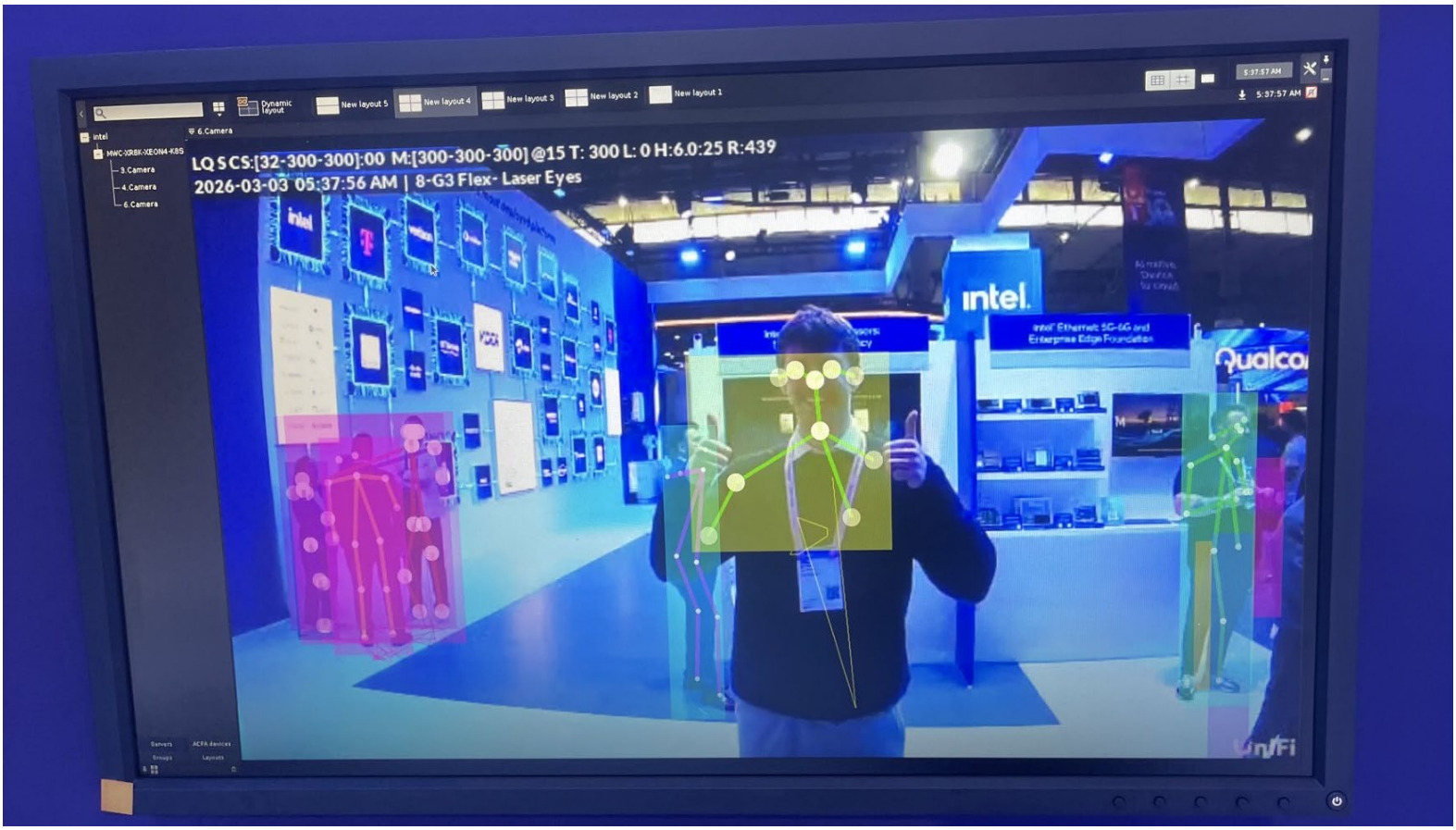


Edge computing at MWC 2026

In contrast to previous years, when edge computing had drifted somewhat to the periphery of the agenda, it was more prominent at MWC 2026, with discussions and demonstrations across the Fira reinforcing what we see as the key dynamic reenergising the market: the inferencing era of AI will prove a critical tipping point for the adoption of edge computing.

George Glanville, Senior Analyst and Edge Practice Lead

March 2026

While devices and hyperscale clouds will continue to support a large share of inferencing workloads, there are situations where it is neither practical nor feasible to run inference in these environments. These use cases sit in the 'edge sweet spot': they require more compute capacity and flexibility than can be delivered by devices or embedded systems, but constraints around latency, data backhaul and data sensitivity preclude deployment in the hyperscale cloud.

Many of these situations were visible around the conference, but not necessarily under the term of edge computing. Instead, the marketing machine has reframed these situations under a range of different guises – physical AI, distributed data centres, cloudified operational technology (OT), private AI and more. Given edge computing's synonymity with latency, this evolution may better convey that propositions are now more focused on data control, a finding illustrated by our edge industry survey.

Demonstrating the (computer) vision

Computer vision was at the forefront of edge computing demonstrations at MWC 2026. This supports our view that computer vision represents the most mature and commercially developed edge AI opportunity today. For example, our edge AI market forecast projects that computer vision use cases accounted for nearly 50% of the global edge AI addressable market in 2025.

Vendors showcased computer vision across a wide range of contexts

Source: STL Partners

Across the exhibition halls, vendors showcased computer vision deployments across a wide range of contexts. For example, Google Cloud and Intel showcased solutions to measure dwell time in retail environments, Intel targeted intrusion detection, Qualcomm highlighted PPE compliance and defect monitoring in industrial settings, Telefónica showed how computer vision could alert authorities to emergencies on trains, while Wind River illustrated how computer vision can support robotic operations. The applicability of computer vision across such a diverse array of industry verticals illustrates why many ecosystem players view it as the most immediate value driver for edge AI.

A growing dynamic observed in deployments of computer vision is the integration of small language models (SLM) as an interface enabling users to query computer vision outputs conversationally. For example, Intel and

Cisco showcased how a nine-billion-parameter SLM – represented by a holographic version of Howie Mandel – can enable users to query queue lengths at bars and toilets at live events.

Let's get physical!

While computer vision represents the most immediate edge AI opportunity, many vendors were also keen to showcase the longer-term potential of physical AI. The premise is that as connected robots, vehicles and drones proliferate, there will be a growing need for low-latency, off-device processing to coordinate and orchestrate these systems.

One example came from Ericsson, working with the Automotive Edge Computing Consortium, which demonstrated how off-vehicle edge processing could enable improved inter-vehicle awareness for safety applications. A key attraction of edge computing for physical AI is that it can support the deployment of AI across brownfield infrastructure. For example, Telefónica's partnership with rail operator CAF enables intelligent decision-making for rail vehicles that have a lifecycle of up to 30 years.

In this context of edge computing for physical AI transport use cases, several vendors view the potential of edge compute deployed at the RAN as particularly compelling due to its ability to combine low latency with wide-area coverage. As John Saw, CTO at T-Mobile US, stated, the network could ultimately form the "nervous system of humanoid AI". Capgemini illustrated this concept showing how two AI-enabled RAN sites could support the safer operation of autonomous buses in Barcelona.

Physical AI will not only materialise in wide-area transport environments; more localised deployments are also emerging. For example, Google Cloud and KDDI demonstrated an agentic AI deployment for the Japanese retailer Lawson. Google Distributed Cloud runs in-store, enabling Gemini-powered AI agents that monitor shelves, manage inventory and instruct a robotic arm to retrieve products.

While these demonstrations illustrate the long-term potential of edge computing in orchestrating physical AI systems, significant hurdles remain. These include the need to establish deeper partnerships with transport and robotics original equipment manufacturers (OEMs), secure regulatory approvals and safety certifications, deploy edge infrastructure at greater scale, and develop clearer pathways to monetisation.

Chipping away at the edge

As enterprises stack different use cases across computer vision and physical AI at the edge, a growing battleground is emerging in respect to the silicon required to power these use cases.

Intel, unsurprisingly, emphasised the ability of its CPUs to support many edge AI workloads without requiring discrete AI accelerators. A useful anecdotal benchmark in this discussion is that one CPU core supports one camera and approximately ten decisions. Discussion at the Fira also suggested that with CPU-only deployments, there is often a threshold of five video streams that can be analysed on a standard CPU, at which point a GPU becomes necessary.

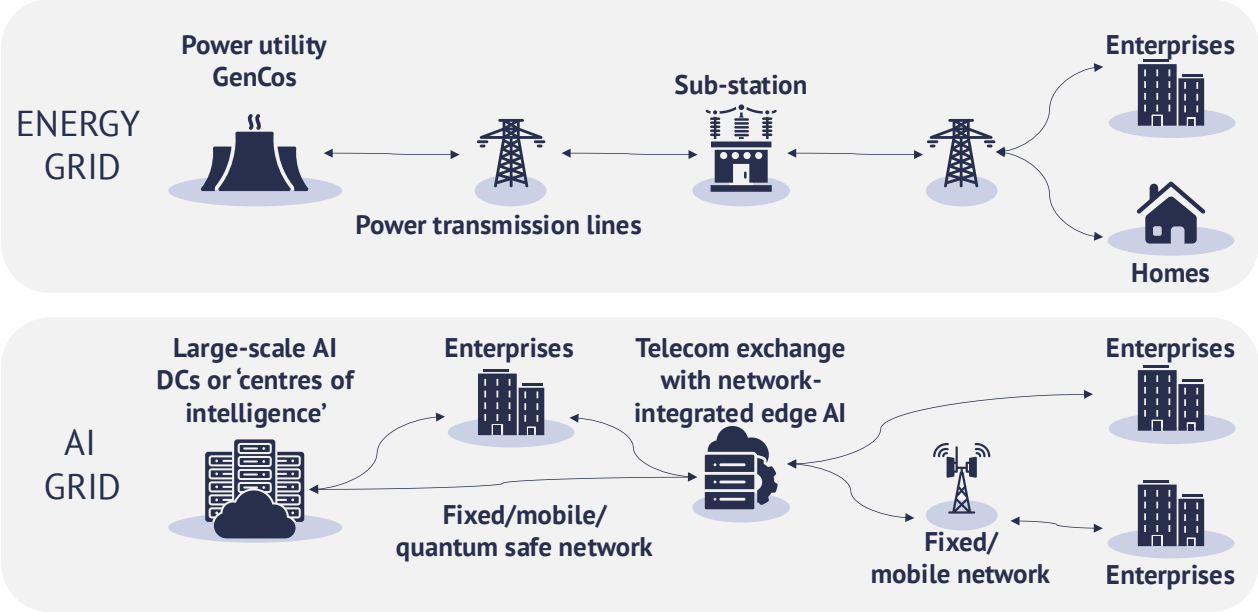
However, Nvidia GPUs are not the only option emerging in this space. For example, Alibaba showcased its on-premises AI Stack, equipped with its proprietary AI accelerators. According to the company, two of these accelerators can support models with up to 32 billion parameters, while larger configurations can scale to support models approaching one trillion parameters.

Network edge and AI factories: There is electricity in the air

Illustrating its approach to developing compute infrastructure capable of hosting AI workloads, Singtel Digital Infrastructure presented its strategy as analogous to an electricity grid. In this architecture, it is developing central sites for the training of AI models (the power plants), connected via its fixed network (the transmission grid) to converted central offices that enable the inferencing of AI models (the electricity substations) that then ultimately deliver outputs to enterprises.

Indeed, as outlined in our Telco AI data centre forecast, establishing AI factories and network edge sites as complementary layers is an approach we have seen adopted by several operators in South-East Asia, such as China Telecom. While this strategy serves to position the telco as a more comprehensive leader of sovereign AI compute infrastructure, it is most viable for those with an appetite for higher capital intensity, operating across large or geographically dispersed markets where the latency advantage of deploying workloads at the network edge is more pronounced, and where existing cloud provider presence is more limited. For these reasons it may not fully translate to Europe.

Singtel Digital InfraCo's AI grid



Source: STL Partners recreation of graphic from Singtel Digital InfraCo event on 1 March 2026

Edge federation is only useful if there is an edge to federate

Building on last year's edge federation initiative between Telefónica, Deutsche Telekom and TIM, two additional operators – Vodafone and Orange – joined the demonstration at MWC 2026. The initiative aims to allow developers to access multiple operators' edge nodes through a single, consistent portal.

Despite this progress, the initiative remains largely at the pilot stage, constrained by a fundamental issue: successful edge federation requires, first, an edge to federate. With the exception of Telefónica which has five commercially active nodes and plans to reach 17 by year-end, the other operators in this consortium only have a couple of pilot sites each – undermining the utility of this proposition. With the introduction of a new European Commission IPCEI scheme focused on developing edge infrastructure, network edge build may galvanise, but these programmes are likely to move slowly.

In true EU fashion, what is better than one federation project? Two federation projects. The conference also saw the announcement of EURO-3C, a new initiative aimed at developing a federated edge cloud in Europe, with a stronger focus on commercialisation than the IPCEI programmes. Notably, the project includes a broader set of participants, including operators such as BT and KPN alongside European cloud providers like OVHcloud. Together, these initiatives reflect the European Commission's belief that aggregating capacity across multiple providers is the most viable path to competing with the entrenched hyperscale cloud platforms.

What next?

STL Partners will continue to closely monitor the evolution of the edge computing market, and how telecoms operators fit within it. To view STL Partners' broader perspective on MWC 2026 – covering topics such as network innovation, enterprise platforms and private networks – click here.

George Glanville is a Senior Analyst at STL Partners.

Get in touch with the author to learn more

george.glanville@stlpartners.com

Or visit STL Partners' Edge Computing Hub

https://stlpartners.com/edge-computing/