engineering.com

White paper: AI in the Telecom Industry

IEEE
Advancing Technology for Humanity

This whitepaper is sponsored by IEEE.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This white paper discusses the telecommunication industry's growing use of AI, including the ways in which:

- AI-based network automation improves efficiency, reliability and customer service, and opens up new customer markets.
- Data integrity techniques like explainability and automated security help ensure AI systems are trustworthy.
- Refined AI modeling through digital twins, reinforcement learning and machine reasoning leads to more capable AI systems.
- AI optimization reduces costs through efficient processes, optimal designs and improved energy efficiency.
- AI is addressing challenges such as integrating AI with legacy systems, building hybrid networks and handling varied frequency bands.

The white paper also introduces the future of personal connectivity devices, such as a wearable digital assistant, and examines the new AI-based chips designed to bring these technologies to fruition. It concludes with a discussion of how AI is enabling a modern phased design approach with continuous feedback loops, reducing costly redesigns and bottlenecks.

BACKGROUND

AI is enabling faster and more powerful chips for the telecom industry. This allows engineers to imagine and design products such as wearable smartphones that take the concept of personal digital assistants to new levels. These and other IoT devices will be able to connect to the cloud and each other at high speeds with little latency thanks to the AI-based ICs that control and use 5G technology. Additionally, system designers are using AI to build more useful, efficient, and robust communication networks using a combination of the new chips and improved AI techniques.

The Cellular Telecommunications Industry Association (CTIA) estimates that American wireless companies are investing $275 billion into building 5G networks, creating 4.5 million new jobs and adding $1.5 trillion to the economy. Compared to its predecessor, a 5G network is 100 times faster, allowing a full movie to be downloaded in 15 seconds, as opposed to 6 minutes over a 4G network. The number of devices that can connect to 5G could exceed that of 4G devices 100-fold, allowing connectivity between vehicles (manual and autonomous), smartwatches, baby monitors, drones, virtual reality headsets, and more. 5G's low latency will provide the quick reaction time needed to allow continual patient monitoring, remote surgery and other medical breakthroughs, potentially lowering the cost of healthcare by $305 billion per year.

Smart cities will use 5G to monitor infrastructure including buildings, roads, bridges, streetlights and traffic signals, which improves safety, reduces street congestion and decreases fuel consumption, resulting in $160 billion in cost savings and benefits. Virtual reality (VR) and augmented reality (AR) will offer immersive and cost-effective educational experiences for students. Finally, 5G will improve monitoring of the power grid, increasing energy efficiency, improving reliability, reducing costs, and adding $1.8 trillion in revenue to the U.S. economy.

This white paper explores the expanding role of AI in the telecom industry—its applications, benefits and challenges—and describes how the industry is taking advantage of the technology. Topics will include:

- Opportunities for telecoms to create new products and markets, enhance user experiences, improve performance and reliability, and decrease expenditures.
- The challenges of maintaining data quality and security and incorporating AI with existing technology.
- The ways in which engineers are using AI to address the challenges of creating new products and services.

OPPORTUNITIES

The telecom industry is adopting AI, 5G, IoT and edge computing in various capacities, driven in part by three key factors identified by telecom provider Ericsson: network automation, data integrity and refined AI modeling. Network automation streamlines IT processes such as customer service and employee onboarding, as well as network design and dynamic reconfiguration based on desired outcomes and real-time conditions. Automation reduces costs, improves network performance and decreases down time. Data integrity assures that the AI's output is reliable and secured against external threats through a series of checks and balances with human oversight. Real-time feedback, reinforcement learning, and machine reasoning refine AI models so they focus on the overall intent of the system at a holistic level, allowing the designer to specify a goal while AI handles the details. Finally, AI decreases expenditures and opens the market to new products and new customers, delivering many economic benefits for telecoms.

NETWORK AUTOMATION

Ericsson, as well as industry monitor Enterprise Viewpoint, note that improving network orchestration — the act of setting up user accounts, specifying devices and software, and maintaining applications — is essential to delivering end-to-end services to operators, employees and customers. Automating network operations reduces tedious and redundant administrative work, enhances user experiences, builds dynamic and self-correcting networks, and helps to design more robust systems.

This allows operators to provide more user-friendly customer services, deliver better performance, and decrease capital and operating expenditures at the same time. AI will transform the current "chatbot" self-service model into a more intuitive and responsive "virtual kiosk," allowing telecoms to identify customer needs, develop personalized services and open up new business opportunities.

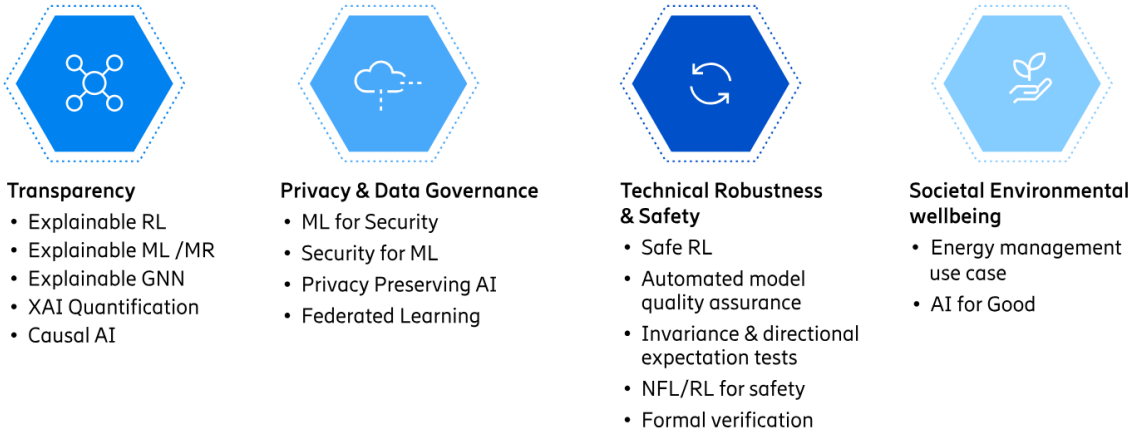
Network automation. (Image: Ericsson.)

System administrators are implementing AI-based zero-touch operations (ZTO) to automate IT processes, which makes network operations more efficient, accurate and user-friendly. Time-consuming processes that reduce productivity while users wait for various services to be completed manually can now be automated, decreasing employee downtime. For example, AI can streamline the user experience with processes such as employee onboarding, IT requests for equipment and software, anti-malware updates and more. ZTO also improves network security by ensuring and maintaining compliance with security and other regulations.

AI allows the end users—system administrators, employees and customers—to interact with a product or network at a more abstract level, alleviating the need to get bogged down in details. Rather than wading through countless menus of services, a user can simply tell the AI the intent. For example, employee onboarding involves setting up user credentials, ordering equipment, setting up custom software and network permissions, and maintaining application and anti-malware software. This is a time-consuming administrative process that is prone to human error. ZTO automates these functions, freeing network administrators to focus on high-level priorities and allowing employees to be more productive.

Network automation through AI helps to improve network performance by analyzing network traffic patterns and optimizing resource allocation on-the-fly, thus improving throughput and reliability. Dynamic networks in high service demand areas optimize data routing to reduce congestion, which maximizes available bandwidth and decreases latency. Highly trained AI models use a combination of historical data and real-time conditions to model, analyze and predict network traffic patterns, helping operators to design, update and dynamically reconfigure networks. AI helps in troubleshooting, reducing downtime. In many cases, it can anticipate pending faults and compensate before they actually cause a failure. Eventually, AI will lead to self-healing networks.

When Spanish telecom MásMóvil wanted to decrease an area's congestion during peak usage hours, the company conducted a pilot study of a high-use zone in order to find effective upgrades that could easily be implemented. An AI, initially trained on a digital twin followed by an iterative optimization process, predicted that during congested periods, using remote electrical tilt (RET) to adjust certain antenna angles would lead to optimized performance. The AI recommended certain nodes to upgrade based on historical usage data and other factors. When those nodes were upgraded, download speeds improved by 12%.



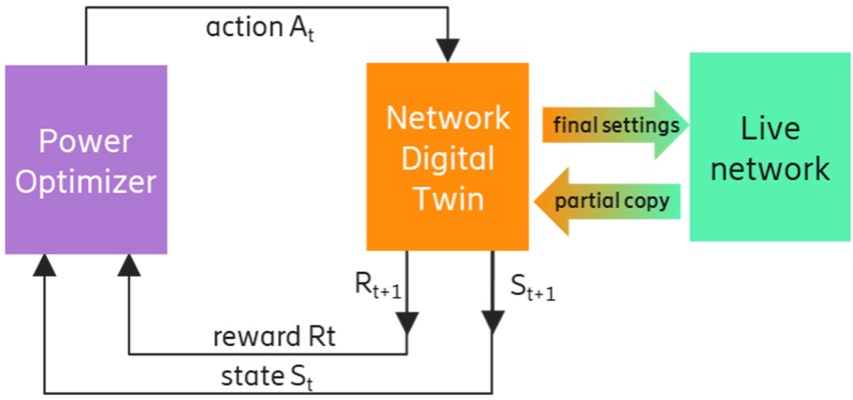
Trustworthy AI.
(Image: Ericsson.)

DATA INTEGRITY

Data integrity begins with trustworthy data, so the telecom industry uses techniques such as explainability (following the AI's decision path that led to a certain conclusion), automated security and built-in safety mechanisms. Some companies are adopting the upcoming European Union AI Act, which follows the "Ethics Guidelines for Trustworthy AI" principles written by the European Commission High-Level Expert Group. These guidelines can reduce security and compliance issues by detecting and preventing security threats such as cyberattacks and fraud, which, in turn, protects customer data.

REFINED AI MODELING

AI agents are constantly refined through digital twins, reinforcement learning and machine reasoning techniques. A digital twin provides a two-way interface between a computer model and a product in the field, allowing the model to help design the product and then using the product's performance to revise the model. Reinforcement learning "rewards" the AI for positive outcomes, leading to gradual improvements in its decision-making. Machine reasoning applies human-like common sense to complex problems to determine the overall intent of the solution. In effect, reinforcement learning emphasizes the "how" while machine reasoning focuses on the "what" and "why."



Refined AI modeling.
(Image: Ericsson.)

An intent-based network allows a telecom operator to define a desired outcome, such as "improve network performance." Then, using machine reasoning, it translates that directive into a series of real-time actions, such as dynamically allocating resources based on existing conditions, in order to meet the goal. Closed-loop control provides immediate feedback, allowing the AI to automatically make real-time forecasts about whether the overall intent is likely to be met, and revise the actions on-the-fly.

DECREASED COSTS

In addition to the aforementioned enhancements in productivity, AI is also improving energy efficiency, helping telecoms save money and boost their sustainability status. Enterprise Viewpoint sees AI's ability to predict and control energy consumption as the easiest way to reduce a telecom's CapEx and OpEx, citing Google's use of AI to optimize its data centers' cooling systems, which reduced energy consumption and cost by 40%.

Ericsson estimates that the energy used to power a network represents between 10% and 30% of its network operating expenses, costing up to $25 billion annually. The company is using AI to improve energy efficiency at all levels, from the use of renewable energy to the system design. A digital twin uses actual network data to teach the AI, which then tweaks the model according to various parameters, including energy usage. These adjustments are made on the model, refined by reinforcement learning, and optimized until they're ready to go live. Coupled with smart grid technology, AI can perform intelligent load-shifting to choose the least expensive power source based on time-of-day pricing. Telecoms who invest in on-site renewable energy generation can use weather forecasts, real-time pricing and grid status in order to maximize the use of their own energy sources.

Swisscom conducted a case study of their effort to decrease the downlink transmission power of a cluster of cells in the Ticino area of Switzerland. AI predicted that power optimization combined with remote electrical tilt of select antennas would significantly reduce the energy consumption of certain network nodes, and then identified the optimal places to install the technology. Test results showed not only a 20% decrease in transmission power, but also a simultaneous 5.5% improvement in network throughput.

SMARTER PRODUCTS

In addition to improving designs, processes and infrastructure, AI is opening markets for new consumer products and taking existing products to the next level of functionality and usability. What we now think of as "smart" devices are often cumbersome to set up and use, even for experienced technology enthusiasts. Average consumers are left unaware of, or unable to use, many product features, hampered by the inability to communicate a desired outcome to a device in its own terms. AI's natural language processing ability, coupled with its use of machine reasoning, translates human intent into a series of actions that accomplish the tasks automatically, allowing users to get the most from the product.

As devices become smarter, more people will adopt them. Early personal digital assistants (PDAs) were little more than electronic day-planners. Today's smartphones incorporate basic PDA functions into a handful of apps, but neither device lives up to the true level of an assistant. Technology newcomer Humane wants to change that by simplifying and enhancing the smartphone experience.

AI Pin.
(Image: Humane.)

The company introduced the AI Pin, a high-tech wearable smartphone replacement that responds to voice commands, much like Siri, Alexa or another natural language AI assistant. Cosmos, the AI Pin's OS and user interface, features a virtual assistant that's similar to ChatGPT. Cosmos automatically directs questions and requests to the appropriate tools, eliminating the need to download and manage apps manually. This turns the device into a true assistant, allowing the user to specify what needs to be done and let the AI Pin handle the details, just as a manager delegates a task to an administrative assistant.

Its laser ink projector, which operates at a distance of 20 to 40 cm, turns a hand into a virtual touch screen. The badge interprets finger gestures such as touch, tap, swipe and double tap, as well as voice commands. Its built-in Personic Speaker uses a Head Related Transfer Function (HRTF) to create a personally optimized bubble of sound at a fixed distance, regardless of the audio intensity. The Pin weighs 55 grams and features a 13MP ultrawide camera, as well as light and depth detectors.

AI Pin virtual touchscreen.
(Image: Humane.)

Like any smartphone, users can choose to accept or decline an incoming call. Unlike other smartphones, the AI Pin remembers history and learns which calls are prioritized, so if it's told to only trigger an audible notification if the caller is a "trusted" contact, it automatically knows who the contacts are. (Trusted contacts can be edited manually through voice commands, as well.) Silent notifications are provided for non-prioritized contacts. The phrase "Catch me up" will cause the AI Pin to sort through texts, calls and emails, and give a briefing of what's important.

Unlike Alexa and Siri, which are constantly listening for voice commands, the badge only "listens" when the user taps the screen. The AI Pin's color-coded "Trust Light" informs the wearer and those nearby when the device is in operation and indicates the task that it's performing, such as photo capture, a phone call in progress, an activated microphone and more. Another light indicator, the ultra-private beacon, notifies of incoming messages and essential alerts.

The badge allows messages to be composed sans keyboard; a wearer can simply provide a contact and the gist of a message, and the AI Pin will instantly write a full message in the user's writing style. The user can tell it to revise the text to reflect different tones (for example, friendly, urgent, formal), and elaborate on or condense any part of the message. In a live conversation, its real-time language translator automatically detects which language is being spoken and then translates between the user's native language and the local dialect.

The pin features a Qualcomm Snapdragon chipset, 4GB of RAM, and 32GB of eMMC storage. For additional storage of photos, notes and lists, Humane created the cloud-based ".Center" (dot Center) which is accessible through any web browser. The badge communicates via cell, Wi-Fi, Bluetooth and GPS. It's powered by a built-in lithium-ion battery that can be recharged wirelessly or by USB-C.

At the time of this writing, customers can purchase an AI Pin for $699 and must subscribe to Humane's service for a monthly fee of $24, which includes voice and data coverage through T-Mobile's network.

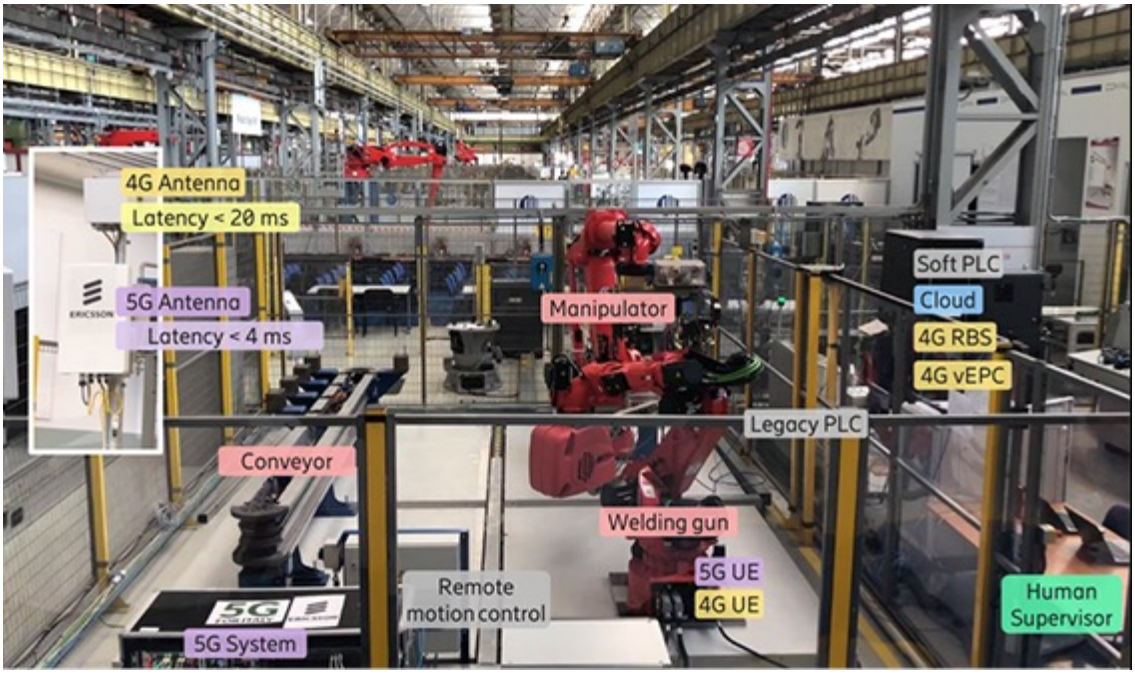
CHALLENGES

In addition to data quality, Ericsson says that AI and connectivity present other challenges, such as integrating AI with legacy systems, building hybrid networks and handling varied frequency bands. But just as more powerful computers help engineers design even *more* powerful computers, AI is helping to mitigate these issues by building more intelligence into new products, allowing them to be seamlessly incorporated into systems where they make the most sense.

LEGACY SYSTEMS

Network upgrades must ensure that the new system maintains backward compatibility with legacy equipment, software and protocols. Legacy software, limited scalability, lower reliability and slow network performance interfere with the desired outcomes to varying degrees.

Upgrading network infrastructure improves performance and reliability across the board, so that's often the starting point for modernizing networks. Designers are using AI to identify network components whose upgrades offer the most impact for the least cost, starting with the low-hanging fruit like individual nodes that can be easily replaced, continuing with larger system improvements, and gradually moving into higher-level designs as the technology becomes less costly and more ubiquitous.



Mixing new and old technology. (Image: Ericsson.)

HYBRID NETWORKS

Hybrid networks are a blend of wired and wireless communication technologies that provide the most complete customer services in an efficient and cost-effective manner. Integrating these technologies, maintaining a variety of equipment, and preventing security breaches are among the numerous challenges of implementing hybrids. Additionally, telecoms must ensure that network signals don't interfere with each other and that each piece of the network can grow to accommodate more traffic as needed.

Advanced planning and fault-tolerant redundancy, as well as modular and integrated testing, lead to a robust and reliable network. Real-time AI-based network analysis tools regulate bandwidth, locate faults, detect potential security threats and automatically reroute data, which improves throughout, reduces latency, and maintains network quality and data integrity.

VARIED FREQUENCY BANDS

Ericsson is employing generative AI to optimize the use of multiple-input/multiple-output (MIMO) channels and estimate available channel occupancy, allowing the network to dynamically redistribute resources to meet current demand. The real-time data and results are then fed back to the AI training to refine its future predictions.

USING TODAY'S AI TO CREATE TOMORROW'S AI

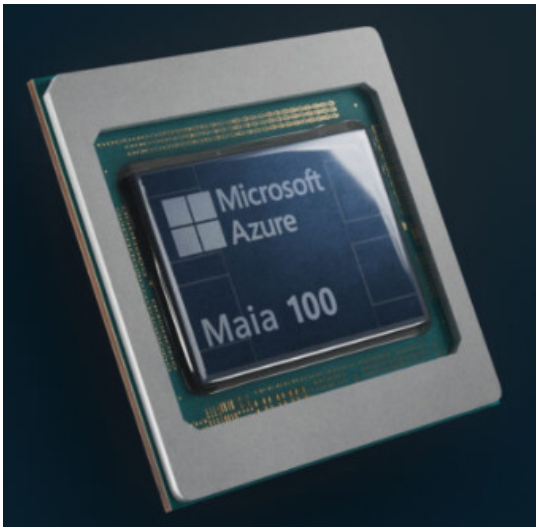
Printed circuit boards were once laid out manually, as were the traces and components in early ICs, with every generation leading to more powerful software that took over much of the manual design. In the same way, AI is being used to create the next generation of AI, at both the hardware and software levels. In addition to component design, generative AI is improving the overall design process, providing more checks and balances throughout all design phases.

AI IN CHIP DESIGN

Semiconductor manufacturer Broadcom recently introduced an on-chip, neural-network unit called NetGNT (Networking General-purpose Neural-network Traffic-analyzer) in its new Trident 5-X12 chip. The standard packet-processing pipeline looks at one packet as it takes a specific path through the chip's ports and buffers. NetGNT is a machine learning inference engine that can be trained to identify different types of traffic patterns across the chip.

For example, NetGNT can look for patterns in which many packet flows all converge on the same port and buffer at roughly the same time, causing congestion. NetGNT — running strictly in hardware to reduce latency — recognizes this pattern in real-time and employs techniques that improve network performance. Trident 5-X12 is a programmable, field-upgradable device that provides 16 Terabits/second of bandwidth and supports 800G ports, which allows direct connection to Broadcom's Tomahawk 5, a device designed for AI/ML data centers.

According to Broadcom, the Trident 5-X12 uses one-fourth the power per 400G port than its Trident 4-X9, and supports the transition to next-generation CPU and GPU servers using 400G NICs. The addition of NetGNT enables new capabilities and improves network efficiency and performance. Its telemetry capabilities allow deeper real-time insights into network operations, which can then be used to train NetGNT. It also maintains legacy API compatibility with current Broadcom switch chips. The Trident 5-X12 is field-programmable using NPL (Network Programming Language), enabling the addition of new features such as telemetry, security and traffic engineering.

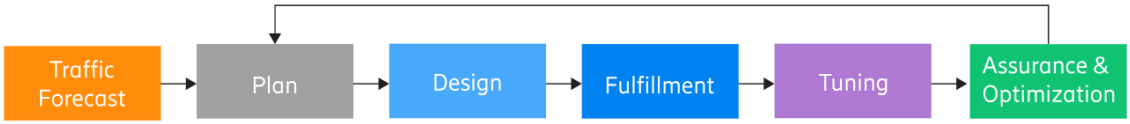


Microsoft Azure Maia.
(Image: Microsoft.)

Meanwhile, Microsoft is introducing its own datacenter processors which are optimized for AI and whose designs have been guided by feedback from its primary AI partner, OpenAI (the creator of ChatGPT). Its new Azure Maia AI Accelerator is optimized for AI tasks, including generative AI. The Azure Cobalt CPU is an Arm processor built for general-purpose cloud-based workloads. The company plans to use them to run its own services, including Microsoft Copilot and Azure OpenAI Service, and then expanding to other applications.

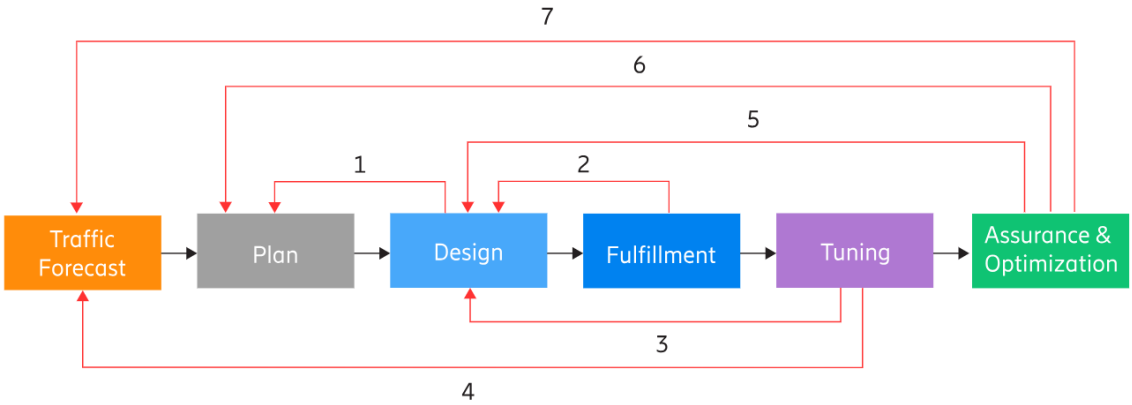
PHASED APPROACH

According to Ericsson, a modern phased approach — a step-by-step process with many feedback loops — to system upgrades can smooth the transition from today's technology to the next generation of telecommunication networks. Unlike a traditional phased design, in which the bulk of the feedback occurs near the end of the process, the modern phased approach emphasizes feedback at each stage.



Traditional phased approach. (Image: Ericsson.)

Besides the relative lack of feedback, the phased approach shown above is rigid and less amenable to changing requirements and technology evolution throughout the product's life cycle. Conversely, the modern phased approach is flexible and resilient, and provides continual feedback that minimizes costly and time-consuming system redesigns. Intent-based AI design agents at every stage coordinate the feedback loops, provide the best available data and predictions based on past experiences and machine cognition, and assure that each phase, as well as the overall system, efficiently meets the desired objectives.



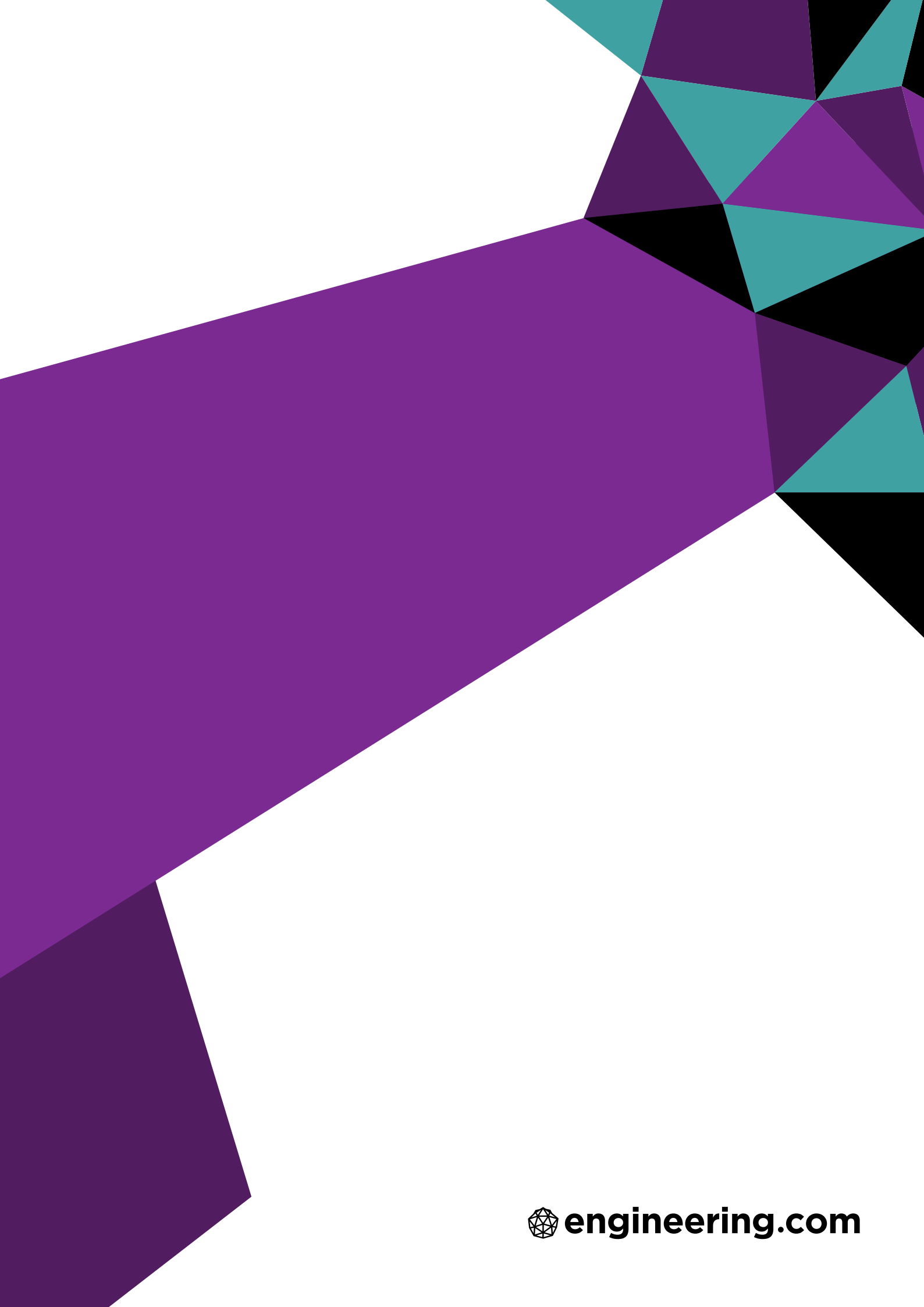
Modern phased approach. (Image: Ericsson.)

LEARN MORE

AI is poised to transform the telecommunication industry by improving services, decreasing costs and opening new consumer markets. The opportunities abound, and expert resources are available to help engineers move into the AI realm. As this technology evolves, it is more important than ever for engineers and researchers to have access to cutting-edge and reliable information that applies to their workflows. IEEE DiscoveryPoint for Communications is an all-in-one platform built for engineers in the telecommunications industry who design and develop technology solutions. Curated by subject matter experts, this platform brings communications-focused information together from hundreds of vetted sources. Quickly find technical information and extract key insights from results with machine learning. With IEEE DiscoveryPoint, spend less time searching and more time discovering solutions to optimize, improve and innovate high-quality products.

IEEE
Advancing Technology for Humanity

To learn more, request a demonstration to see the ways IEEE DiscoveryPoint for Communications can help improve your team's product design workflow.



engineering.com