The AI-native service provider

From economic pressure to a strategic advantage

A pragmatic framework for a successful AI journey

The need for value: Generating revenue versus spending to modernize

The telecommunications (telco) industry is caught between the intense pressure to reduce operational expenditure (OPEX) and the necessity to generate new revenue growth.

Artificial Intelligence (AI) is no longer an experimental technology, but the primary catalyst for navigating these challenges. AI-native value promises to revolutionize current business and operational models with better network operations, customer experience, and security.

Moving from isolated experiments, pilots, and proof-of-concepts to enterprise-wide deployment has its challenges. Those challenges range from high infrastructure costs and fragmented data complexity to an overall lack of trust in AI reliability. AI tools and software need specific guardrails and governance. Many promising AI initiatives stall, failing to deliver a return on investment (ROI) which reinforces organizational skepticism.

The economic value of AI is outlined in this overview, along with a pragmatic framework for a successful AI journey. The framework argues that for AI technology to flourish, its deployment needs to be economically viable, practical, and transparent across the datacenter and network edge.

The service provider AI imperative: A market in transformation

Two powerful and simultaneous imperatives are shaping the telco market.

On one hand, service providers face escalating operational complexity stemming from managing intricate multivendor network architectures and maintaining aging systems to defend against competition and sophisticated cyber threats. This complexity has pushed traditional, manual processes beyond their breaking point and created an urgent need to streamline operational processes to reduce operational expenditure (OPEX) and free up resources for innovation.

On the other hand, the telecommunications industry is fighting to avoid the commoditization of connectivity. The strategic imperative is to find new growth engines, increase average revenue per user (ARPU) through personalization, and reduce costly customer churn.

AI is not a single solution but a dual-purpose tool. For network operations, it is a measure to ensure efficiency and survival. For commercial teams, it is a tool for growth. Any successful strategy must address both, framing every AI investment in the language of its specific economic impetus, be it cost reduction or revenue generation.

From theory to practice: High-value target areas for AI

Many service providers are actively deploying AI across several high-value domains where it can make a tangible effect on the dual pressures of cost and growth.

1. **Network automation and predictive maintenance:** This is the most critical area for AI adoption. AI is essential for creating autonomous intelligent networks that reduce manual intervention, predictnetwork faults before they occur to lower mean time to repair (MTTR), and optimize networks in real time. This directly addresses the challenge of cost reduction.

If it does not make cents, it does not make sense.

2. **Customer experience and churn reduction:** AI-powered chatbots are lowering support costs, while predictive analytics identify at-risk customers, allowing for proactive retention campaigns. Hyper-personalization of services and offers is a key lever for increasing ARPU and profitability.
3. **Security and risk management:** As networks become more distributed, the attack surface expands. AI is no longer optional. It is an essential defense mechanism for detecting threats, identifying fraud, and managing vulnerabilities across a complex network in real time.

The reasons why AI is essential become clear and compelling. However, the challenge that stalls progress for many service providers is how to get started.

The path to value: Pillars for a pragmatic AI strategy

Bridging the gap between promising pilots and delivering value in production requires a deliberate, pragmatic AI strategy.

The following 3 pillars provide some guidance to more successfully scale service provider AI initiatives.

The economics of service provider AI

The value of AI is in production, not in the lab.

The first question for any AI initiative must be about its economic viability. The significant cost of specialized AI infrastructure like graphics processing units (GPUs), coupled with the risk of running inefficient, oversized models, can quickly erase potential gains. Service providers must shift their focus from chasing the biggest, most hyped models to deploying the right-sized models that deliver a clear and compelling ROI, and align to both their business and operational objectives.

This requires a pragmatic approach, constantly asking: Is the problem being solved worth the cost? It means embracing an AI shrink-ray[1] philosophy and championing smaller, more cost-effective, and powerful open-source models that can be run efficiently through techniques like quantization. A sustainable AI strategy must be economically viable and relentlessly focused on optimizing the total cost of ownership.

An operational blueprint for the modern service provider

For service providers, the production environment is a distributed, hybrid, and multicloud reality. AI cannot live solely in a central cloud. For low-latency use cases, real-time analytics, and to meet the various facets of digital sovereignty, AI workloads must have the flexibility to be deployed where they are needed and make the most impact. This requirement creates immense operational complexity.

The key to unlocking AI's potential is to industrialize it. A systematic deployment of AI tools and systems, transforming traditional workflows and business models. This requires a common platform that can orchestrate both AI and network workloads from the core datacenter to a cloud environment and out to thousands of edge locations. Success depends on building an "AI factory", a dedicated environment or platform designed to streamline the development, deployment, and management of AI models and applications. Such a factory can reliably manage the entire AI/MLOps lifecycle and agentic AI, including build, train, serve, and monitor, across this complex, hybrid environment. The potential for AI is only unlocked when it is operationalized and deployed where it is needed most.

1 *AI in telecom."AI shrink-ray finds its telco edge." RCR Wireless 18 July 2025.*

The open source path to trusted AI is collaboration

Many service providers are trapped in a vicious cycle. An overly cautious engineering-focused cycle combined with a fear of regulatory noncompliance, high technical debt, and a lack of in-house skills, leading to standalone, underperforming projects. These failures reinforce the initial skepticism and stall further investment.

Breaking the recurring cycle of distrust.

Trust is the currency of AI adoption, and openness is the antidote to distrust. An open-source approach to AI fosters the transparency needed to build confidence. It promotes vendor diversity, giving service providers control over their own destiny. Most importantly, it provides the foundation for auditable, governable, and understandable solutions that can meet strict regulatory requirements for transparency and human oversight, turning a compliance hurdle into a source of competitive advantage.

The value of Red Hat: Your trusted partner for the AI-native future

Helping you execute a strategy that goes from principles to the complete platform.

Red Hat AI provides the complete platform and pragmatic strategy to help service providers execute on these 3 pillars and turn architectural principles into production reality.

- **Mastering economics:** Red Hat® AI Inference Server delivers on the AI shrink-ray promise. With tools like LLM Compressor and efficient GPU optimization, it allows service providers to run models at scale with drastically lower infrastructure costs. A benefit that directly addresses the core economic challenge of production AI at scale.
- **Operationalization:** Red Hat OpenShift® AI provides the single, consistent MLOps platform to build, deploy, and manage AI across the entire hybrid landscape. It is the foundation for the AI factory, which supports the industrialization of AI from the core datacenter to the farthest telco edge.
- **Building trust:** An open source foundation is the bedrock of trusted AI. Offerings like validated models with capacity planning guidance reduce adoption risk by removing guesswork. Red Hat platforms have built-in governance features that provide the transparent, auditable foundation service providers require to meet regulatory demands and build confidence across their organization.

Red Hat provides the complete foundation and a pragmatic AI strategy to help service providers navigate the complexities of AI, turning the greatest challenges for the telco industry into its most significant opportunities.

Begin your pragmatic AI journey today

To learn more about how Red Hat can help you build your AI-empowered future, visit us at Red Hat AI or contact our experts.

About Red Hat

Red Hat helps customers standardize across environments, develop cloud-native applications, and integrate, automate, secure, and manage complex environments with award-winning support, training, and consulting services.

facebook.com/redhatinc
@RedHat
linkedin.com/company/red-hat

North America
1 888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com

redhat.com
#2562570_0825