Intelligent operations

How AI plays a critical role in network operations

ericsson.com/
intelligent
automation
platform

Intelligent Automation Guide Series
June 2022

Evolution is required for future networks and use cases

5G networks demand new, highly automated solutions that support a wider range of use cases in an increasingly complex environment.

Ericsson Intelligent RAN Automation is a series of technologies, solutions and services that use artificial intelligence (AI) and machine learning (ML) to automate repetitive tasks, reducing complexity in the Radio Access Network (RAN). It transforms RAN software and service life cycles, enabling intelligent operations to improve network performance and user experience whilst reducing operational costs.

The role of AI is to unlock more advanced network automation, performance and analytics capabilities to make RAN functions more autonomous, replacing manual processes with intelligent tools that augment humans. It also makes both RAN functions and tools more robust for deployment in different environments.

Call for change: A wider scope for automation

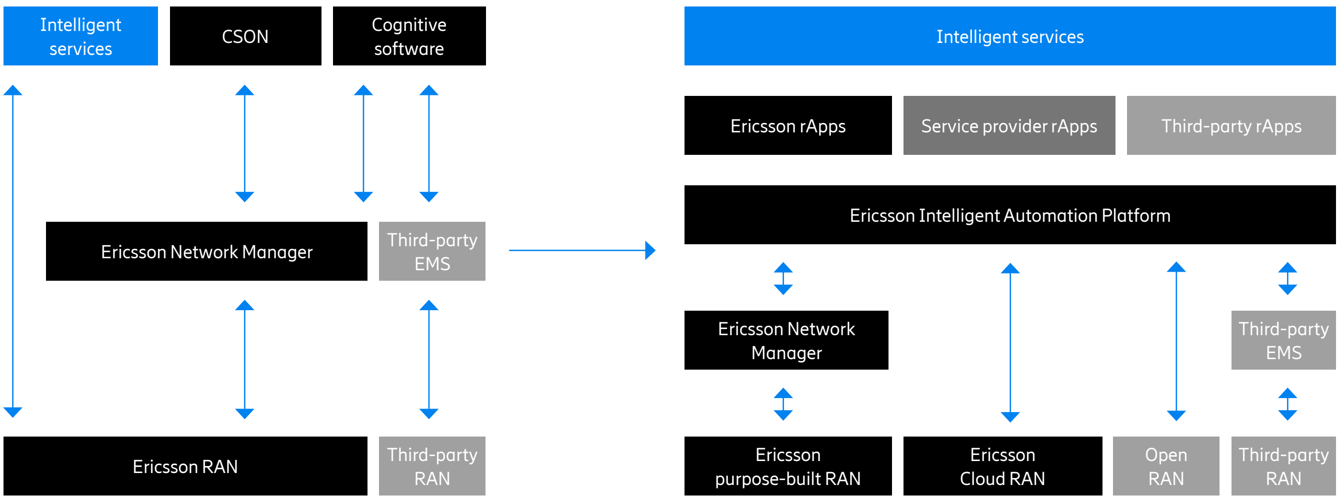
As defined in 3GPP, self-organizing networks (SON) that include self-configuration, self-healing and self-optimization functionalities are not sufficient to meet the new demands. These include:

- higher operational efficiency to cope with increased complexity, such as a mix of virtualization and physical workloads, increased network density, and multi-technology
- shorter time to market for new business opportunities such as network slices, multi-tenancy and dedicated networks
- optimal performance to provide superior user experience
- increased openness and innovation

Ericsson Intelligent RAN Automation provides end-to-end network automation. It includes both centralized and distributed SON solutions with new capabilities that empower innovation and support many use cases with a shorter time to market. These solutions are highly adaptable, supporting both existing and future networks.

The objective of RAN automation is to boost RAN performance and operational efficiency by replacing the manual work of developing, deploying, installing, optimizing and retiring RAN functions with automated processes.

Figure 1: The evolution of current architecture for innovation



Creating a competitive advantage with Intelligent RAN Automation

Ericsson's automation solutions, with AI foundations, provide communication service providers with the platforms and advanced life cycle management (LCM) of RAN software and services required to evolve networks efficiently to successfully meet ever-changing demands. The aim is to deliver improved network performance, accelerate time to market for new capabilities, target the right investment for improved return-on-investment (ROI), and enhance operational efficiency.

AI algorithms applied to control loops with different response times

Ericsson Intelligent RAN Automation solutions provide the right automation where it makes sense and is the best value for money. The task of efficiently operating the RAN to best utilize the deployed resources can be divided into different control loops, acting together according to different time scales and scopes.

Intelligent RAN Automation solutions utilize AI/ML algorithms that interact and integrate with engineered algorithms and existing processes, where applicable, in all these control loops. The two fastest control loops are related to traditional Radio Resource Management (RRM). Examples include link adaption implemented as RAN function in the fastest control loop and cell supervision in the second fastest.

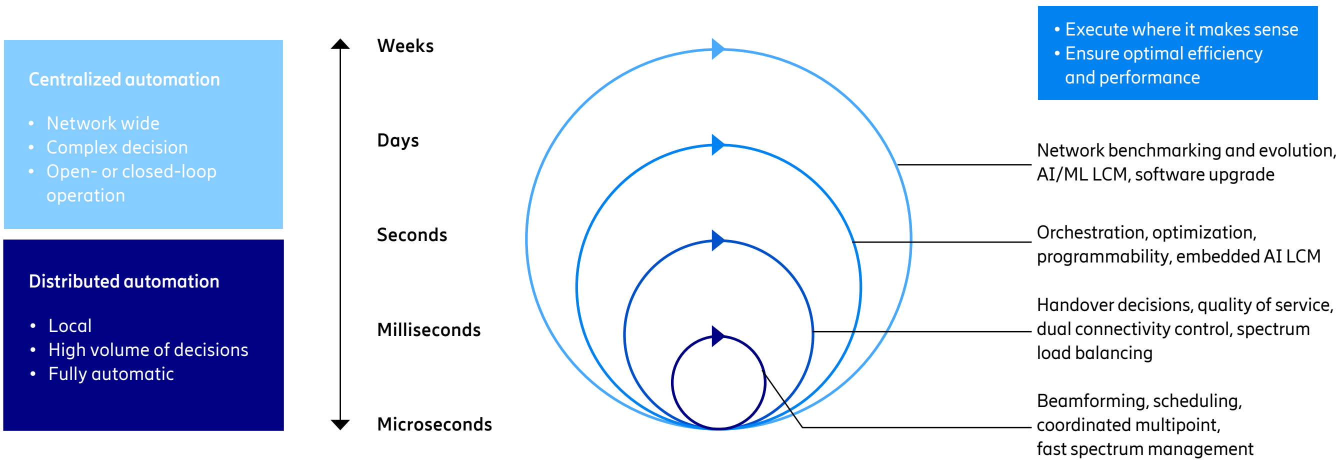
Functionality in these control loops is mostly autonomous and is driven by engineered algorithms which require complex configurations in a given timeframe, ranging from mere milliseconds up to several hundred milliseconds.

In many cases, AI and ML make it possible to enhance the functionality in fast control loops to make them more adaptive and robust for deployment in different environments. This, in turn, minimizes the amount of configuration optimization required in slow control loops.

The slower control loops relate to traditional network design, optimization and management use cases which are implemented in non-real time. This functionality resides in the rApps within the new architecture. In contrast to the two fast control loops, these slower loops are largely manual. The slow control loops encompass the bulk of the manual work that will disappear as a result of RAN automation, explaining why AI and ML are especially attractive in these loops.

In the short term, AI/ML-powered solutions based on self-trained algorithms will be more accurate and efficient for certain use cases than the traditional, rule-based, human-made solutions. In the long term, AI/ML-powered solutions will be ubiquitous in that RAN and AI/ML will be just another set of tools used to achieve the best-performing and most cost-effective network.

Figure 2: Holistic take on RAN automation



Evolved life cycle management alternatives for service providers

Compared to traditional software, AI/ML technology introduces the elements of training, model drift concept, federated learning, and a stronger need for access to high volumes of data. The LCM processes set the roles of suppliers, integrators, and service providers – in essence, who is responsible for what, and who sells what to whom.

As an industry, we must adjust the traditional software LCM to include AI/ML-based software LCM. A very high-level AI/ML LCM process will have different alternatives depending on the vendor and service provider's responsibility.

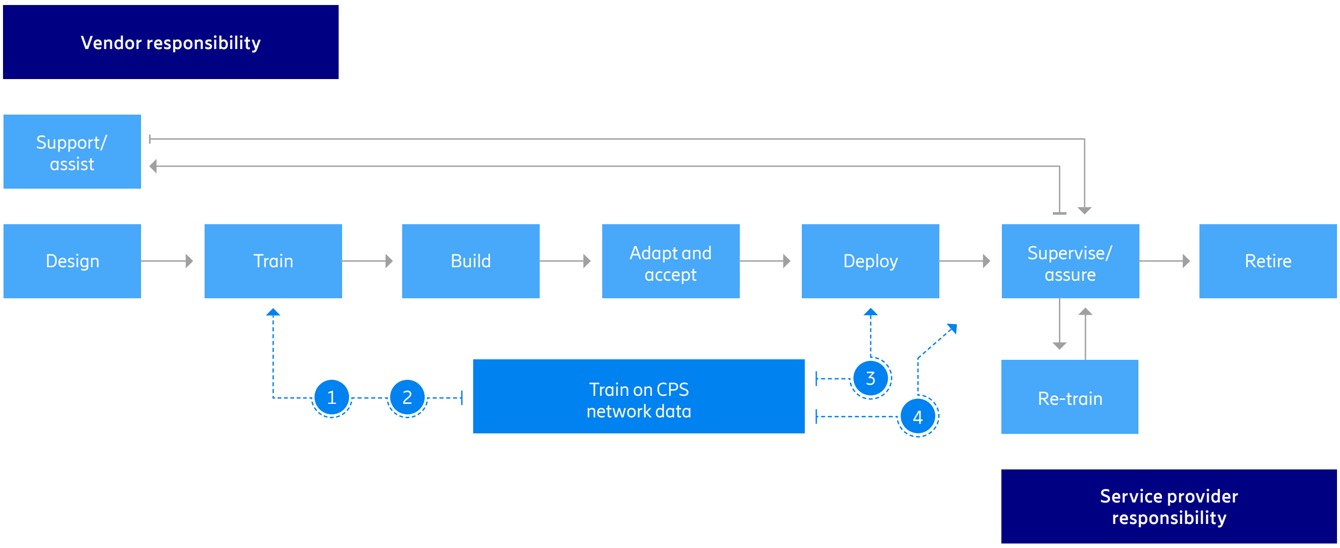
The four main LCM alternatives are:

1. The trained global model – training is the responsibility of the vendor, for example, the rApp is provided as an application by the vendor.
2. The local model – local training is the responsibility of the vendor, though the developer might deliver an initially trained model with vendor data.
3. The initially trained global model with retraining capability – training or re-training is the responsibility of the service provider.
4. Embedded model with automatic re-training – re-training on local data is done autonomously in the service provider's network.

AI/ML software provides flexibility for the right level of global versus local adaptation.

These four models show that AI/ML software lends itself much better to choosing the right level of global versus local adaptation, in comparison with traditional rule-based solutions. AI/ML software provides different levels of training depending on the globalization of the model. An example would be a model that is trained once in the vendor environment before being deployed into multiple networks and situations. It would have the ability to adapt correctly to new, previously unseen data, as well as perform local adaptations and train or re-train the AI models with unique local data.

Figure 3: High-level LCM for AI/ML software



How Ericsson leverages data-driven AI algorithms

Data collection and management are key challenges to scaling the AI/ML software and tools for service providers.

For the AI/ML-based algorithms, both public and non-public data are needed. Public data (for example, performance monitoring) is standardized and the exposed data available from a product or service is supplied by a vendor to service providers for the purpose of product operations and/or service delivery.

Non-public data (for example, AI model debug trace), on the other hand, is data containing sensitive information relating to intelligent property rights (IPR) and is used by the vendor for innovation and/or service development, verification, and deployment. Since the generated non-public data is typically 100,000 times larger in volume than the public data, Ericsson has developed mechanisms to bring data that is needed solely for the relevant use cases from specific network elements. This makes data management easier, simplifies operations and enables the AI/ML application to work at scale.

Once the data has been acquired, a virtual environment — consisting of a digital twin and associated simulators and emulators — is often used in the development of AI/ML-based algorithms that can use public and non-public field data or simulated data to train the final model.

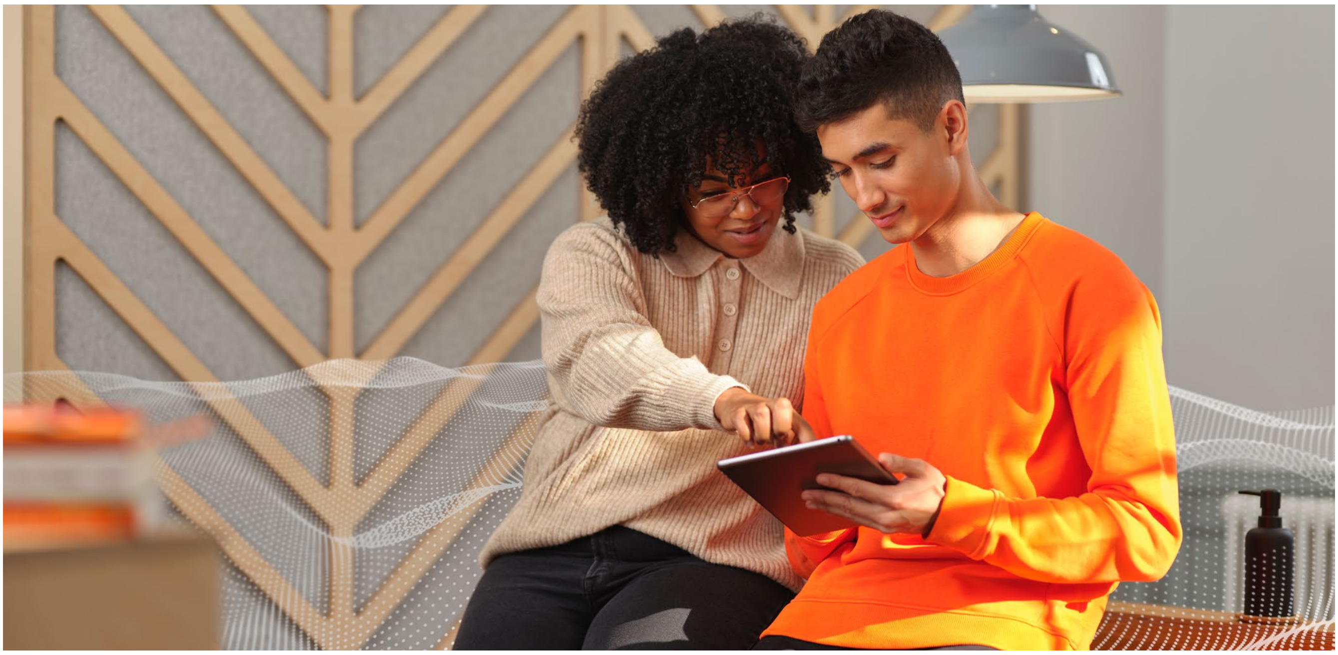
The digital twin is a solution that allows representation of the real-life process or object in the digital domain. The benefit of digital twins is that they can be used to predict and avoid the negative impacts in a live network – for example, during the initial exploration of the advanced AI techniques. All the simulations can be done using digital twins and when performance of the algorithm is good, it is exported to the intelligent agent in the live network.

Once the agent has acquired all the necessary knowledge from the digital twin, the achieved policy can be safely applied to the live network. From that moment onwards, the agent will decide the optimal actions on the live network, while continuing to learn from its feedback and allowing a configurable degree of controlled exploration.

Training and re-training the AI/ML algorithms

We at Ericsson make a distinction between the initial training of AI/ML algorithms, defined as creating and training an ML algorithm in the design phase, or training and re-training in the maintenance phase. Once the AI/ML features are identified for the initial training of the AI/ML model, we know what data is needed for re-training when the model starts to drift, which might impact the efficiency of network function or a process. The re-training can either be done offline in data-driven development at Ericsson or within the service provider's network. In the latter case, the re-training is done with customer-unique data and often with the purpose of adapting to local environments that are hard to generalize with the data available offline.

The AI/ML algorithm may be improved over time, or complemented by the addition of other algorithms to make the predictions more accurate, or by re-training the model with local data in the network where it operates. In the long term, this iterative development may result in centralization of certain AI/ML resources as the system architecture and capabilities evolve. Data-driven development is a key element in the LCM evolution of RAN software, together with the training and re-training processes.



Intelligent enablers for real-life use cases

The majority of service providers' business objectives, such as improving network performance, require acting at scale.

Certain network parameters are configured on a per-cell level but may have a strong impact on the performance of surrounding cells, such as antenna electrical tilt and downlink transmission power. A change in any of these parameters also affects users served by the surrounding cells. Finding the optimal configuration for these types of parameters is a complex exercise for humans, so we need the support of AI techniques to achieve a network-wide coordinated strategy. It is needed to ensure implicit coordination and the correct operations strategy in which intelligent agents will aim to improve not only each cell individually, but the network as a whole.

Intelligent agents capable of handling complex processes are needed to optimize trade-offs between long- and short-term benefits. These operational strategies with the intelligent operations approach have been implemented for several service providers with, for example, antenna tilt optimization and downlink transmission power optimization.

Deep reinforcement learning (RL) is the specific area of ML suited to the challenge of autonomous learning without human expertise.

RL delivers long-term rewards in dynamic environments

RL techniques take inspiration from human psychology to learn. They have been applied in many different fields, from video games, to chess and self-driving cars.

The RL agent accumulates knowledge about the dynamics of the environment — in our case, the mobile network — through different interactions that may result in positive or negative outcomes depending on how the goal is defined.

To train the system, a software agent interacts with the environment by repeatedly observing its state and then — based on the knowledge available to the agent at each stage — taking actions that are meant to maximize a long-term reward, that is, the improved situation based on defined criteria. In each iteration, the agent will learn from the outcome of the suggested actions and will become increasingly "wiser". At the beginning of the process, exploration of the environment will naturally be highly erratic, and then gradually become more focused and precise as the iterations proceed and the knowledge about the environment's dynamics is improved.

At the end of the training phase, the agent should contain enough knowledge to facilitate a decision for each possible state of the environment. Later, when applying the agent to a specific network, the RL system will continue learning and a configurable degree of exploration can be carried out at the same time.

In mobile network optimization, most existing solutions are based on rules defined by highly skilled network operations experts who need to translate that knowledge into the proper automation frameworks. These rules are typically static and universal for all networks. The complexity of 5G makes it very challenging to manually devise rule modifications that benefit a specific network case. In this complex and dynamic environment, an RL agent creates the biggest value, since it can be pre-trained with general knowledge, and then continues to learn in production, allowing an optimal policy for each specific scenario.

Power Optimization use case

The best way to understand the application of RL is with the Power Optimization solution. This use case can be developed as an rApp for centralized automation.

Downlink (DL) Power Optimization is a network optimization solution that uses deep RL technology to identify if cell transmit power can be reduced without compromising coverage or performance. At the same time, the solution identifies cells where power increase is required for performance improvement. Power Optimization saves energy and allows radio capacity to be maximized in markets with strict radio frequency emissions regulations. Continuous closed-loop optimization automatically maintains the optimum settings as the network evolves and traffic distributions change. This results in DL power reduction on the coverage layer while maintaining traffic volume and improving DL and uplink (UL) performance.

The application of AI techniques has been achieved with great results. For a European service provider, Ericsson Power Optimizers helped reduce overall transmitted power by 20 percent, with 3.4 percent saved on the electricity bill per base station. Enhanced antenna tilt improved with DL throughput by 5.5 percent and UL throughput by 30 percent.

Conclusion

The expanding scope of 5G applications is increasing, not only in the demand for improved availability, reliability, latency and security, but also for creating a wider range of network scenarios to satisfy the different applications' requirements.

To cope with this increasing complexity, we need to unlock further advanced network automation and make RAN network functions more autonomous, replacing manual processes with intelligent tools that augment humans.

In many scenarios, to achieve results at scale, we need to act simultaneously across several network elements, and this is simply not possible without this "intelligent" automation to ensure a "network-wide coordinated strategy".

To meet the business goals of service providers, new intelligent operations need to have an extremely elevated level of automation, learning autonomously from the live network, with high accuracy and working on a large scale. Supported by AI foundations, Ericsson Intelligent RAN Automation gives service providers the platforms, and evolved LCM of RAN software and services to evolve networks efficiently to successfully meet ever-changing demands. Additionally, new Intelligent RAN Automation use cases will enable service providers to create business value in terms of improved performance, higher efficiency, enhanced customer experience and ultimately creating new revenue streams.

Technology leadership in AI/ML principles (such as RL and digital twins), combined with network data expertise, are applied to RAN automation to form the foundation of the Ericsson Intelligent RAN Automation ecosystem.

Ericsson is working with multiple service providers on implementing AI-based software and services with great results (60 percent reduction in network performance issues, 40 percent reduction on bad quality cells, and 30 percent improvement on service availability). We expect these results from the field will accelerate the adoption of AI/ML in operations. The future of networks evolution has started with innovative uses cases that will enrich the intelligent operations ecosystem (especially in the energy efficiency domain) and pave the way towards zero-touch operations.

About Ericsson

Ericsson enables communications service providers to capture the full value of connectivity. The company's portfolio spans Networks, Digital Services, Managed Services, and Emerging Business and is designed to help our customers go digital, increase efficiency and find new revenue streams. Ericsson's investments in innovation have delivered the benefits of telephony and mobile broadband to billions of people around the world. The Ericsson stock is listed on Nasdaq Stockholm and on Nasdaq New York.
www.ericsson.com

Ericsson
SE-164 80 Stockholm, Sweden
Telephone +46 10 719 0000
www.ericsson.com

0341-8/22112-FGM1010 Uen