Futurewerk

A Futurewerk Perspective

RETHINKING APPLICATION SUPPORT OUTSOURCING IN THE AGE OF AGENTIC AI

FROM CAPACITY-BASED DELIVERY TO AN AI-ENABLED SERVICE MODEL

The opportunity is not to replace FTEs with agents. It is to redesign application support around clearer work visibility, service-based pricing, shared productivity gains, and stronger governance.

CONTENTS

WHY AGENTIC AI REQUIRES MORE THAN ASKING PROVIDERS TO USE IT

Agentic AI can improve application support, but not by being inserted into the old FTE model. The real opportunity is to rethink how support is structured, priced, and governed.

Application support has long been one of the core areas of IT outsourcing. Many large support contracts have been built around FTE- or capacity-based offshore delivery models that gave enterprises scale, predictability, and managerial simplicity.

That model worked because it was easy to buy, govern, staff, and report against. But it also created a limitation: providers were responsible for keeping support running, while the contract did not always incentivise them to redesign delivery in ways that structurally improved productivity.

Application support is a strong candidate for AI-led productivity improvement, but enterprises cannot treat AI adoption as simply a question of asking providers to use agentic AI.

Operating within the existing FTE model while expecting providers to adopt agentic AI will only deliver limited benefits.

It places the burden of productivity improvement on providers whose revenues may be directly affected by reduced manual effort. More importantly, it risks using agentic AI only as a tool for incremental effort reduction, rather than as a way to rethink application support itself.

To capture the full potential of agentic AI, enterprises must move beyond the FTE model and redesign support around transparent pricing, shared accountability, and clear gain-sharing rules.

THE APPLICATION SUPPORT OPPORTUNITY

Application support is a natural starting point for agentic AI because much of the work is repetitive, knowledge-based, and process-led. But value depends less on the technology alone and more on the readiness of the support environment.

Application support teams handle recurring incidents, known-error patterns, monitoring alerts, ticket histories, runbooks, escalation paths, dependency maps, and application-specific workarounds. These are the conditions under which AI can assist, accelerate, or partially automate work.

But the opportunity should not be reduced to simple ticket automation. Application support is not a uniform body of repeatable tasks.

It includes low-risk, high-volume issues that may be suitable for automation; knowledge-intensive incidents where AI can support engineers with context and recommendations; and high-risk production issues where human judgement and accountability remain essential. The value lies in knowing the difference.

Agentic AI may help classify and route tickets, retrieve relevant knowledge, summarise incidents, identify similar past issues, suggest resolution paths, support root-cause analysis, and execute approved remediation for bounded problems.

It may also help reduce repetitive L1 and L2 effort, improve first-time resolution, and accelerate incident communication.

But it is far less suitable for poorly documented issues, complex L3 incidents, regulatory-sensitive workflows, high-impact production changes, or architecture-level failures without strong human oversight.

This distinction matters because the opportunity is not "AI in application support" in general.

The opportunity is to understand which parts of support can be AI-enabled, which parts should remain human-led, and which parts require a new operating and commercial model.

A support estate with reliable ticket histories, clear runbooks, integrated tools, and known resolution paths is far more AI-addressable than one where knowledge is fragmented, tacit, or poorly documented.

IT OUTSOURCING MARKET DYNAMICS: THE CHALLENGER PROPOSITION

AI-led challenger offers promise lower costs, but enterprises should not treat them as simple vendor replacement decisions. The real issue is whether AI-led productivity can be evaluated, priced, governed, and sustained.

Some challenger providers claim 30 to 40 per cent cost savings by replacing existing application support providers. But these offers are not as straightforward as they appear.

Many remain rooted in the same underlying logic as the model they seek to replace: reduce the manual effort currently provided by someone else, use agentic AI to absorb part of that work, and convert the reduction into a lower price.

That can create value, but it is still a narrow productivity argument. It does not, by itself, amount to application support transformation.

Nor should the market be understood as a simple divide between AI-led challengers and traditional incumbents.

Many incumbents are also investing in AI-native support models, automation platforms, copilots, agents, and delivery accelerators. Their challenge is different.

Incumbents have to transform delivery from within existing relationships, often where the commercial model is still tied to FTE capacity, effort, and staffing structures.

A challenger's lower-cost proposal must be tested against application knowledge, transition risk, documentation quality, service continuity, platform consumption, data security, and accountability.

An incumbent's AI proposition must be tested against whether productivity gains will actually flow through commercially, or whether they will remain trapped inside the existing commercial model.

The risk is that enterprises treat the challenger proposition as a simple substitution decision.

The real question is not only which provider can reduce effort. It is how the enterprise should redesign its support model as AI changes the relationship among work, effort, quality, cost, and accountability.

Many enterprises can report SLA performance but cannot explain the underlying work with enough precision to evaluate AI-led productivity claims.

STEP 1: BASELINE

Understand the work and AI-addressability

AI-enabled pricing cannot start with the provider's productivity claim. It has to start with the enterprise's support baseline.

A baseline should answer practical questions. Which tickets repeat most often? Which applications generate the highest effort? What is the split across L1, L2, L3, and production support? What is the current cost per ticket by severity, application, and issue type? How much effort is spent on manual triage versus resolution? Which incidents follow known resolution paths? Which runbooks are current and reliable? Where is documentation weak? Which issues are high-volume but low-risk? Which issues are low-volume but business-critical?

These questions matter because AI value is uneven.

A recurring access issue with clear resolution steps is different from an intermittent production defect affecting a payment flow.

A well-documented known-error pattern is different from an incident where knowledge sits only in the heads of a few experienced engineers.

A ticket that can be safely auto-remediated is different from one that requires judgement, stakeholder alignment, or regulatory awareness.

"We are seeing enterprises become more careful about where agentic AI is applied. The question is not only whether the technology works, but whether the support baseline is clear enough to show which workflows are ready, what risks are involved, what data is available, and how value can be measured commercially."

– Ali Touré, Sourcing Transformation Specialist

The baseline is the evidence base for deciding which work can be AI-enabled, which work should remain human-led, and which service units can be priced differently as AI changes the effort required to deliver them.

At minimum, enterprises should baseline ticket volumes by severity, category, and application; effort distribution across support levels; current automation; manual effort hotspots; resolution metrics such as mean time to resolve, first-time resolution, and reopen rate; documentation quality; runbook maturity; application criticality; and risk boundaries.

They should also identify the support activities that are too sensitive for automation because of security, regulatory, customer-impact, or operational-continuity considerations.

Not every enterprise will have this visibility at the outset, and not every provider will monitor or report it in the way required for AI-enabled pricing. In some cases, building the baseline may itself need to become part of the transition plan.

Traditional application support contracts have often been priced around capacity: the number of people, the location of delivery, the skill mix, the service window, and the agreed SLA structure. That model can still provide predictability, but it is not well suited to capturing AI-led productivity.

STEP 2: SERVICE UNITS

From FTE-based support to AI-enabled service units

Service units turn the baseline into a commercial model. They allow enterprises to move from buying capacity to buying defined support responsibilities.

If AI changes the effort required to perform support work, pricing only by FTE becomes increasingly limited. It makes capacity visible, but not necessarily productivity. It allows enterprises to govern staffing, but not always service improvement.

It can also make it difficult for providers to reduce manual effort without reducing the revenue base of the contract.

A service unit is a measurable unit of support work or support responsibility around which pricing, performance, and accountability can be organised.

In application support, possible service units may include tickets by severity, known-error resolution, application support by complexity tier, support by service tower, release support, monitoring and remediation, major incident support, or business-critical application coverage.

This is similar in principle to how enterprises have become used to consuming cloud services, but application support requires stronger quality, complexity, and risk guardrails.

A compute unit is relatively standardised. A support ticket is not. One ticket may take five minutes; another may require days of investigation. One application may be low-risk; another may support customer payments, regulatory reporting, or business-critical operations.

Service units therefore need to be designed from the baseline, not imposed generically.

This is where AI-addressability matters. If a category of tickets is high-volume, well documented, and suitable for automation, it may be priced differently from complex L3 support. If AI can assist with triage but not resolution, the service unit should reflect that distinction.

If an application is business-critical and requires human oversight regardless of AI capability, the pricing model should recognise the risk and accountability involved.

The move from FTEs to service units does not mean people disappear from delivery. FTEs may remain part of the provider's internal staffing model.

The enterprise no longer buys only capacity. It buys defined services, governed outcomes, productivity improvement, and the ability to benefit when AI reduces the effort required to deliver support.

For this shift to work, enterprises need to define their service catalogue more clearly. They need to understand the support activities they are buying, the variability across applications and ticket types, the level of human oversight required, and the quality measures that should apply. They also need providers to price services rather than simply provide capacity.

"To unlock AI productivity, enterprises must pivot from purchasing headcount to defining clear service units. This requires making volume, complexity, criticality, and risk visible—so performance is measured against the work delivered, not just internal capacity."

- Prem Balasubramanian, CTO and Head of AI, Hitachi Digital Services

The price of an AI-enabled service unit is shaped not only by the work performed, but also by platform ownership, AI consumption, governance effort, human oversight, quality risk, and the way productivity gains are shared.

STEP 3: PRICING MECHANISMS

Pricing mechanics for AI-enabled support

AI-enabled support changes what sits behind the price. Enterprises need to understand the AI environment, consumption costs, human oversight, and productivity gains that shape the real commercial model.

Consider a simple example. An enterprise may define "known-error resolution" as a service unit after analysing its baseline. The data shows that a set of recurring incidents accounts for a meaningful share of L1 and L2 effort, that the fixes are documented, and that many of them can be supported by agentic AI.

At first glance, this looks easy to price: the enterprise could move from paying for FTEs to paying a fixed price per known-error resolution.

But AI changes the pricing underneath that unit. If the provider uses its own AI platform, the cost of model usage, orchestration, monitoring, and tooling may be embedded in the service price. If the provider deploys agents inside the enterprise's Azure OpenAI, AWS, Google Cloud, ServiceNow, or other AI environment, token and compute costs may appear on the enterprise's bill instead.

The enterprise may then be paying the provider for the service unit while also paying separately for the AI consumption required to deliver it.

That changes the commercial question. The enterprise needs to know whether the price per service unit is calculated before or after AI consumption costs, how usage is attributed to specific support activities, what level of human oversight is included, how quality will be audited, and how savings will be shared if AI reduces the provider's manual effort below the original baseline.

"AI is moving from experimentation into enterprise-scale adoption and larger transformation engagements. The conversation is shifting, from isolated productivity gains to how organizations build, govern and scale AI to deliver measurable business outcomes. Clients are now focused on linking usage to value, with the right guardrails in place. That is where a clear commercial and value-realization model becomes critical to scaling AI with confidence."

- Manoj Mehta, President of EMEA, Cognizant

Service-unit pricing should not reward volume for its own sake. A pure cost-per-ticket model may create an incentive to process tickets rather than prevent them.

It should therefore be paired with guardrails such as incident reduction, first-time resolution, reopen rate, mean time to resolve, application stability, user experience, and business impact.

A productivity adjustment mechanism should be built into the agreement. If AI reduces effort over time, the commercial model should specify how and when the benefit is recognised.

The enterprise should receive a fee reduction or commercial credit when productivity improves beyond an agreed threshold, while the provider should receive recognition for creating that improvement.

Gain-sharing should be explicit. If a provider invests in AI-enabled support and reduces manual effort, the provider should not be penalised for reducing the very capacity on which the old contract was priced. At the same time, the enterprise should not pay the same fee while all productivity gains become provider margin.

A well-designed gain-sharing model allows both sides to benefit for a defined period, after which the improved run-rate becomes the new baseline.

AI consumption must also be governed. Where the provider owns the AI platform, consumption may be bundled into the service price. Where the enterprise owns the environment, consumption must be visible, attributable, and included in the net savings calculation.

Without this discipline, enterprises risk shifting cost from the provider contract into the cloud or platform bill without knowing whether the overall cost position has actually improved.

These mechanisms are harder to implement than they appear. They depend on reliable baseline data, agreed measurement methods, pricing transparency, and governance discipline.

STEP 4: CONTRACT TERMS

Turning the commercial model into contract terms

A commercial model only matters when it can be governed. The contract has to convert service units, AI usage, pricing, accountability, and knowledge ownership into enforceable terms.

The contract should begin with the service catalogue. It should define the support services in scope, the service units used for pricing, the application tiers, the ticket categories, the support levels, and the responsibilities of both enterprise and provider.

It should make clear which services are priced by unit, which remain capacity-based, which are outcome-linked, and which require separate treatment because of complexity or risk.

Not all AI use is the same. There is a difference between an agent that summarises incidents, one that recommends a resolution, one that executes an approved remediation, and one that closes a ticket.

The contract should specify what agents are permitted to do, what autonomy level applies, what human oversight is required, and what approval process is needed before AI use expands into higher-risk areas.

Pricing clauses need to reflect the new mechanics. The agreement should specify the service-unit price, the baseline against which productivity is measured, the treatment of AI consumption costs, the rules for calculating net savings, the gain-sharing mechanism, and the rebasing schedule.

If AI consumption sits on the enterprise's platform, the contract should define how usage is reported, attributed, capped, and included in commercial calculations.

Accountability must remain clear. AI should not dilute provider responsibility for service quality. If the provider uses AI as part of its delivery model, the provider remains accountable for the quality of service unless the contract explicitly defines a different control model.

The agreement should cover reopen rates, error handling, audit rights, escalation rules, incident review, and liability for incorrect or harmful resolution.

The contract should also address data, IP, and knowledge portability.

Enterprise-specific prompts, workflows, configurations, resolution playbooks, incident classification models, and support intelligence built on the enterprise's data and incident history should either be owned by the enterprise or portable through a perpetual usage right.

Provider-owned generic accelerators can remain with the provider, provided they do not embed enterprise-confidential knowledge.

Providers will often argue that configuration and orchestration layers constitute their IP even when built on client data. Enterprises should expect this negotiation and address it directly.

Application support is overdue for an AI reset. The commercial model must evolve alongside the agentic delivery model. Enterprises can no longer rely on legacy SLAs and productivity-based contracts. Future engagements should be built on transparent pricing, shared accountability, and value-sharing outcomes. Contracts must clearly define the role of AI agents, the guardrails governing their actions, how performance is measured and audited, and ownership of the intelligence they create. This is more than AI-enabled delivery. It is a fundamentally new operating model with measurable business impact.

- Omkar Nisal, CEO, Wipro Europe

GOVERNING THE TRANSITION

As AI-enabled support scales, enterprises need governance that tracks where agents are used, how much autonomy they have, whether quality is improving, and whether productivity gains are being captured over time.

Bringing agentic AI into application support means moving from one delivery and commercial model to another. That transition needs active governance.

Visibility. Enterprises should maintain an AI deployment register that identifies which agents are in use, which support activities they cover, what data they access, what autonomy level they operate at, and what human oversight applies. This register should be reviewed regularly, especially as agents move from low-risk augmentation to higher-impact support activities.

Quality and risk monitoring. AI-assisted and AI-resolved work should be tracked separately. Reopen rates should be monitored. Samples of AI-resolved tickets should be audited. High-risk categories should require human approval. Escalation thresholds should be defined. Enterprises should have visibility into where AI improves performance and where it introduces new risk.

Consumption transparency. Token, compute, orchestration, and platform usage should be reported against support activities, not buried inside a broader cloud bill or provider service charge. Enterprises should be able to see whether AI-enabled delivery is improving the overall cost position, not simply shifting cost from the provider contract into another bill.

Rebasing discipline. Productivity gains should not remain permanently outside the contract baseline. If AI reduces effort and improves service performance, the improved run-rate should eventually become the new baseline. This prevents gain-sharing from becoming a permanent premium for a one-time improvement and ensures the enterprise captures the structural benefit over time.

Change control. As providers introduce new agents, expand autonomy, connect additional systems, or move into more sensitive workflows, the enterprise should approve those changes through defined governance. Agentic AI should not quietly expand from low-risk support assistance into high-impact decision-making without explicit review.

"As DevOps blurs the traditional divide between development and support, AI cannot simply be bolted onto old operating models. The next shift is to integrate AI and agents end-to-end within product and platform models, with clear accountability for speed, cost performance, quality, and outcomes."

- **Ashish Gupta,** Head of EMEA, HCL Technologies

THE TARGET STATE: AN AI-ENABLED SERVICE MODEL

The target state is not AI-assisted capacity. It is a support model where productivity is visible, pricing is transparent, accountability is preserved, and gains are shared.

The target state is a different application support model, one where the enterprise has greater visibility into the work, the provider has a clearer path to improve delivery, and both sides have a commercial structure for sharing the value created.

In this model, AI-enabled productivity is no longer treated as a vague promise or a hidden delivery improvement. It is linked to defined services, transparent pricing, measurable quality, visible consumption, and agreed commercial adjustments. FTEs may still exist in the delivery model, but they are no longer the primary way value is understood or priced.

This changes the relationship between enterprise and provider.

The provider is rewarded for improving delivery, not preserving capacity. The enterprise captures productivity gains without simply pushing margin pressure onto the provider.

AI can be introduced and expanded, but within clear boundaries of oversight, accountability, risk, and governance. That is the real shift: from resource-based support to an AI-enabled service model.

The next generation of application support contracts will not be defined by how many agents replace how many FTEs. It will be defined by whether productivity is visible, pricing is transparent, accountability is preserved, and the gains are shared and sustained over time.

Futurewerk

Accelerating
Enterprise
Innovation

LONDON | HYDERABAD

CONTACT US
info@futurewerk.com