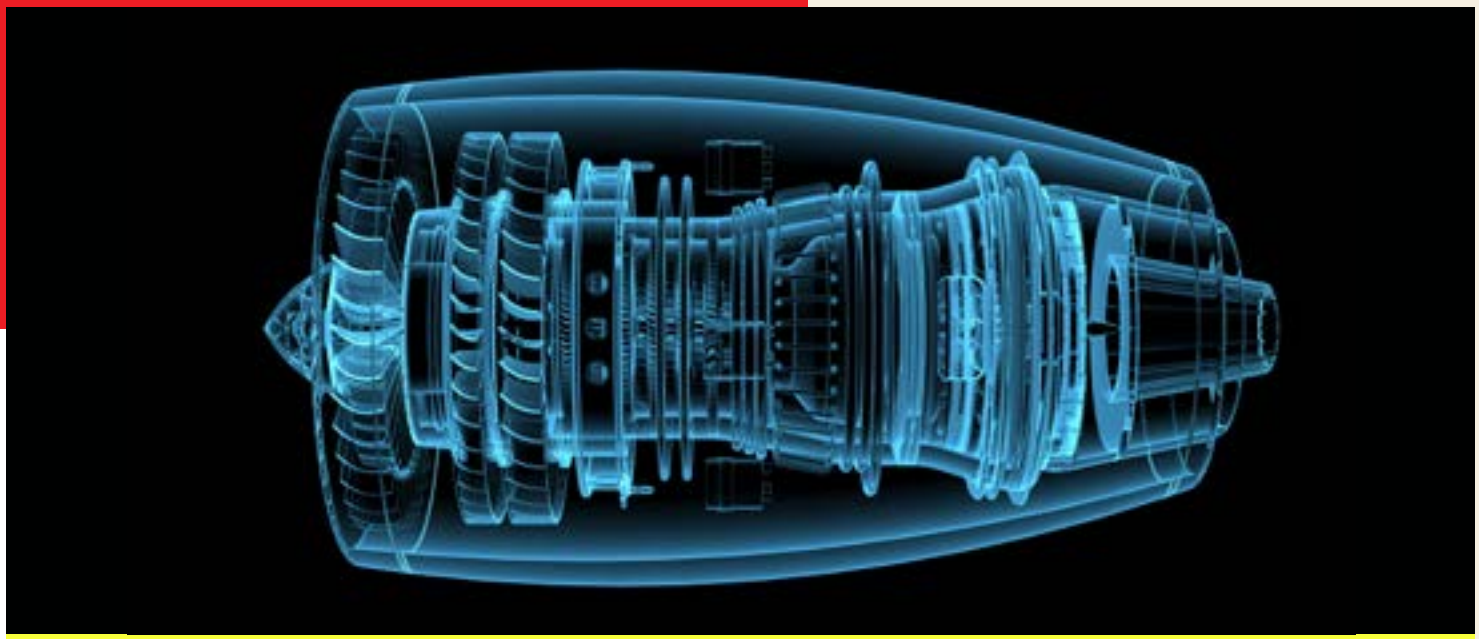
Unlike large language models (LLMs) such as ChatGPT, which are reactive and batch-process data after a human prompt, real-time AI continuously ingests and acts on information without waiting for human input.[1] This proactive approach is vital in safety-critical environments like shared factory floors, operating theatres or self-driving vehicles, where even a slight delay can be dangerous.

Key characteristics of real-time AI:

- Low latency—extremely fast processing, measured in milliseconds
- Continuous learning—adapts to new data on the fly
- Automation—acts without waiting for human input
- High-capacity processing—can analyze massive volumes of data simultaneously

> “
>
> According to McKinsey, 60-70% of AI workloads are expected to shift to real-time inference by 2030, creating an urgent need for low-latency connectivity, compute and security at the edge beyond current demand.[2]

[1] Striim, The Future of AI is Real-Time Data, 2026
[2] Verizon, Verizon unveils AI strategy to power next-gen AI demands, 2025



How are today's leading organizations using real-time AI?

Real-time AI analytics equips organizations with the insights needed for moment-by-moment decision-making. The technology solves issues that previously seemed insurmountable on a daily basis, producing innovations that have an impact across multiple industries.

In finance, Barclays revolutionized fraud detection by consolidating disparate systems onto a unified 'Aerospike' real-time AI platform.[3] This centralized approach eliminated data silos, integrating rules, models and customer profiles for consistent decision-making.

Leveraging Aerospike's predictable single-hop data access and Hybrid Memory Architecture (HMA), Barclays achieved remarkable results. Latency was reduced by 80%, consistently delivering sub-100ms response times at peak loads. The platform now handles 4 times the throughput and accommodates 10 times the data growth to over 30 terabytes without replatforming. [3]

This significantly improved fraud detection accuracy and lowered false positives and false negatives, creating a high-performance, scalable solution.

In aviation, Rolls-Royce's Intelligent Engine programme embeds AI and IoT sensors directly into jet engines, where the physical engine that surrounds the digital engine is indivisible.

Digital twin capabilities accurately predict maintenance demand weeks or even years ahead, maximizing 'time on wing' and reducing disruptions.[4] The entire digital service is underpinned by the 'Blue Data Thread', an open industry solution creating a connected ecosystem for airlines, aviation maintenance, repair and overhaul services (MROs), and original equipment manufacturers (OEMs) of the aircraft.[5] This ensures valuable data is shared safely, helping lead the industry in solving today's challenges.

[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

What are the consequences of making real-time AI operational?

Real-time AI lives or dies by how fast data can travel and be processed. Traditional cloud-based architectures are too slow for many of these applications. Data simply cannot afford the round trip to a distant data center.

Why it matters
Operationalizing real-time AI is fundamentally reshaping infrastructure demands, particularly concerning latency and data throughput.

- Netflix's real-time AI recommender system exemplifies this by demonstrating the critical need for timely data processing and actionable insights to retain users—influencing 80% of hours streamed and preventing abandonment.[6]
- Latency, however, is only half the challenge. Real-time AI, especially for computer vision and continuous sensor monitoring, demands enormous and sustained data flows, with bandwidth limitations now cited by 59% of organizations as a major AI constraint.[7]

What this means
This shift necessitates a profound re-evaluation of network capabilities and data center interconnects.

- A minimum 6 times increase in Data Centre Interconnect (DCI) bandwidth is anticipated over the next five years to support these burgeoning data volumes.[8]
- Even the fastest AI model is constrained by the network carrying it. An insufficient network will make even lightning-fast AI compute feel sluggish and unresponsive to the end user, undermining the very purpose of real-time processing.

And the consequence
Businesses must aggressively upgrade and re-architect their network infrastructure to meet these new, stringent demands.

- Enterprises need to prioritize robust, low-latency network designs that extend from core data centers to distributed edge nodes, ensuring seamless, high-speed data flow for real-time AI workloads.
- Investing in high-bandwidth interconnections and resilient edge infrastructure becomes non-negotiable to prevent AI performance bottlenecks and fully leverage AI's immediate decision-making capabilities.

Need to know
The distributed nature of real-time AI also introduces significant new cybersecurity challenges.

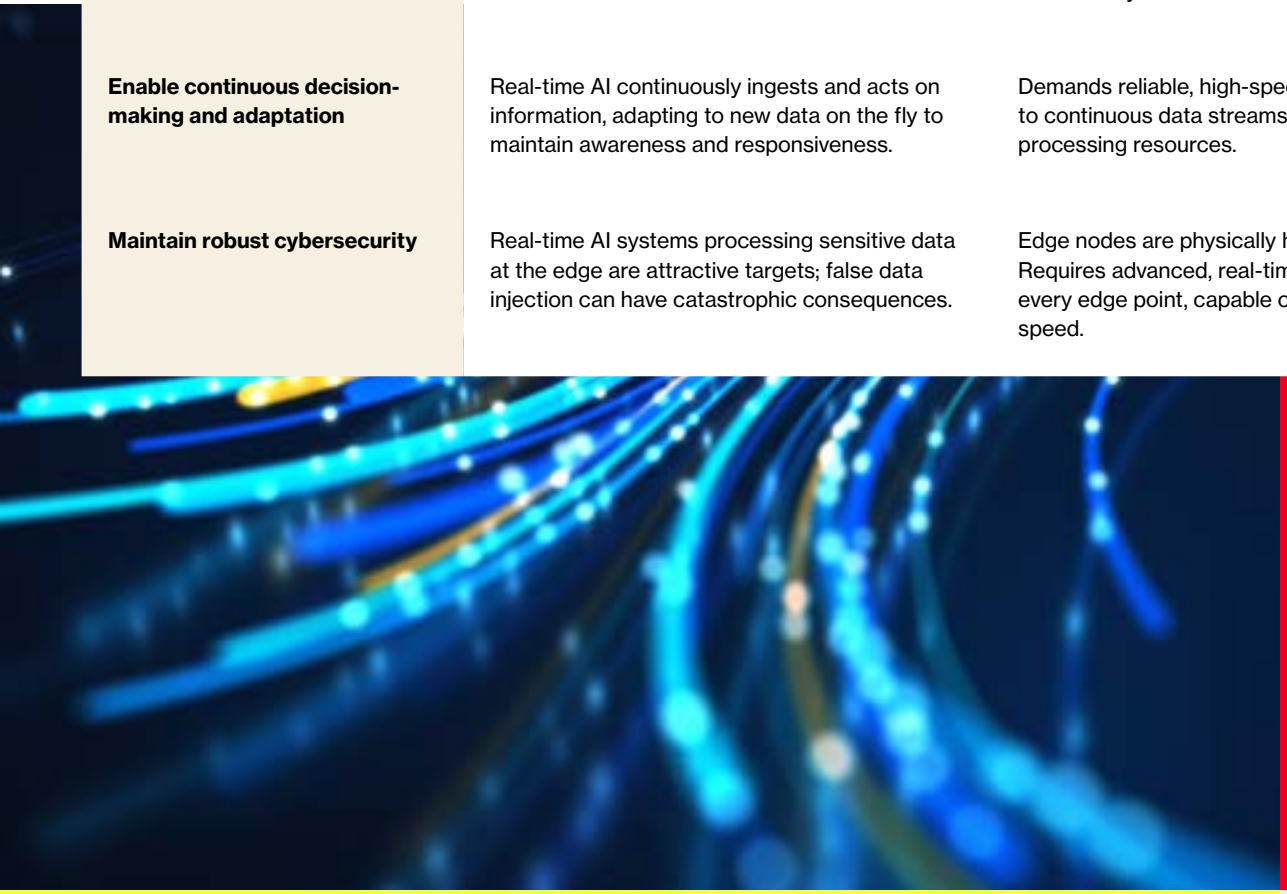
- Processing data at the edge—in vehicles, factories, or public spaces—creates new vulnerabilities. Edge nodes are physically harder to secure and more exposed to adversarial interference.
- Real-time AI systems are attractive targets: feeding false data to an autonomous system at speed can have catastrophic consequences, necessitating advanced, real-time security measures at every edge point.[9]

[6] Gomez-Uribe & Hunt, Netflix Tech Blog: The Netflix Recommender System: Algorithms, Business Value, and Innovation, 2015
[7] Wondrasek, Softwareseni: 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

Real-time AI demands a foundation of speed and security

Real-time AI's revolutionary capabilities, from autonomous systems to industrial automation, offer immense potential, but unlocking it requires tackling significant infrastructure challenges. This involves a proactive overhaul of network capabilities, edge computing and cybersecurity.

Business need	AI need	Network impact
Support instantaneous, proactive operations	Real-time AI processes live data streams, making instantaneous decisions and responses (milliseconds) without human input	Extremely low latency network communication is critical, demanding edge computing, as traditional cloud round trips are too slow to handle synchronous transfers.
Handle massive and continuous data flows	Real-time AI demands enormous and sustained data flows for its continuous learning and decision-making.	Requires massive bandwidth capacity and robust Data Centre Interconnects (DCI). Bandwidth limitations are a major AI constraint.
Optimize through strategic placement	Need for ultra-low latency and instantaneous responses dictates that some real-time AI workloads must run physically closer to data source.	The network must push AI workloads to the edge, physically close to where data is generated and decisions are made, requiring robust core-to-edge connectivity.
Enable continuous decision-making and adaptation	Real-time AI continuously ingests and acts on information, adapting to new data on the fly to maintain awareness and responsiveness.	Demands reliable, high-speed network access to continuous data streams and distributed processing resources.
Maintain robust cybersecurity	Real-time AI systems processing sensitive data at the edge are attractive targets; false data injection can have catastrophic consequences.	Edge nodes are physically harder to secure. Requires advanced, real-time security measures at every edge point, capable of operating at machine speed.



Powering real-time AI at scale

Deploying advanced AI workloads, particularly those demanding real-time responsiveness and massive data throughput, necessitates immense network capacity and computational power. Yet, for the majority of executives, infrastructure remains the single biggest barrier to scaling AI effectively.

Verizon Business is uniquely positioned to be the partner businesses need. We also use AI in our own business for predictive network optimization, real-time customer service intelligence and data-driven decision-making. This gives us a practical understanding of AI that goes far beyond theory.

Find out more

Understanding how to generate value from real-time AI and build the network to support its unique demands is essential. Verizon works with businesses globally, helping them leverage AI to accelerate innovation, enhance services, harvest insights, and drive business intelligence. Learn more about how we can help you.

*This content was created with the assistance of artificial intelligence and validated by humans.

**All company names are used for identification only and do not imply endorsement.