Charlotte Patrick
Research

Scaling Agentic AI

Charlotte Patrick, Independent Analyst

August 2026

Sponsored by Netcracker Technology

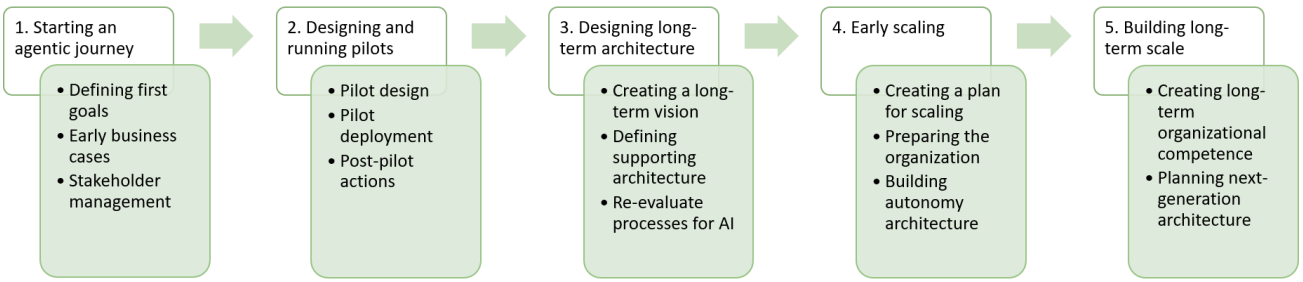
Delivering the Digital Future
Netcracker
An NEC Company

Abstract: This research note sets out a roadmap for the deployment of agentic systems in telecoms. In the early stages, the roadmap resembles that of many other telco transformation programs; however, the complexity of these systems makes scaling a significant challenge. The note looks at what a successful telco should be working on at each stage.

A pathway to building agentic scale

Today, telcos typically have a number of agents in pilots and production, sitting somewhere around Stage 2 of a 5-stage process for developing agentic systems, as shown in Figure 1. The first section of the report considers the requirements for creating solid foundations that will help agentic systems scale in future.

Figure 1: Five stages to developing agentic systems



Source: Charlotte Patrick Research

Stage 1: Preparing the agentic journey

At Stage 1, the roadmap looks much like any other deployment of new technology: a suitable first use case is identified, the business case is tested, stakeholders are consulted, success metrics are defined, and sufficient governance, security, and oversight are put in place to ensure that the initiative does not carry significant risks.

This stage is also about learning rather than scaling:

> *What are the characteristics of good early use cases – and what processes are not suitable?*
>
> *Where are the biggest challenges associated with agentic AI going to be?*
>
> *What capabilities are required beyond simply creating a few agents*

Stage 2: Designing and running pilots

Bringing together learnings from both the telcos and other verticals also creating agentic systems, a pilot should:

Take baby steps: Starting small with a "minimum viable pilot" rather than an ambitious transformation program. Early deployments should focus on simple, low-risk use cases within a single department or workflow, allowing teams to gain experience while limiting risk. Rather than creating entirely new processes, the first agents should be introduced into existing workflows and automations where past work has already dealt with issues such as data availability and lack of defined processes.

Plan for a successful agentic pilot: A successful pilot focuses on a well-defined set of limited tasks automated under human supervision - and refines, refines, refines. This iterative improvement, via human feedback and retraining loops, allows agents to steadily improve performance, reduces errors, and avoids undesirable behaviors such as excessive consumption of infrastructure resources. The pilot should also be used to establish the first reusable autonomy architecture assets (this concept encompasses the set of technologies

and processes needed to support an agentic system). Typical projects include additional developments to existing data stores and the first knowledge assets such as a knowledge graph.

Basic operational risks should be anticipated from the outset with agent rollback and fallback mechanisms, human intervention points, and appropriate governance and monitoring.

All of this considerable development work will mean that the pilot provides the telco with its first set of technical, operational, and business skills around agents; as well as practical tools such as reusable templates, standards, and reference architectures that will allow subsequent agentic projects to scale more quickly and consistently.

Measure pilot success: It is important to find ways to measure success to demonstrate business value and support future investment. Clear success criteria and KPIs should be defined from the outset, covering operational performance, financial benefits, customer experience and employee productivity. There are also new types of measures needed specifically for agents, including the ability to monitor, log, and trace agent behavior, decision quality, and whole-system performance. Establishing these types of robust measurement and observability from the outset helps distinguish pilots that deliver demonstrable value from those that don't.

Take post-pilot learnings: As the pilot concludes, telcos should take a little time to reflect on what has been learned and identify what needs to be addressed before broader deployment. This includes gathering feedback from users and stakeholders, assessing operational, technical, governance and adoption challenges, and understanding how agents behave in real-world environments rather than controlled demonstrations.

Building the long-term roadmap in Stage 3

As pilots progress, discussions turn to the scaling of agentic systems and a question arises:

"A scaled agentic system requires significant investment; how can we minimize cost while pushing forward?"

To tackle this question, telcos should take a step back before the pilot phase ends and spend time developing a longer-term vision. The left-hand side of Figure 2 sets out the big-picture business decisions that are needed, and the right-hand side considers some of the major technical decision areas, many of which are new to a telco.

Figure 2: The elements of a long-term autonomy roadmap

Source: Charlotte Patrick Research

Considering the various activity areas shown in Figure 2:

Prioritize projects by value, risk, and feasibility

It is still "early days" for agentic system deployment, and the telco will benefit from a short, formal document describing what the company expects to realistically achieve with autonomy, a list of governance and other concerns that need to be kept in mind when making decisions on projects, and a list of priority projects. This list should have a calculation of the potential financial upside, risks and a view of feasibility for each project.

Pain point identification

A formal exercise which looks at the range of potential risks generated by an agentic system – and the available resolutions. This looks set to be an important part of the long-term roadmap due to the wide variety of risks (highlighted in recent work); and should include issues seen during pilots which still need work (e.g. data or knowledge gaps) and potential major future issues. This will help prioritize work items on the roadmap and enable a view of potential costs.

Long-term business model

Taking the financial upside and potential costs for each project from the steps above, a long-term business model can be generated. Projects including agentic systems are likely to be subject to the same sort of value generation rules as past automation projects, with projects around energy saving or reducing the workload of larger teams generating good value. The complexity of creating certain agentic systems (e.g. where the environment often throws up unusual scenarios or there are complex sets of factors that impact the outcome of a decision) may make certain projects non-viable in the early days. Non-viable business cases should then become part of an exercise which considers a future where costs decline due to factors such as the use of smaller, more efficient models, better model routing, experience with creating knowledge assets and suites of pre-built tools for managing agent error.

What about sustainability?

While cost is crucial, environmental impact should also remain a high priority. Exercises such as calculating energy budgets for individual processes, particularly where agents operate continuously or use complex multi-agent reasoning, will be helpful.

Netcracker ROI Framework

A Positive ROI for Every Use Case: AI in telecom can only succeed when it delivers measurable value. Netcracker guarantees a positive business outcome for every deployment through a structured framework based on experience and best practices:

- √ *Identify and prioritize the right use cases based on value to the business, clear KPI improvement goals and ROI expectations*
- √ *Assess readiness and dependencies, including data and API availability, data quality, LLM selection and MCP integration*
- √ *Deploy with telecom expertise, ensuring seamless integration with BSS/OSS systems and customer-facing channels*
- √ *Continuously measure agent performance, correlate with business KPI goals and optimize agents*

With these measures in place, aided by dynamic and adaptable user interfaces providing a real-time view of LLM and agent performance, Netcracker helps service providers move quickly from pilot to production with the confidence that each use case is driving measurable returns.

Identify best long-term technologies and target architecture

A well-defined autonomy architecture is needed to support the long-term agentic roadmap set out in the steps above. This exercise is likely to take longer in those parts of the architecture, such as the knowledge plane, which have the potential to quickly accumulate technical debt. The exercise should begin with assessing the current maturity of the various layers (particularly the more complex data and knowledge requirements) to identify gaps that will constrain future progress. Based on this assessment, telcos should define a target architecture and migration and modernization roadmap. This roadmap will need to be regularly revisited as it contains a range of emerging concepts and technologies that technical teams will need to understand and track as they mature over the next few years.

Harnesses and boundaries

OpenAI defined the term harness engineering in 2026 to describe a working environment around an AI agent that allows it to perform reliably while humans steer. Several (sometimes overlapping) concepts set out different risk reduction activities; for example, bounded autonomy allows an agent to act independently, but only within predefined goals, permissions, policies, and approval thresholds. These practices are fairly nascent but key to creating agentic systems that work and will require telcos to develop new skill sets.

Scaling and reuse

As discussed previously, options for telcos to reduce cost while allowing scaling will be key. These include using proven agents from other verticals and pre-created vendor agents, whilst also growing their own agent catalog and a library of reusable agent skills. Thinking in this area should also focus on delivering high-quality data and knowledge assets: complete documentation, well-structured knowledge, and consistent data models etc.

Memory and state

As autonomous systems become more capable, they need to maintain an understanding of what has happened previously, what is happening now, and what should happen next. Telcos will need to develop explicit approaches for managing both memory and state. Memory enables agents to retain and retrieve relevant knowledge from previous interactions, decisions and outcomes, while state allows them to understand the current context of tasks, workflows, network conditions and other agents.

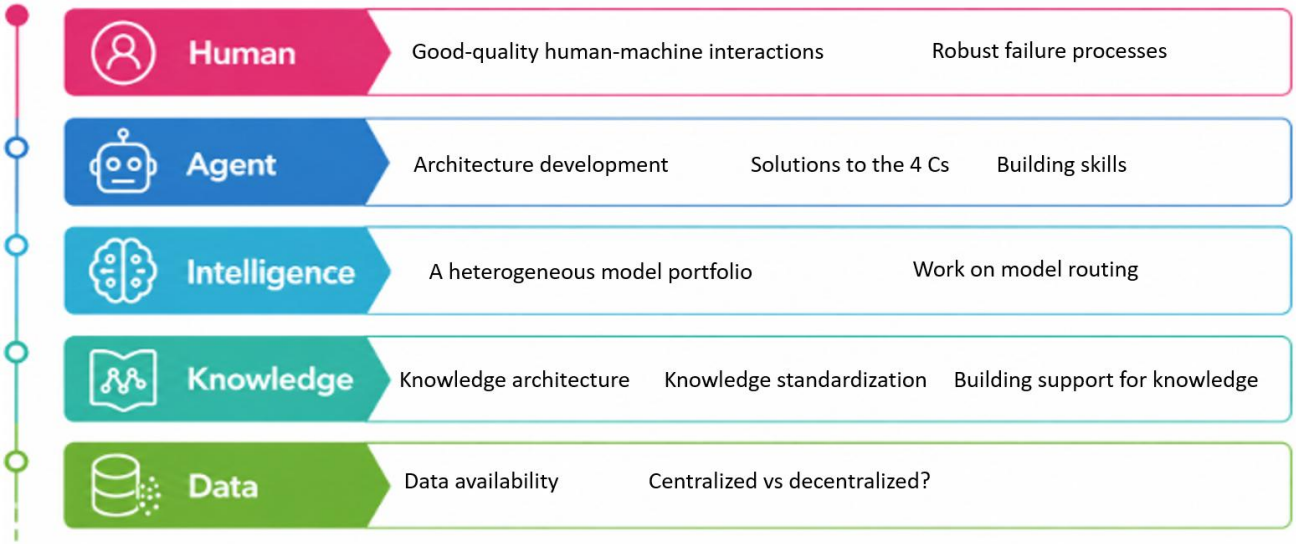
Autonomy-first design

Autonomy-first design is an emerging mindset for designing processes around autonomous systems rather than around people. While no settled industry term yet exists, the underlying principle is pragmatic: first ask *Can AI do this?* and *Should AI do this?* Where autonomy is appropriate; processes, information, interfaces, controls and operating environments should be redesigned to support machine execution, rather than adapting agents to workflows originally created for humans. Humans are then introduced deliberately where intent, judgment, accountability or exception handling are required.

Developing a solid foundation to enable scale in agentic systems

Stage 4 of the roadmap sees the telco beginning to tackle challenges on multiple fronts in order to enable future scaling. Figure 3 sets out the main layers of an autonomy architecture. There are many actions needed at each layer, and this section picks out some of the most pressing for a telco looking to scale their agents and discusses them in more detail below.

Figure 3: Autonomy architecture action items to enable scaling



Source: Charlotte Patrick Research

Human

The human is placed at the top of the autonomy architecture to emphasize their central role in agentic systems. Work items in this layer include creating good-quality user experiences and enabling human-facing systems (such as the copilot) to continue learning and improving over time.

Another key area considers how humans and machines work together to prevent/deal with situations where agentic systems fail or produce unexpected outcomes. Telcos need to develop clear policies on acceptable error rates and decisioning around when human verification is required. In addition, formal interactions between humans and machines will include feedback loops that allow users to explain what went wrong and how outcomes should have differed, and "clarification loops" that enable agents to recognize potential issues before they happen and seek guidance from humans or other agents before taking irreversible or high-impact actions.

Agents

Most discussions in this layer relate to the architecture and design of agents. As previously, architectural developments will need to support the scaling of agentic systems in areas such as the efficient use of compute and model resources. Also, resilient and predictable agentic systems require the four Cs: control, coordination, conflict resolution, and correction (error!).

To enable this, decisions will be needed around:

- how agents should be structured, how their roles and capabilities should be defined, and the boundaries within which they should be permitted to operate
- orchestration: how tasks are decomposed, agents and tools are selected, and activities are best coordinated (especially if agents begin to work in more complex peer-to-peer relationships)

- communications between humans, agents, and tools - how can tasks, context, and results flow between these different entities.

Tackling these topics will require a broad mix of skills:

- some solutions are still relatively nascent, and the planners and architects of agentic systems need to keep up-to-date with developments and the vendors that look likely to provide solutions
- some activities require process redesign or knowledge that is held in operational teams working at the day-to-day level; they need to be bought into discussions around agentic system design
- external skills may also need to be bought into the organization; again, planners should keep up to date with discussions in this area to enable them to make informed choices of external consultancy
- sitting underneath these discussions of agentic system architecture are a wide range of tools needed to support the agentic system; including agentic lifecycle management, governance, monitoring and observation, security, access management, etc. Skills will need to be bought in via vendor solutions and associated consultancy.

Intelligence

The term "minimum viable intelligence" is used to suggest that the least intelligence possible should be deployed to achieve each individual task in an agentic system. To achieve this, telcos will increasingly adopt a heterogeneous portfolio of models that support agents with rules, machine learning, symbolic reasoning, small and medium-sized (on-premises) models, and large foundation models. During recent research interviews, the development of small language models was discussed less often than expected; however, broader market discussion argues that many agent requests are narrow, repetitive, and well-bounded, making these small models increasingly economical and operationally appropriate for agent systems.

Model routing is also emerging as a core capability of the intelligence layer. Rather than relying on a single large model, systems will dynamically select the most appropriate model based on cost, capability and operational requirements. This selection will increasingly need information from the other layers of the autonomy architecture for context (e.g. task state, previous execution failures and available tools).

Knowledge

The knowledge plane transforms raw operational data into structured knowledge that provides the context agents need to make good-quality decisions. It is likely to become one of the most technically challenging

layers of the autonomy architecture, requiring telcos to combine knowledge graphs, vector databases, digital twins and other technologies into a coherent knowledge ecosystem.

One of the largest areas of work will be designing a scalable and maintainable knowledge architecture. Knowledge graphs are emerging as a core technology, but they can become complex, expensive to maintain and difficult to scale if not carefully designed. Graph design techniques under discussion include keeping graphs small and specific, with federation of knowledge across domains, as needed. Also implementing AI to build and maintain them. There is also an industry-wide conversation about standardizing a common ontology to enable easy knowledge ingestion in a multi-vendor environment.

As the knowledge plane becomes central to autonomous decision-making, its performance, resilience and trustworthiness will become as important as the intelligence models that consume it. Another major work area, therefore, includes building common semantic models, governance, privacy, and security controls, as well as mechanisms to explain where knowledge originated and why it influenced an agent's decisions.

Netcracker's Telecom Knowledge Layer

With deep telecom know-how built on 30+ years in the industry, Netcracker's telecom knowledge layer grounds AI agents in real domain expertise, delivering significant results to the telecom business.

Knowledge Base: *Connects agents to the vast body of unstructured telecom knowledge, including documentation, procedures and domain expertise.*

Semantic Ontology: *Creates a common understanding of telecom concepts, relationships and dependencies, enabling AI to understand how products, services, customers, resources and network elements relate to one another.*

End-to-End Reasoning: *Combines causal analysis with real-time operational data to ground AI decisions in the broader business and network context, beyond probabilistic LLM reasoning.*

In conclusion

In order to support the development of future agentic solutions, telcos need to work in a structured manner:

1. Begin with clearly defined, low-risk use cases where processes, data and operational responsibilities are already reasonably mature
2. Stage 2 pilots should combine limited agent deployment with the creation of reusable foundations for the next generation of agents. The pilot should end with a formal assessment of what must change before broader deployment
3. Step back and address the long-term architecture. This includes building an understanding of the wide range of risks likely to be seen and developing solutions around the most challenging
4. Greater autonomy requires stronger—not weaker—controls. Telcos should focus on understanding and building skills in areas such as harness engineering, bounded autonomy, AgentOps, and the design of agentic architectures.
5. Scaling of systems will require development of reusable agent skills, knowledge assets, interfaces and operational controls rather than repeatedly building complete systems from scratch. Consideration will be needed across all agentic projects to ensure that the correct assets are built and technical debt does not accrue.