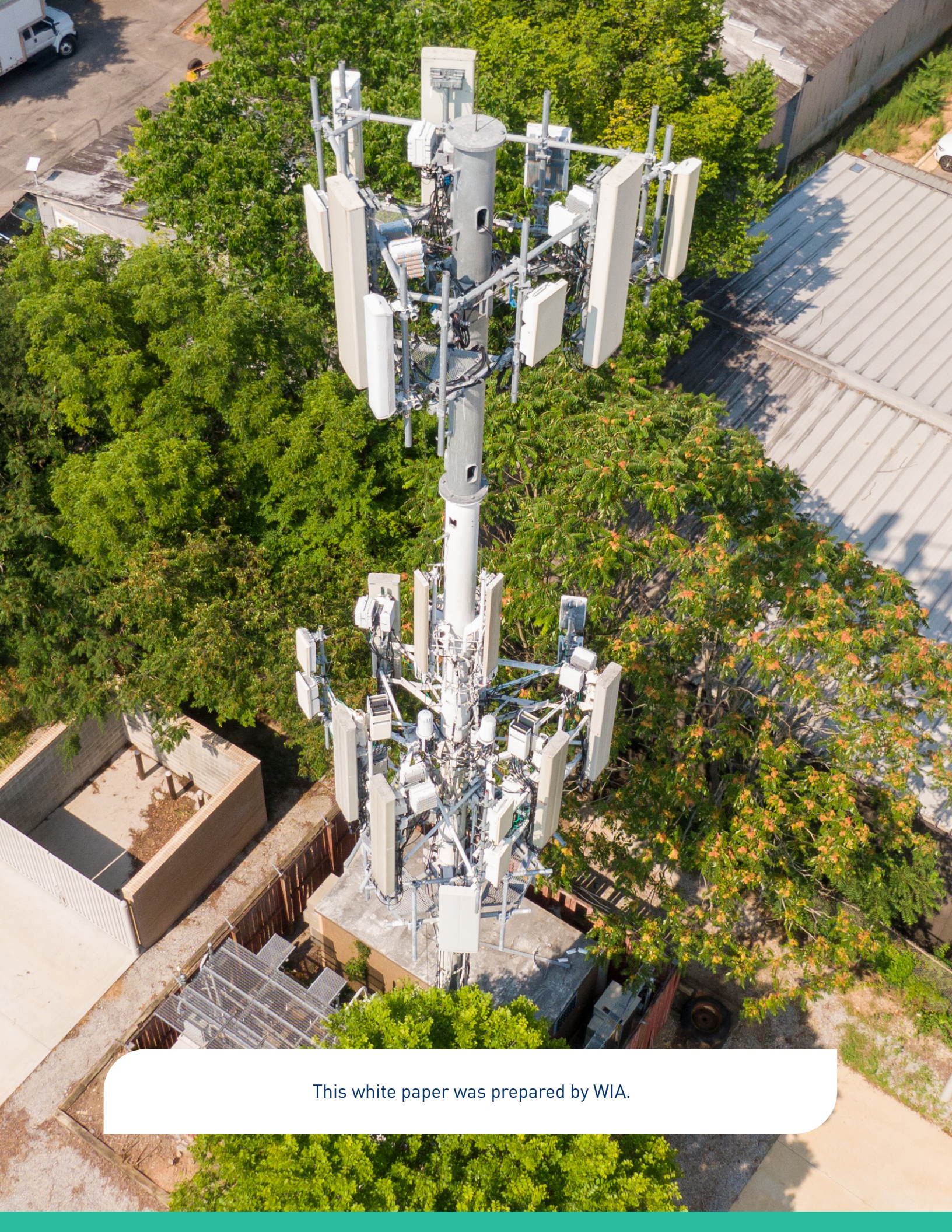




Edge Infrastructure

Defining the Opportunity and Path Ahead

Published Q4 2026



This white paper was prepared by WIA.

Content

Executive Summary	1
What Is Edge Infrastructure?	3
Where Edge Infrastructure is Deployed	3
Why Edge Infrastructure? Why Now?	4
Three Conditions Now Exist At Once	
The Role Of The Edge Infrastructure Initiative	6
General Findings: Sizing The Opportunity	7
What Is Actually Driving Demand Today	7
An Industry Driven Definition	8
Working Definitions	9
25 Word Definition	9
100 Word Definition	9
Power: The Central Constraint	10
Key Findings	11
Open Questions The Initiative Is Working Through	12
Conclusions	13
Key Takeaway	13
Get Involved	13





Executive Summary

“ Edge infrastructure places compute, storage, and connectivity equipment at active communications sites, using existing infrastructure to deliver resilient, low-latency processing, telecom, and digital services closer to users and devices.

The concept of edge infrastructure is not new, but novel use cases have changed the game. Edge infrastructure is a near-term opportunity created by the convergence of inference demand, latency-sensitive applications, and the availability of distributed communications sites.

Rather than working toward one single edge architecture, the opportunity today is to make computing resources and connectivity available in the places they are most needed. The market for edge processing is today driven by demand for: distributed compute; network points of presence and fiber hubs; AI and

compute applications; and emerging applications. A consistent thread runs through all four categories: infrastructure built today to serve current workloads becomes the foundation the next generation of applications depends on. Build decisions made now shape what is possible later.

Edge infrastructure is a genuine, near-term opportunity but capturing it will require the ecosystem to identify and address shared questions and challenges. That collective work is the purpose of the WIA's Edge Infrastructure Initiative (EI2). The Initiative's work to date points to three main conclusions:

- The demand behind edge infrastructure today is driven by the latency requirements imposed by AI inference, requirements that centralized hyperscaler facilities cannot meet for physical reasons, not simply cost reasons.
- Because infrastructure build cycles are long, the window for organizations to secure sites, power, and fiber positioning is open now.
- Power availability is the near-term gating factor.

EI2 is a working group convened by WIA that brings together a representative cross-section of the wireless and communications infrastructure ecosystem to work through the practical, technical and policy questions that need to be answered to allow a frictionless market to develop. This paper describes the Initiative's collective work and findings to date.

What Is Edge Infrastructure?

Edge infrastructure moves computing power out of a small number of very large, centralized facilities and distributes it closer to the end user devices — at cell towers, telecom buildings, commercial buildings, ILA (In Line Amplifier) sites, transportation hubs, utility facilities, municipal sites and other locations that already exist. Instead of sending every request on a long round trip to a distant, campus-scale computing facility, edge infrastructure processes information close to where it is created: on the device, at a cell tower, or a few miles away.

This proximity is what makes it possible to return results in single-digit milliseconds rather than tens or hundreds of milliseconds, while drawing a small fraction of the power, water and land that a large, centralized facility requires. Edge nodes are typically closet-to-room sized and fit within existing buildings and telecom and utility facilities.

Where Edge Infrastructure Is Deployed

Edge nodes are compact — often no larger than a server rack or a small room — which makes them deployable in a wide variety of locations that already exist: telecommunications sites such as cell towers, In-line Amplifier (ILA) sites and 5G base stations that already have power, connectivity, and land that has already been permitted; transportation hubs and corridors, including airports, transit stations and rail yards; municipal and utility buildings such as city halls, public safety facilities and substations; and commercial and industrial facilities, including office buildings, warehouses, hospitals and manufacturing plants. In nearly every case, adding edge computing equipment is a modest addition to an existing site, rather than a new campus built from the ground up.

Edge Infrastructure	
Physical Size	Closet to room-sized
Power Draw	Approximately 200 kW-3,000 kW
Water Use	Minimal to none
Typical Zoning	Telecommunications, light industrial or commercial

Why Edge Infrastructure? Why Now?

Edge infrastructure is not new. The concept has existed for more than a decade, but earlier efforts — most notably multi-access edge computing (MEC) — arrived ahead of real demand. The applications that were to justify distributed compute at scale did not materialize on the timelines anyone projected. In short, MEC was a solution looking for an application.

That has changed. AI inference — running a trained model against new, real-world data — is the first workload that centralized architecture genuinely struggles to serve for a growing set of use cases. Real-time conversation, autonomous systems, multi-step AI agent workflows and physical AI interactions (with drones, robotics and automatic guide vehicles, for example) require response times measured in single-digit milliseconds (ms). At this threshold, the distance data must travel becomes the limiting factor: the speed of light imposes a ceiling that no amount of additional capacity in a distant facility can remove.

Three Conditions Now Exist At Once

Condition	What It Means
Enterprise AI spending is real	Production deployments with defined latency budgets are underway today - not pilots.
Applications are specific and identifiable	AI referencing, real-time video analytics and industrial private networks are current examples; autonomous vehicles and robotics, sensing-and-communication and 6G infrastructure are near-term additions.
The network foundation is in place	Current telecom architectures make edge compute technically viable at telecom sites rather than theoretical.

Because infrastructure build cycles — site selection, power, fiber, permitting and equipment — typically run two to three years, siting and capacity decisions today will lay the foundation for tomorrow’s edge infrastructure capabilities.



The Role Of The Edge Infrastructure Initiative

The Edge Infrastructure Initiative is a working group convened by WIA that brings together a representative cross-section of the wireless and communications infrastructure ecosystem — site and tower owners, network operators and technology and equipment providers among them — to work through the practical, technical and policy questions standing between today’s edge infrastructure opportunity and a mature, investable market. This paper describes the Initiative’s collective work and findings.

Work Plan Pillar	Objective
Determine physical parameters	Define the latency, bandwidth, power, cooling, security, and space parameters that describe edge infrastructure today and over the next five to ten years.
Enable ecosystem development	Reduce friction for site owners and prospective tenants through thought leadership and education.
Track legislative and regulatory action	Monitor federal, state, and local legislation, regulations and executive orders to ensure frictionless deployment of edge infrastructure.

The group has met on a regular cadence since April 2026 to address these issues — building a consistent, shared body of analysis month-over-month rather than a one-time survey. Looking ahead, the final months of 2026 will be spent turning its analysis into practical resources for the ecosystem.

General Findings: Sizing The Opportunity

What Is Actually Driving Demand Today

Current and near-term edge infrastructure demand can be grouped into a small number of recurring categories, summarized at a high level below. The Initiative's underlying materials track each category in more depth, including projected timelines and technical drivers.

Category	What's Driving It
Distributed compute demand	Power and permitting constraints on large, centralized facilities are pushing shared compute storage, sovereign and private cloud deployments and decentralized architectures toward smaller, distributed installations.
Network points of presence and fiber hubs	As fiber density increases, more communications sites, from towers to ILA sites, become viable aggregation points; these deployments generally have lower power needs than compute-heavy installations.
Current AI and compute applications	Industrial and manufacturing sensing, healthcare imaging and monitoring, private large-language-model deployments with data-locality requirements, autonomous systems, real-time video analytics and dedicated AI inference hosting.
Emerging Applications	Ambient AI and always-on wearables, immersive and holographic communications, sensing-and-communication integrated with 6G, and multi-step agentic AI workflows that chain many inference calls together.

A consistent thread runs through all four categories: infrastructure built today to serve current workloads becomes the foundation the next generation of applications depends on. Build decisions made now shape what is possible later.

An Industry-Driven Definition

Edge infrastructure does not have a single universal definition. A three-tier framework of deployments is emerging to capture the range of deployments the market is actually building toward. There appears to be an unusually high degree of consistency across the ecosystem, with no significant outliers — an early sign that the market, while still young, is converging on a shared mental model of what “edge” means in practice.

Tier	Typical Footprint	Illustrative Applications	Latency Character
Small (Local edge)	Closet to room-sized; fits within a cabinet, shelter or existing building space	Real-time conversation, AR/VR, autonomous control loops, healthcare monitoring	Sub-10 ms; single hop to the user
Medium (Regional edge)	Single hall or modular building; a handful of deployments per metro area	Agentic AI workflows, real-time analytics, metro aggregation	Roughly 10-30 ms
Large (Edge hub)	Building or campus-scale, but still far smaller than a hyperscale campus; a small number per state	Heavier reasoning workloads, batch analytics, multi-site overflow	Tolerant of 30-100+ ms

The framework is deliberately descriptive rather than exhaustive: it is meant to orient a newcomer to the range of deployments under discussion, not to serve as a design specification.

Working Definitions

25 Word Description

“Edge infrastructure places compute, storage, and connectivity equipment at active communications sites, using existing infrastructure to deliver resilient, low-latency processing, telecom and digital services closer to users and devices.”

100 Word Description

“Edge infrastructure is distributed, secure computing equipment deployed close to users, devices and the places where data is generated, at existing communications sites such as cell towers, buildings and campuses. Compact and typically neutral-host, it houses compute, storage, networking, power, cooling, backup power, security and fiber connectivity in small-footprint installations. This includes equipment racks, modular enclosures, or single-story structures, minimizing noise and physical impact while right-sizing electrical draw to each location’s needs. Facilities are usually unstaffed and remotely monitored, reusing site infrastructure and existing electrical service rather than new greenfield expansion. By processing data locally, edge infrastructure delivers low-latency, resilient performance for AI, public safety, enterprise and smart-building applications.”

— WIA Edge Infrastructure Initiative, August 2026



Power: The Central Constraint

Power has emerged as a dominant theme. Existing communications sites have real, but limited, electrical headroom and that headroom varies enormously by site type and location — a rural low-band tower looks very different from a dense urban site carrying multiple carriers; a hospital or airport has more electrical capacity than a commercial building. Because a third-party edge deployment typically requires its own utility connection rather than sharing an existing circuit, adding compute can trigger new permitting and interconnection steps even at sites that otherwise look ready.

In some markets, the ability to secure a new utility connection at all — separate from the underlying capacity available — is becoming a binding constraint in its own right, as utilities work through a growing queue of requests. That queue is compounded by the volume of requests hitting utilities from every direction at once, not just from edge deployments, which several companies have observed firsthand at industry power-planning forums.

Beyond raw capacity, several site-level engineering questions came up repeatedly. Backup power and battery sizing for edge compute follow different

rules than long-standing telecom equipment practice: mission-critical AI workloads generally need generator support, and runtime and power-quality requirements differ from the backup profile telecom equipment that is typically designed around. Many communications sites also run on DC power for existing radio equipment while edge compute hardware typically expects AC, which adds a conversion step and cooling choices — from simple air cooling to direct-to-chip liquid cooling — depend heavily on rack density and the local climate. None of these are unsolvable, but they mean an edge deployment is rarely a simple bolt-on to an existing site and each one benefits from planning informed by what other members have already learned.

Key Findings

Edge infrastructure investment should be planned alongside — not independently of — a site's existing and anticipated power profile.



Open Questions The Initiative Is Working Through

Alongside technical and policy findings, there are numerous practical frictions currently slowing edge infrastructure deployment. None of these are unusual for an infrastructure category at this stage of maturity, and we expect more questions will arise as the market develops.

- Limited technical standardization across a distributed footprint made up of many vendors and site types.
- No settled business model yet for monetizing edge compute capacity.
- Open governance and data-handling questions, particularly around data sovereignty, privacy, and security in neutral-host models where multiple tenants may share the same infrastructure.
- A workforce and skills gap for deploying and operating distributed compute at telecom sites.

Key Takeaway

Edge infrastructure is a genuine, near-term opportunity — but capturing it will require the ecosystem to work through questions of siting, power, and governance together.

Conclusions

The Initiative's work to date points to three main conclusions:

- The demand behind edge infrastructure today is different in kind from the demand that drove — and ultimately did not sustain — the last wave of edge investment. AI inference imposes latency requirements that centralized hyperscaler facilities cannot meet for physical reasons, not simply cost reasons.
- Because infrastructure build cycles are long, the window for organizations to secure sites, power, and fiber positioning is open now.
- Power availability — more than the underlying technology — is the near-term gating factor.

Get Involved

The Initiative meets monthly and welcomes additional participants who bring a distinct perspective to the ecosystem — including from adjacent industries not yet represented. Members receive the full underlying data behind this paper, participate in shaping upcoming homework and work streams. For more information about joining, please contact Iain.Gillott@WIA.org

This paper will be updated as the Initiative's work progresses. Prospective members are welcome to request the most current version, along with the full parameter data, regulatory tracking detail and discussion notes referenced throughout.



WIA.ORG