Ericsson White Paper
GFTL-26:001102 UEN
July 2026

ERICSSON

Telecom foundation world models—Toward higher levels of network autonomy

Content

Introduction

Communications service providers (CSPs) are on a journey toward autonomous networks. The motivation rests on three pillars [1, 2]:

1. cost savings, where current manual tasks are automated
2. additional revenues from improving the quality of existing services through automated service assurance
3. new revenues from new services, enabled by solutions that allow versatility and thus offer a wide range of tailor-made services to end customers with a fast time-to-market

TM Forum has introduced different levels of autonomy, and the primary goal for CSPs is to transition toward full autonomy. To achieve higher levels of autonomy, improvements must be made in all phases of the autonomous network lifecyle: planning, deployment, service provision, maintenance, and optimization [2].

In autonomous networks, predictive analysis plays a major role [3]. Artificial intelligence (AI)-driven models predict action outcomes to guide network decisions, with continuous learning. The challenge CSPs face is to implement and integrate AI models and solutions that target a wide range of use cases without increasing complexity and cost.

To succeed in the journey toward fully autonomous networks, AI agents are needed to support decision-making by predicting and reasoning about the future state of the network. Today's technology features large language model (LLM)-powered AI agents that offer flexibility. However, they lack the reliability required to accurately predict network key performance indicators (KPIs) or assess the impact of actions on the network. On the other hand, traditional machine learning (ML) prediction models are often scoped to a narrow set of network KPIs and fail to capture the full complexity of the network.

In this paper, we promote the vision of telecom foundation world models: models trained on a wide variety of network data that can predict how the network state evolves in response to candidate actions, enabling improved decision-making and increased autonomous capabilities.

A world model is an AI model that has the capability to predict the state of the world—in this case, the network and its environment—when applying a sequence of actions and is designed to assist in decisions to carry out modifications in the network. World models can be used to evaluate the impact of candidate modifications before they are applied.

Foundation models are AI models trained on massive datasets. They are widely used in the text and image domains (LLMs and image processing) and are starting to be adopted by the telecom industry. Telecom foundation models [4, 5] are pre-trained and fine-tuned on various network data ranging from network management and orchestration to physical layer data. After extensive pre-training, and adaptation to specific use cases, these models can be applied to a wide range of tasks.

World models can rely on foundation models to predict of the impact of actions. We refer to them as foundation world models. Utilizing foundation models to build world models brings the advantage of generalizing to a wide range of use cases.

Telecom foundation world models will enable CSPs to move toward fully autonomous networks thanks to their broad understanding of the network architecture, deployments, performance, and characteristics, and primarily, their understanding of how performance depends on network configuration.

Telecom foundation world models in the management domains are enablers in many automation use cases, such as intent management functions, network digital twins (NDTs), and radio access network (RAN) applications (rAPPs) in service management and orchestration.

We believe these challenges can be addressed using telecom foundation world models, which enable prediction of the impact of changes in the network for a wide range of use cases in all phases of the autonomous network lifecycle.

Telecom foundation world models

World models

A world model predicts how an environment will respond to a chosen action. Given the current state of the network and a candidate action, it predicts the resulting state, as shown in Figure 1. This action-conditional prediction is what distinguishes a world model from a model that only forecasts the future based on past input data, and it is what makes a world model useful for proactive decision making. The world model does not itself decide or plan. It provides predictions that agents, optimizers, and control algorithms use to choose actions and to evaluate their consequences.

In telecom, the world is the network and its environment. A telecom world model captures how network configuration changes and external factors such as shifts in user behavior or unexpected weather conditions affect network KPIs. The training data must include actions taken in the network together with observations of the network state before and after, and it must be sufficiently diverse so the model can disentangle the effect of actions from external factors. Such data is scarce in real deployments today. World models for telecom have been identified as a powerful engine for NDTs and predictive support for operational decision-making in 5G Advanced and 6G [4, 6], and they have been demonstrated in specific use cases such as RAN energy savings and RAN optimization [7, 8].
The approach is promising, but industrial-scale deployment remains open.

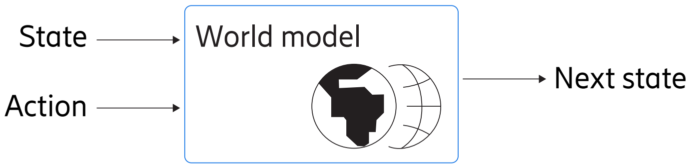

Figure 1. World model

Foundation models

A foundation model is a large neural network trained on broad and diverse data that can be specialized for a wide range of downstream tasks, as illustrated in Figure 2. Once pre-trained, the model can compress its inputs into a compact internal representation, an embedding that captures the most important patterns in the data. Foundation models are mainstream in text and image use cases and can also be trained on time series or other types of data.

Telecom foundation models can be pre-trained on telecom data such as network topology, fault, configuration, accounting, performance, and security records (FCAPS), internal data from network functions, infrastructure data, and air-interface data such as in-phase and quadrature (IQ) samples [4, 5]. Other useful sources include text from product documentation and standards specifications, and synthetic data produced by network simulators. Telecom foundation models can be adapted to a wide range of network management, automation, and research and development (R&D) tasks, including forecasting KPIs such as cell throughput or latency, anomaly detection, and root cause analysis. A foundation model often predicts what is likely to happen next given past data and does not by itself predict the effect of a future action—a capability that proactive decision making requires. Acquiring that capability depends on rich action-response examples in the training data, which typical pre-training corpora do not contain.

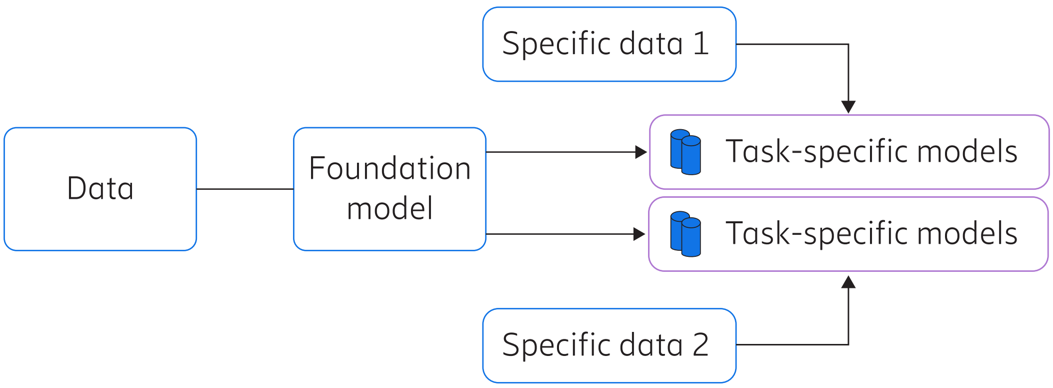

Figure 2. Foundation model

Telecom foundation world models

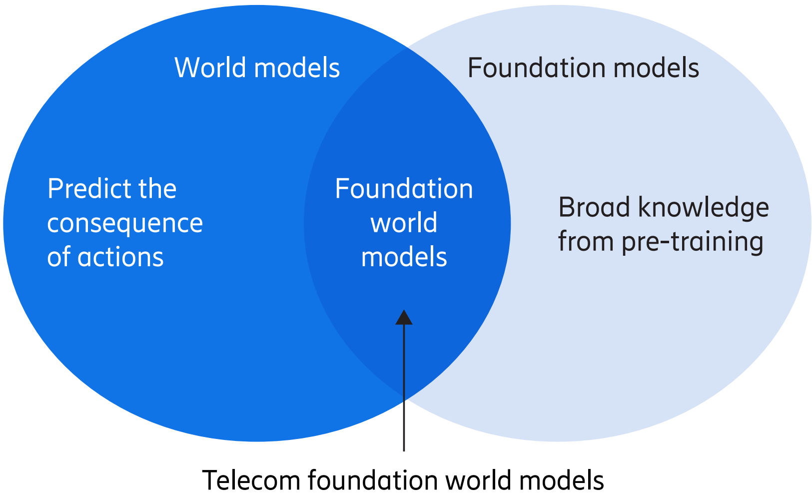

Figure 3. Telecom foundation world models

A telecom foundation world model brings the two together, as depicted in Figure 3: a foundation model that provides a compact representation of the network, and a world model that predicts how the network responds to actions on top of that representation. The result is a generic AI model that can serve many automation use cases, reducing the number of models to maintain and offering a more holistic view of the network.

The telecom foundation world model can be obtained by fine-tuning an existing foundation model on action-rich data, or by training a smaller predictor on top of the foundation model's frozen latent representation. The two approaches can also be combined, so that the representation and dynamics are learned together [9]. In each case, the foundation model provides a compact summary of multimodal network data, and the world model captures the action-conditional dynamics on top of it.

Building a world model on top of a foundation model offers several advantages. The same pre-trained representation can serve many downstream tasks, so a single foundation world model can be adapted to multiple automation use cases, reducing the number of models to maintain and offering a more holistic view of the network. The foundation model also captures broad telecom knowledge that the world model would otherwise have to learn alongside the action-response dynamics. With this representation in place, the world model can focus on learning how the network state changes under actions, making more efficient use of the limited action-response data available.

Figure 4 shows the high-level architectural blueprint for autonomous domains (ADs) [2], where an autonomous domain is a system or set of systems or agents that is capable of autonomous behavior, for example, resolve tasks, and adhere to objectives without manual human intervention [10].

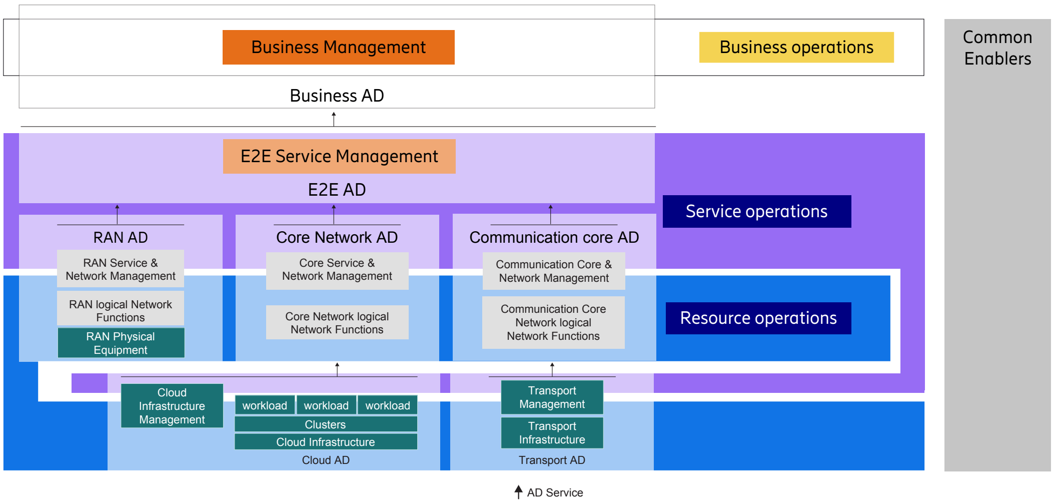

Figure 4 High-level architectural blueprint for Autonomous Domains

In the business and upper service-operations layers, timescales vary depending on the use case, and the focus is on end-to-end planning, customer service management, and proactive fault handling. LLMs fine-tuned on log files, standards specifications, product documentation, and trouble tickets may be well suited here, together with telecom foundation world models, for example for service level agreement (SLA) management.
In the lower layers of service operations and in the management part of the resource layer, for example in RAN service and network management, timescales range from seconds to minutes. This is where foundation world models, fed with different types of data, such as topology, FCAPS data, and other management data, are most relevant. In the lowest part of resource layers, decisions are taken at millisecond level and foundation world models trained on low-level radio data [11] could support tasks such as channel estimation and beam selection. This paper focuses on the business, service, and management parts of the resource-operations domain.

The impact on automation use cases

In this section, we highlight use cases toward higher levels of autonomy and discuss the role of telecom foundation world models in them.

- Intent-based management [3]: Intent allows network operations and management to move from the how to the what. It can be used to formulate what we want to achieve in the network, and the network then figures out how to do so, determining the configurations in an autonomous manner adapted to the current situation in the network. Conflicts may arise: a key part of mitigating them is to predict the effect of actions before they are taken in the network. Using the predicted outcomes, the system can select the action that ensures the highest compliance with the objectives. Intent management functions will be able to take actions in the network to fulfill the intents by checking the effect of different actions using telecom foundation world models, which can be part of an NDT.
- Network digital twins: By keeping an up-to-date mirror image of the live network, or part of the live network, with a given level of detail, NDTs can be used for a wide range of use cases, answering what-if questions. For example, what would a fault's impact be on the services provided by the network? Would it be possible to accept a certain SLA on a new service, and what would the consequences be? Telecom foundation world models can be used under the hood, as the AI engine of NDTs.
- Network planning and deployment: While telecom software is highly customizable, this flexibility comes with a large number of configuration parameters and setting them correctly is a time-consuming and tedious task. Utilizing telecom foundation world models, the configuration process can ensure that the right configurations are selected to meet specific customer needs. Another example is network planning, where cloud-native RAN deployments allow a high degree of flexibility. Currently, it is difficult to find the best plan based on various performance and availability requirements, considering different hardware and software options while minimizing cost. For this use case, telecom foundation world models can be used to evaluate automatically generated valid options.

- Network optimization: rAPPs can be employed to optimize network performance, for example by controlling uplink and downlink power or antenna tilts. Given multiple objectives to achieve for example, capacity versus coverage versus energy, different optimizer rAPPs can be invoked to address configuration changes in the RAN. In this scenario, telecom foundation world models provide a safe training environment for reinforcement learning-based AI agents.
- AI agents in OSS/BSS: In the BSS domain, many manual process flows can be automated with the use of AI agents. One example is automating the customer order process, including SLA negotiation. A telecom foundation world model can support effective decision-making by predicting the impact of agents' decisions on SLA.
- Predictive fault management: If certain faults can be predicted in advance, they can be mitigated before they happen, so their impact is minimized. For example, alarms or measurements of CPU temperatures in the data center may indicate that a server is likely to fail, so it is best to preempt traffic and workload from that server to avoid disruption of service. Telecom foundation world models can help answer questions such as: will a certain element fail in the next minutes or hours, given its current conditions? Would we be able to mitigate the failure with a given set of remedial actions?
- Improved handling of software errors: Software errors and vulnerabilities require fast resolution. After detecting a software error or vulnerability in a live network, a new version of the code must be delivered and validated before deployment. Telecom foundation world models can support this process by predicting the impact of the patch in the context of an operator. This helps ensure that the patch does not crash in the operator's system due to different features, configurations, and context that are unique to the operator.

The road to get there

Building telecom foundation world models is challenging, with difficulties in modeling, data, and operations. The model must capture spatial information about the network topology as well as variations in time, ranging from short-term peaks in network utilization to seasonal patterns. In addition, it should generalize across deployments with different combinations of radio hardware. A central difficulty is to separate the impact of a management operation on the network from variation in KPIs caused by external factors and unforeseen events.

Data quality is equally critical. Training data must be relevant, complete, and traceable while preserving regulatory compliance, data sovereignty, ownership, and contractual boundaries. Privacy is enforced by anonymizing end users' identities, geographical location of sites, cell identifiers, and CSP names. Anonymizing spatial information, however, restricts the model's ability to learn the spatial and topological structure that the modeling task requires. On the operations side, training and inference come with non-negligible infrastructure needs and costs. Care must be taken to design energy-efficient, scalable, and sustainable solutions.

Pre-training a base telecom foundation world model and specializing it for the customer network

In a typical scenario, a vendor builds a telecom foundation world model based on its domain expertise. The vendor uses available data and knowledge, including synthetic data gained from simulations and content from product documentation to pre-train the base model. Thanks to this broad knowledge, the base model may be reused across many deployments and directly applied to common network automation problems.

When customer-specific accuracy is needed, the base model can be specialized into a customer-specific telecom foundation world model enriched with data from the customer network. Specialization captures the unique characteristics of the CSPs’ network and the type of problems they wish to solve, while keeping the data safe: a CSP’s data is not used in any way in other CSPs’ models.

Open research questions and potential avenues

Several open research questions remain on the way to industrial-scale telecom foundation world models. We highlight the main ones below.

Data availability: Data availability for training remains one of the core challenges. Privacy is one concern, but getting access to diverse, high-quality data, and understanding what scale of data is needed to train telecom foundation world models itself is a challenge. Beyond availability, the data must be rich enough to support disentanglement between actions taken on the network and external factors. Without this, the model learns correlations rather than the action-response relationship that network automation depends on.

One model or many: A core question is whether to operate many use case-specific models or a few general ones. Specialized models may perform best on narrow tasks, but operating many of them is costly and complex. Whether a few general models can match that performance while simplifying operation is itself open. More concretely, it remains unclear how many network parameters and KPIs a single model can effectively capture, and whether context is sufficient to achieve high performance across diverse network sites while still being cost-efficient. A practical path is to start with dedicated models where the data is strongly different and to combine domain knowledge with data-driven models to navigate the performance versus model-complexity trade-off.

Lifecycle management: Lifecycle management of telecom foundation world models includes training, deployment, retraining, adaptation to different past and current tasks via fine-tuning or in-context learning, and retirement. Reusing the model across network domains and use cases would help simplify lifecycle management, but since fine-tuning is likely to be needed, managing the fine-tuned versions of the model will be required too. Methods such as in-context learning, which adapt the model output given a specific context without retraining, may further ease the lifecycle management.

The role of synthetic data: Network simulators based on a mix of first principles and data-driven models can be used to generate data for exploring rare or risky situations, or to generate privacy-preserving data. Telecom foundation world models need to be trained on rare data showing changes and their consequences, and synthetic data is a key asset to generating these examples today, but it needs to be combined with field data. Open questions include the optimal strategy for combining real and synthetic data, and how the role of synthetic data will evolve as closed-loop agents become more present in the network and routinely apply diverse changes, generating richer field data over time.

Handling different types of data: Good foundation models are available for text, images, speech, and time series. Telecom data introduces less-explored types of input, such as time series of network KPIs, maps, or topology information which carry irregular and rare events. Reusing pre-trained models when working with these inputs and judging whether adding new types of data justifies the increase in complexity are still open questions. Efficient training and adaptation to this new data in itself is part of the challenge.

Trustworthiness and security: Trustworthy and secure use of telecom foundation world models must be ensured, with proper guardrails in place to maintain the network's integrity. Trustworthiness covers the reliability of the model predictions, including uncertainty estimates. Security covers the protection of the training data and model weights, as well as the integrity of inputs and outputs. To what extent can these aspects be enforced in training or through external guardrails remains an open question.

Conclusion and way forward

Telecom foundation world models are an essential building block on the road to higher levels of network autonomy. Trained on vast amounts of diverse telecom data, they support proactive decision-making and self-adaptation. They will play a fundamental role in areas such as intent-based management, NDTs, optimization of network resources, and predictive fault management. Moving to autonomous networks will come with cost savings and increased and new revenues for CSPs.

Reaching this future will require collaboration between CSPs and vendors. CSPs should support the effort by allowing their historical operational data to be used in the training process of telecom foundation world models. To be successful, telecom foundation world models will require rich data availability, ranging from operational network changes to end-user experience. When special care must be taken to preserve privacy, specialized, customer-specific models can be built.

Telecom foundation world models offer a unique opportunity for CSPs. Now is the time to engage in a dialogue to shape the future of these models so that they address CSPs' pain points and help them move towards higher levels of network autonomy.

References

1. How zero can we touch? - Ericsson
2. Autonomy by design — enabling self-managing networks across the life cycle
3. Autonomous network operations: from reactive management to intent-driven optimization
4. Multi-Modal Data-Enhanced Foundation Models for Prediction and Control in Wireless Networks: A Survey
5. Telecom Foundation Models: Applications, Challenges, and Future Trends
6. Telecom World Models: Unifying Digital Twins, Foundation Models, and Predictive Planning for 6G
7. M. Bouton, I. Mistioni, S. Lindståhl, J. Jeong, Mobile Network Control with a World Model, IEEE NOMS 2026.
8. Amadeu Nascimento Junior, Meltem Civas, Jean Martins, Adriano Mendo, Juan Ramiro, Evolving Prediction Models for Radio Access Networks IEEE ICC 2026.
9. Assran, M., et al. (2025). V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning.
10. TM Forum, Autonomous Networks Technical Architecture (IG1230), December 12, 2022
11. 6G gains momentum with AI and industry alignment - Ericsson

Authors

András Zahemszky is a Master Researcher at Ericsson Research whose work focuses on autonomous networks. His current research interest lies in the intersection of mobile networks and AI. In his current role, he is leading and conducting various research activities in intent-based management. His previous research work included core network concepts and prototyping. He joined Ericsson in 2007 after completing an M.Sc. in Computer Science at Budapest University of Technology and Economics in Hungary.

Maxime Bouton is a Senior Specialist at Ericsson Research on AI for safe and scalable control. His research interests lie in applied reinforcement learning, AI safety, and autonomous and multi-agent systems. Maxime received a Ph.D. degree from Stanford University where he worked at the Stanford Intelligent Systems Lab on safe planning under uncertainty for autonomous systems. Prior to doing research, Maxime obtained his master's degrees in Aeronautics and Astronautics at Stanford University and CentraleSupélec.

Jaeseong Jeong is a Senior Specialist AI for Optimization at Ericsson Research. Since joining Ericsson in 2016, he has been working on advanced AI technologies that enable autonomous, self-optimizing networks. He received a Ph.D. degree from the Korea Advanced Institute of Science and Technology. Prior to joining Ericsson, he was with Automatic Control Department at KTH Royal Institute of Technology in Sweden as a postdoctoral researcher.

Selim Ickin joined Ericsson Research in 2014 and currently serves as a Senior Specialist in Distributed AI. He has authored over 30 publications and holds patents across telecommunications, applications, devices, and AI. His research interests encompass distributed learning, sustainable AI/ML modeling, and AI-native network architectures. He obtained his B.Sc. in Electrical and Electronics Engineering from Bilkent University in Turkey, and his M.Sc. and Ph.D. degrees from Blekinge Institute of Technology in Sweden.

Dr. Melike Erol-Kantarci, Strategic Product Manager, AI in RAN at Ericsson, is an influential technology leader in AI-native RAN product development, driving forward-looking solutions and AI RAN strategy in Ericsson Networks. She holds the prestigious Canada Research Chair in AI-enabled next-generation wireless networks and is a highly cited full professor at the University of Ottawa, Canada. Throughout her more than 15 years in communications, she has received numerous awards and recognitions from technical societies, including IEEE Fellowship for her contributions to AI-enabled wireless networks.

Juan Ramiro, leads AI Product Research within the Cognitive Network Solutions Area in Ericsson's Business Area Cloud and Software Services. He holds a Telecommunications Engineering degree from the University of Malaga (with awards to the best academic record and master thesis), a Ph.D. in Electrical and Electronic Engineering from Aalborg University, and an Executive MBA from San Telmo Business School. He has co-authored more than 30 patent applications and scientific publications, and he is also an Honorary Associate Professor at the University of Malaga and co-author of a book on self-organizing networks.

Vishal Garg is an AI evangelist with over 18 years of experience in the telecom OSS/BSS domain, spanning diverse roles across technology and innovation. His current focus is on agentic AI, generative AI, and their potential to transform telecom systems and operations. He is passionate about applying emerging AI capabilities, including foundation models, to create intuitive, efficient, and practical solutions for complex telecom challenges.