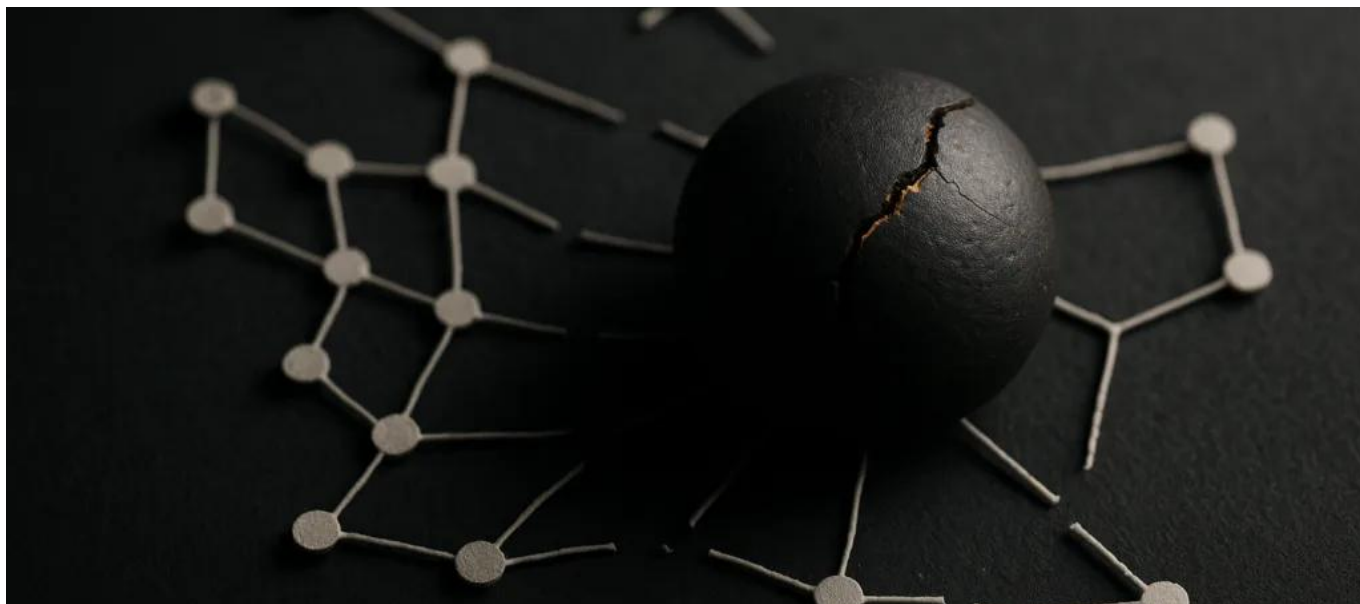




Postmortem: Azure's Sweden Central AI Outage and the 18-Region Gateway Failure 24 Hours Later

At 10:03 UTC on 29 September 2026, Azure OpenAI stopped answering reliably in Sweden Central. The region's whole AI data plane went with it, for five hours and fifty-five minutes.

Azure's status history records a 5h55m platform failure in Sweden Central on 29 September 2026 (10:03 to 15:58 UTC) that took down Azure OpenAI Service, Foundry Agent Service, Foundry Models and Cognitive Services at once. Roughly 34 hours later, at 20:30 UTC on 30 September, a separate incident degraded networking across 18 regions. Different causes, same consequence for European teams.

The duration is not what makes this pair worth writing about. Five hours and fifty-five minutes is a bad day, not a historic one. It is the correlation nobody has in their risk register: the set of regions chosen for legal reasons is a small set, and small sets fail together more often than large ones, for reasons that have nothing to do with a shared root cause.



Four AI services failed together behind a one-line “platform issue”

Here is what Microsoft actually published, and it is less than you want. The [Azure status history entry for 29 September 2026](#) gives a confirmed impact window of 10:03 to 15:58 UTC, names four affected services in Sweden Central (Azure OpenAI Service, Foundry Agent Service, Foundry Models, Cognitive Services), and describes customer-observed symptoms as intermittent request failures, increased latency and HTTP 5XX errors against affected models and data-plane APIs. The stated root cause is a “platform issue.” That is the whole technical disclosure. Microsoft noted the impact was mitigated and that a Post Incident Review is generally published within 14 days for impacted customers.

As of 6 October 2026, no tracking IDs, no Preliminary Post Incident Review and no final Microsoft root-cause statement for either incident were publicly retrievable. So anyone telling you what broke inside Sweden Central is guessing.

What the notice tells you by omission matters for compliance teams. The reported impact was to models and data-plane APIs hosted in Sweden Central, and nothing in the notice indicates prompts, completions or customer content were processed outside the region. This was an availability event. If your DPIA names Sweden Central as the processing location, that assertion survived the outage intact. Your SLA did not.

1 29 Sep, 10:03 UTC: Sweden Central AI plane starts failing

Intermittent request failures, elevated latency and HTTP 5XX against models and data-plane APIs. Four AI services affected simultaneously, which points at something shared beneath them rather than four coincident service bugs.

2 29 Sep, 15:58 UTC: impact window closes

Total confirmed duration 5h55m. Microsoft's published cause remains a generic platform issue; the PIR is promised inside the standard 14-day window.

3 29 Sep (date of snapshot): Azure Front Door shows unavailable

A status page snapshot around 29 September showed Front Door as unavailable. Available evidence does not link it to either incident. Treat it as a third, separate event and do not build a narrative on it.



Postmortem: Azure's Sweden Central AI Outage and the 18-Region Gateway Failure 24 Hours Later

4 30 Sep, 20:30 UTC: 18 regions, networking layer

A second incident begins, affecting 18 Azure regions. Reported affected regions include North Europe, West Europe, France Central, UK South, UK West, Germany North and Switzerland North. Services hit: ExpressRoute Gateway, VPN Gateway, Azure Firewall, Application Gateway, Web Application Firewall and Azure VMware Solution.

5 What Microsoft said about the second one

Public wording was that customers using ExpressRoute and/or VPN Gateway could experience "degraded or interrupted network connectivity." Third-party reporting attributes the event to infrastructure operating-system servicing activity and estimates roughly 5h45m of disruption. That attribution and that duration are secondary sourcing, not Microsoft-confirmed.

The second incident is a different animal from the first. The Sweden Central event hit the AI data plane. The 30 September event hit the paths into Azure: [ExpressRoute and VPN Gateway, plus Azure Firewall, Application Gateway, WAF and Azure VMware Solution](#). If your private circuit into West Europe is how your on-prem application reaches your Azure OpenAI deployment, this outage looks identical to an AI outage from where your users sit, even though every model endpoint stayed healthy.

! Multi-region failover does not help when the network you fail over is what broke
Six networking services degraded at once on 30 September, and the list includes both the private-circuit layer (ExpressRoute, VPN Gateway) and the ingress layer (Application Gateway, WAF). A retry policy that points at a second region inherits the same gateway dependency.

Two outages in European regions in one week is arithmetic, not evidence

On the evidence available, these incidents did not hit European regions because they were European. The 30 September event spanned 18 regions, of which seven named ones happen to be in Europe. The Sweden Central event was a single-region platform issue with no published cause. Nothing in the record supports a story about European infrastructure being weaker.



Postmortem: Azure's Sweden Central AI Outage and the 18-Region Gateway Failure 24 Hours Later

What the record does support is a selection effect on your side, not Microsoft's. EU and Swiss teams do not deploy across every Azure region. They deploy into a short list that satisfies data residency and, for Azure OpenAI specifically, a shorter list again where the models they need are actually available. Sweden Central, West Europe, Switzerland North, Germany North, France Central: plot both incidents against that list and the overlap looks like a pattern. A small denominator produces high hit rates.

5h55m

confirmed Sweden Central AI impact, 10:03 to 15:58 UTC

Azure status history, 29 Sep 2026

18

Azure regions affected by the gateway incident from 20:30 UTC

shattered.io / windowsforum, Sep-Oct 2026

~34h

between the start of the two independent incidents

derived from both status timestamps

~5h45m

estimated duration of the 18-region disruption (unconfirmed)

shattered.io, Oct 2026

Notice the symmetry in those two durations, and then resist it. 5h55m is Microsoft-confirmed with precise timestamps. ~5h45m is a third-party estimate for a different incident with a different cause. They are close enough to look meaningful and they are not. I mention it because that similarity is the kind of thing that ends up in an internal slide as "Azure outages run about six hours," which is a sample of two from one week.

My take: the lesson is not "go multi-region." Most EU and Swiss teams cannot do it meaningfully, because their legal footprint and the model availability map intersect in two or three places. Residency selection and availability design should be one decision made by one person at one time. In a lot of organisations they are two decisions made by two functions who never compare notes. Legal picks the region, platform picks the endpoint, and nobody writes down what happens when that endpoint returns 5XX for six hours on a Tuesday. The answer gets improvised during the incident.

Fix the degradation path, the fallback jurisdiction and the



network drill

Start with the degradation path, because that is the control you own regardless of how many regions you are allowed to use. For the 29 September pattern (intermittent 5XX, elevated latency against model endpoints), the questions are concrete: does your application distinguish a slow model call from a failed one, and does it have a defined behaviour for each? Intermittent failure with increased latency is the worst shape of outage, because naive retry logic amplifies it. A client that retries three times on 5XX against a region already shedding load turns a partial outage into your own denial of service.

Second, write down the residency boundary of your fallback before you need it. If Sweden Central is unavailable and your contingency is West Europe, that is a different jurisdiction decision and it should be pre-approved in writing, not debated at 11:00 UTC. If your contingency is "queue the work and process it when the region returns," then your SLA to the business needs to say six hours rather than minutes, and the business needs to have agreed to that in advance. Either answer is defensible. Having no answer is not.

Third, separate your AI failure drill from your network failure drill. The 30 September event is the reminder that connectivity into a healthy region can fail independently of that region's services. If you reach Azure OpenAI over ExpressRoute, test the public-endpoint path at least once, under policy, so you know whether it works and whether it is permitted. Discovering during the incident that your only permitted route is the broken one is a governance failure, not an Azure failure.

One cheap control: put the 14-day Post Incident Review window into your process. Microsoft states that a PIR is generally published within 14 days for impacted customers. Assign an owner to retrieve it, read it, and update the risk assumption for that region. As of 6 October 2026 no PIR for either incident was publicly retrievable, so for most readers this item is still open.

There is a documentation angle here too, and it will matter more as the EU transparency obligations land. If you are already mapping where your AI system runs and what it discloses, your residency and availability assumptions belong in that same artefact rather than in a separate platform runbook. I wrote about the December 2026 marking deadline in



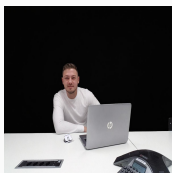
[a guide to shipping Article 50 marking](#). The practical overlap is that the same team is being asked to state, in writing, how the system behaves and where it runs. Add “and what it does when the region is down” to that statement while the document is open.

What the two incidents do not prove about Azure in Europe

I would not conclude that Azure's AI services in Europe are unreliable. Two incidents in 36 hours with no shared published cause is an unlucky week, and without the PIRs there is no basis for a structural claim. I would also not read much into the Front Door unavailability on the status page around 29 September: the available evidence does not link it to either incident, and the honest treatment is to call it a third, separate event and stop there.

What I will say with confidence is narrower. On 29 September 2026, four Azure AI services in one region failed together for 5h55m with no published technical cause. On 30 September, six networking services degraded across 18 regions, reportedly during infrastructure OS servicing. Both windows are long enough that any system with a hard real-time dependency on a single European AI endpoint had a visible, user-facing failure. If your architecture answered that with a queue, a cached response, or a documented degraded mode, you had a quiet week. If it answered with 5XX passed straight through to the user, the architecture made that choice for you, and it made it before the outage started.

Worth checking this week: which of those two descriptions fits the system you actually shipped, rather than the one in the diagram.



WRITTEN BY

Artur Markus

Software engineer in Basel building AI infrastructure and agentic systems, with a



Postmortem: Azure's Sweden Central AI Outage and the 18-Region Gateway Failure 24 Hours Later

background in pharmaceutical automation, GxP compliance, and validation. I help teams turn AI from a chatbot into a reliable operational layer.

[Book a meeting →](#)