

The Intelligent Edge Changes Who Decides

Whitepaper

Bart de Graaff · BdG Advisory · October 2026

The intelligent edge is usually sold as a hardware and latency story. It is a decision-rights story. Cameras, controllers and machines now make operational calls at the point of work, under authority envelopes nobody wrote, through update channels the enterprise often does not hold, and regulators are already asking who can override them.

The decision moved. The authority did not.

Your cameras are already making decisions. Nobody in your organisation signed them off.

The industry describes the intelligent edge as a hardware story: compute moving closer to devices, cameras, controllers and sites, so that latency drops and bandwidth bills shrink. That description is accurate and beside the point. When inference runs on site, the decision runs on site. A choice that used to travel up a chain of systems and people, and come back down as an instruction, is now made at the point of work, in milliseconds, by a model nobody at headquarters has looked at. The enterprise has moved its decisions without moving its decision rights. Ownership, escalation, override and audit still sit where the old chain put them. The decisions no longer pass through there.

This paper argues that the intelligent edge changes the decision chain, not the network diagram. It explains why inference is leaving the data centre, what happens to authority when it does, where control really ends up, and which structural moves let leaders decide where intelligence lives before their suppliers decide for them.

Inference is leaving the data centre for rational reasons

The move to the edge is not a fashion. Each party in the chain has a sound reason to push inference outwards, and those reasons reinforce one another.

Physics comes first. A machine moving through a field, a robot on a line or a camera watching a loading bay produces data faster than it can be shipped to a central region, judged and answered. When the decision has to land before the object has moved, the round trip is the problem.

Economics comes second. Shipping raw video and sensor streams to the cloud costs bandwidth, storage and compute, and most of that data is worthless the moment the decision has been made. Judging locally and sending upwards only the result is cheaper at every layer.

Sovereignty and privacy come third. Data that never leaves the site cannot be intercepted in transit, stored in the wrong jurisdiction or reused without consent. For regulated industries that is a design requirement, not a preference.

The money is following. IDC estimated global edge computing spending at nearly \$261 billion in 2025, growing to \$380 billion by 2028, with artificial intelligence among the fastest-growing segments.

Hardware is being repriced around it. In 2024 Microsoft defined a new class of Windows machine, the Copilot+ PC, around a neural processing unit of 40 or more trillion operations per second, so that small language models run on the device rather than in a data centre. Apple splits Apple Intelligence between a roughly 3-billion-parameter model on the device and a larger model on its Private Cloud Compute servers, and opened the on-device model to developers in 2025. A laptop or a phone is now specified, marketed and priced on how much inference it can do locally.

Industrial equipment got there earlier and with less noise. Siemens announced a module that adds a neural processing chip to its S7-1500 controllers in 2018, so that a programmable logic controller (PLC) could run trained models for visual quality checks and robot guidance next to its control logic. The edge did not arrive with generative AI. It has been accumulating quietly inside operational technology for years.

The models are changing shape at the same time. Gartner predicts that by 2027 organisations will use small, task-specific AI models at least three times more than general-purpose large language models, because general models lose accuracy where tasks need specific business context, and small models are faster and cheaper to run. A domain model that knows one kind of weld, one kind of shelf or one kind of crop does not need a data centre. It fits on the device that sees the weld.

Put those together and the direction is settled. Device makers sell premium silicon. Enterprises cut transport and cloud costs. Regulators prefer data that stays local. Model builders find that small and specific beats large and general for most operational work. Nobody has to push. Everyone is pulling in the same direction.

What none of those parties is responsible for is the decision architecture. That falls to the enterprise, and most enterprises have not noticed it is theirs.

The decision chain inverts

The process-era decision chain ran in one direction. Sense, transmit, decide centrally, dispatch, act. Data flowed up. Instructions flowed down. Authority sat in the middle, with a

system of record and a named person who could be asked why.

The intelligent edge runs the chain the other way. Sense, decide, act, then report. The decision is made where the data is created. What travels upwards is the result, the exception and the evidence, if anyone designed it to travel at all.

John Deere's See & Spray shows the pattern at scale. Boom-mounted cameras and onboard processors identify weeds as the sprayer moves through the field, at up to 15 miles per hour, and trigger individual nozzles only where a weed is detected. In 2025 the technology was used across more than five million acres, and customers cut non-residual herbicide use by an average of nearly 50%, saving about 31 million gallons of herbicide mix, as trade press reporting of the same results confirms.

Consider what that means for decision rights. Every plant in those five million acres was the subject of a separate decision: spray or do not spray. No agronomist approved any of them. No farmer saw them one by one. The farmer set the parameters, chose the mode and owns the outcome. The machine made the decisions. The value came precisely from removing the human from each individual call.

That is the inversion in one example. The centre stops deciding and starts authoring the boundaries inside which the edge decides. Call that boundary the authority envelope: what the device may decide on its own, within which parameters, when it must escalate, who can override it, and what it must report.

The envelope is a governance artefact. In most enterprises it does not exist as one. It exists as a vendor's default configuration, a set of parameters someone in operations chose during commissioning, and a hope that the model behaves.

The data flows invert with the decisions. In the process-era chain, central analytics teams received everything, because everything had to come to the centre to be decided. In the edge chain, the raw data is consumed where it is created and mostly discarded. The centre receives summaries. A central data team that built its value on owning the full data set discovers it now sees the decisions, not the evidence behind them, unless the edge was built to send it.

The economics move with them. The value of a See & Spray decision is herbicide not bought, and it accrues field by field, at the edge. In a factory, the value of a vision model is scrap avoided on one line. In a store, it is shrinkage caught at one door. When value is created and measured at the point of work, the point of work starts to behave like a profit centre. The edge is not only getting faster. It is getting its own P&L.

That matters for authority, because decision rights follow value. A site whose models save measurable money will want a say in how those models are tuned, replaced and funded. A central function that pays for the platform will want a say in how it is used. Both claims are reasonable. Without an explicit allocation, they are settled by whoever has more organisational weight that quarter, which is not a design.

Security follows the same logic. The perimeter model assumed that what mattered sat inside the data centre and the network around it. When decisions are made on a camera at a gate or a controller on a line, there is no useful perimeter. Each device has to prove what it is, what it runs and who changed it last. Identity moves from the location to the device and the transaction.

That shift is structural, not technical. It moves informational power, economic weight and security exposure from the centre to the site. It is the reason the intelligent edge belongs on the leadership agenda, rather than in the infrastructure roadmap.

The site becomes a decision-maker

Domain-specific models do more than move inference. They make competence local.

A generic model knows a little about everything and the same thing everywhere. A domain model trained on one site's data knows that site: its lighting, its product mix, its failure modes, its weather. It gets better at the place by becoming worse at being general. That is the point, and it is the governance problem.

Take a pattern common in distribution. A national logistics operator runs a hundred depots. It deploys a vision model to classify damaged parcels at intake, and the model is fine-tuned per depot because packaging, lighting and conveyor speed differ. Within a year there are a hundred variants. Each depot manager has a model that reflects their site better than any central one could.

Now ask the governance questions. Who approved variant 47? Who knows it has started rejecting a packaging type that depot 12 accepts? When a large customer complains that its goods are classified as damaged at one depot and not at another, which policy applies: head office's, or the model's? Who has the authority to change it, and who will notice when it drifts?

In the process era, those questions had answers because the rule was written once and enforced everywhere. In the edge era, the rule is learned many times and enforced locally. Consistency becomes something you have to engineer, not something you get by default.

That redistributes authority in two directions at once. Site managers gain decision rights they never formally held, because their model's behaviour is shaped by their data and their operational choices. And site managers lose decision rights they did hold, because the model now makes calls their supervisors used to make, and the supervisors cannot see inside it.

Neither shift is written down. The organisation chart says the depot manager runs the depot and head office sets policy. The architecture says the model sets policy at the point of work, and nobody owns the model.

Gartner's own advice on task-specific models points to the crux: enterprise data becomes the differentiator, which demands data preparation, quality checks and versioning. Read structurally, that means whoever owns the local data and its versioning owns the local

decisions. If that is nobody, the decisions are owned by default, and default is the most expensive owner an enterprise can have.

This is where the edge meets the argument in Data-Driven Architecture: who owns the data model owns the decisions. At the edge, there are as many data models as there are sites.

Whoever can change the model holds the decision

The edge looks decentralised. It is not. It is distributed execution under centralised change authority, and the party holding that authority is often not the enterprise.

Every edge model is governed in practice by whoever can change it. The update pipeline that pushes new weights, new thresholds and new software to thousands of devices is the most concentrated control point in the architecture. Whoever holds it can change every decision at every site at once.

Two public cases show both sides of that power.

In May 2024 a driverless Waymo vehicle struck a wooden utility pole in a Phoenix alleyway during a low-speed manoeuvre. No one was hurt. Waymo filed a voluntary recall with the National Highway Traffic Safety Administration covering 672 vehicles, its entire fleet at the time, and fixed it with a software update to how the system responds to poles and a map update. A fault observed in one vehicle was treated as a fault in all of them, because they all ran the same decision logic. The fix travelled the same way.

That is the update channel working as it should. The regulator treated edge decision logic as a product safety matter. The operator had the authority and the mechanism to correct it everywhere, quickly, with a record.

In July 2024 the same mechanism ran in reverse. A faulty content update to CrowdStrike's security software crashed Windows machines worldwide. Microsoft estimated that 8.5 million Windows devices were affected, less than one per cent of all Windows machines, yet airlines, banks and hospitals stopped. CrowdStrike is not an AI system, and that is the lesson. The incident had nothing to do with model quality. It showed what happens when a third party holds a privileged update channel into devices that sit at the point of work, and the enterprise has no staged rollout, no gate and no veto.

Edge AI multiplies the number of such channels. Every camera maker, controller supplier, robot builder and software vendor ships its own model and its own update path. Each one is a party that can change your operational decisions without passing through your governance.

So the question of who decides at the edge has a precise answer. The party that controls the update channel decides. The enterprise needs to know who that is for every class of edge decision it runs.

In many cases it will find that the answer is a supplier. Enterprises rarely design edge decision rights. They buy them, bundled inside equipment.

The original equipment manufacturer (OEM) ships the device, the model, the update path, the diagnostic tools and the data interface. Procurement buys a price per unit and a service-level agreement. Nobody at the table asks who may change the decision logic, who may read the data the device generates, or who may override it in the field. Those rights are allocated anyway. They go to the supplier by default.

Farming shows how far that default can go. In January 2025 the US Federal Trade Commission sued Deere, alleging that its equipment needed software repair tools it made available only to its authorised dealers. In July 2026 the FTC and five states reached a settlement requiring Deere, for ten years, to give farmers and independent repair providers the same repair resources, including software capabilities, that it gives its dealers: reading and clearing fault codes, reprogramming electronic components, restarting a machine after an emissions-related shutdown.

Read that as a decision-rights dispute rather than a repair dispute. The farmer owned the machine. The manufacturer controlled the software that decided whether the machine would run. It took years of advocacy and a federal lawsuit to move part of that control back to the owner. Now imagine the same allocation applied to the model that decides what your production line rejects, what your cameras flag or when your controllers shut a process down.

Regulators have seen this and are moving to the decision itself.

The EU Data Act has applied since 12 September 2025. It gives users of connected products the right to access the data those products generate, and limits the manufacturer's exclusive control over it. For an enterprise running edge devices in Europe, the data behind local decisions is now something it can claim, not something it must negotiate for.

The EU AI Act requires that high-risk AI systems can be effectively overseen by people while in use. Article 14 specifies that the people assigned to oversight must be able to disregard, override or reverse the system's output, and to stop it through a "stop" button or a similar procedure that brings it to a safe state. That is a requirement written for the edge whether or not its drafters thought of it that way. An override that has to travel to a vendor's cloud and back is not an override at the point of work.

The EU Machinery Regulation, which applies from 20 January 2027, is designed to cover AI modules that use learning techniques to perform safety functions in machinery. Learning safety functions on a production line become a matter of conformity assessment, not just engineering preference.

The direction is consistent across three instruments. Regulators are assigning accountability to the providers and users of these systems, and demanding that someone can see, override and stop it. Enterprises that let suppliers hold the defaults will carry the accountability without the control.

The supplier side of this, who controls access to the edge at the infrastructure level and why telcos and cloud providers are fighting over it, is the subject of Battle for the Edge. This paper stays with the enterprise's own decision architecture. The two cannot be separated for long: the partner who runs your edge infrastructure is a candidate holder of your edge decision rights.

The strongest objection: keep it central

The best argument against all of this is simple. Centralise.

Central models are more capable. A large general model handles a wider range of tasks than any small one. One model in one place is easier to evaluate, easier to audit and easier to fix. Edge deployment creates thousands of decision points, many of them in hardware the enterprise does not fully control, running model variants nobody reviews. Connectivity keeps improving. Why accept a sprawl of local decisions when the network can bring every decision home?

It is a serious argument, and much of it is right. Most enterprise decisions should stay central: credit policy, pricing, hiring, anything where consistency matters more than speed and where the input is already in a central system. Nothing in this paper says otherwise.

But the objection misreads the choice in three ways.

First, centralisation does not remove edge decisions where physics or economics require them. A sprayer moving at 15 miles per hour or a controller rejecting parts on a moving line will decide locally whether or not head office has a policy for it. The choice is not between a central decision and an edge decision. It is between a governed edge decision and an ungoverned one.

Second, the most sophisticated operators have already shown that the answer is tiering, not choice. Apple's design puts a small model on the device and a larger one on its own Private Cloud Compute, and routes work between them by design. The placement rule is itself the governance: what may be decided here, what must go there, and what the boundary guarantees. That is the authority envelope, written into the architecture.

Third, the real risk is not the edge. It is the undeclared edge. Most large enterprises already run edge decisions they did not choose: the camera with built-in analytics bought by facilities, the controller with a vision module commissioned by a line engineer, the vehicle telematics configured by a fleet supplier. Centralising your AI strategy does nothing about those. It only means nobody is looking at them.

The objection is right that unmanaged sprawl is dangerous. It is wrong that the cure is to pretend the decisions are not happening. The cure is to decide, explicitly, where each class of decision lives.

Structural moves

Leaders who want the intelligent edge to strengthen their control rather than dissolve it change six things, in this order.

First, inventory decisions, not devices. The asset register lists cameras, controllers and gateways. It does not list what they decide. Build the decision inventory: which decisions are made at the edge today, by which model, from which supplier, with what effect on customers, safety and money. Most leadership teams will find the list is longer than they thought and that half of it was never approved.

Second, place each class of decision deliberately. For every decision class, record where it must be made and why: latency, connectivity, sovereignty, safety or cost. Decisions with no reason to be local come home. Decisions that must be local are declared as edge decisions and governed as such. Placement becomes a governed choice, owned by someone, rather than a by-product of what the vendor shipped.

Third, write the authority envelope. For each edge decision class, define what the device may decide alone, within which parameters, when it escalates, who can override it at the point of work, how it is stopped safely, and what it reports. Make the override local. An override that depends on a vendor's cloud is not an override.

Fourth, take control of the update channel. Every path that can change edge decision logic, whether model weights, thresholds or software, goes through a gate the enterprise controls: staged rollout, test sites, rollback, a named approver and a record. Where a supplier insists on a direct path, treat it as a delegation of decision rights and contract for it as one.

Fifth, buy decision rights, not just hardware. Procurement terms for edge equipment cover data access, model access, diagnostic access, override rights and the right to switch models. In Europe, the Data Act already gives users a claim to device data; contracts should make that claim usable. Elsewhere, negotiate it before installation, while the enterprise still has bargaining power.

Sixth, redesign what flows upwards. The centre no longer receives everything, so specify what it must receive: exceptions, overrides, drift indicators, decision samples and evidence sufficient for audit. The central function's job changes from making decisions to supervising decision systems. Staff and measure it that way.

And stop doing four things.

Stop buying edge intelligence through facilities and operational budgets with no decision review. Stop treating edge AI as an IT pilot, when it is an operating model change at the point of work. Stop letting supplier defaults stand in for policy. Stop measuring progress by devices deployed, when the measure that matters is how many edge decisions have a named owner.

The edge is not getting closer. It is getting accountable. The only open question is to whom.

What to do about it

- **Inventory decisions, not devices.** List which decisions are made at the edge, by which model and supplier, and with what effect on customers, safety and money.
- **Place each decision class deliberately.** Record where each class of decision must be made and why, bring home the ones with no reason to be local, and declare the rest as governed edge decisions.
- **Write the authority envelope.** Define what the device may decide alone, within which parameters, when it escalates, who overrides it locally and how it stops safely.
- **Take control of the update channel.** Route every change to edge decision logic through an enterprise gate with staged rollout, rollback, a named approver and a record.
- **Buy decision rights, not just hardware.** Contract for data, model, diagnostic and override access before installation, and use statutory rights such as the EU Data Act where they apply.
- **Redesign what flows upwards.** Specify the exceptions, overrides, drift signals and evidence the centre must receive, and turn the central function into a supervisor of decision systems.

Sources

- IDC, Global spending on edge computing to reach nearly \$380 billion by 2028, 2025: <https://www.idc.com/resource-center/press-releases/idc-estimates-global-spending-on-edge-computing-to-grow-at-13-8-reaching-nearly-380-billion-by-2028/>
- Microsoft, Introducing Copilot+ PCs, 2024: <https://blogs.microsoft.com/blog/2024/05/20/introducing-copilot-pcs/>
- Apple Machine Learning Research, Updates to Apple's on-device and server foundation language models, 2025: <https://machinelearning.apple.com/research/apple-foundation-models-2025-updates>
- Drives & Controls, Module adds AI to industrial controls and I/O (Siemens S7-1500 TM NPU), 2018: <https://drivesncontrols.com/module-adds-ai-to-industrial-controls-and-i-o/>
- Gartner, Organisations will use small, task-specific AI models three times more than general-purpose LLMs by 2027, 2025: <https://www.gartner.com/en/newsroom/press-releases/2025-04-09-gartner-predicts-by-2027-organizations-will-use-small-task-specific-ai-models-three-times-more-than-general-purpose-large-language-models>
- John Deere, Customers use See & Spray technology across five million acres in 2025, 2025: <https://www.deere.com/en/news/all-news/see-spray-technology-across-5-million-acres/>
- AgNavigator, John Deere's See & Spray saves farmers more than 31m gallons of herbicide mix in 2025, 2025: <https://www.agnavigator.com/Article/2025/11/10/john-deere-uses-ai-to-slash-farmers-input-costs/>

- AZFamily, Waymo recalls 600+ self-driving vehicles for updates after one strikes pole in Phoenix, 2024: <https://www.azfamily.com/2024/06/13/waymo-recalls-600-self-driving-vehicles-updates-after-one-strikes-pole-phoenix/>
- Microsoft, Helping our customers through the CrowdStrike outage, 2024: <https://blogs.microsoft.com/blog/2024/07/20/helping-our-customers-through-the-crowdstrike-outage/>
- US Federal Trade Commission, FTC and states secure settlement with Deere & Company, advancing farmers' right to repair, 2026: <https://www.ftc.gov/news-events/news/press-releases/2026/07/ftc-states-secure-settlement-deere-company-advancing-farmers-right-repair>
- European Commission, Data Act, 2025: <https://digital-strategy.ec.europa.eu/en/policies/data-act>
- European Commission AI Act Service Desk, Article 14: Human oversight: <https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-14>
- EU-OSHA, Regulation (EU) 2023/1230 on machinery: <https://osha.europa.eu/en/legislation/directive/regulation-20231230eu-machinery>

Read online: bdgadvisory.com/perspectives/intelligent-edge