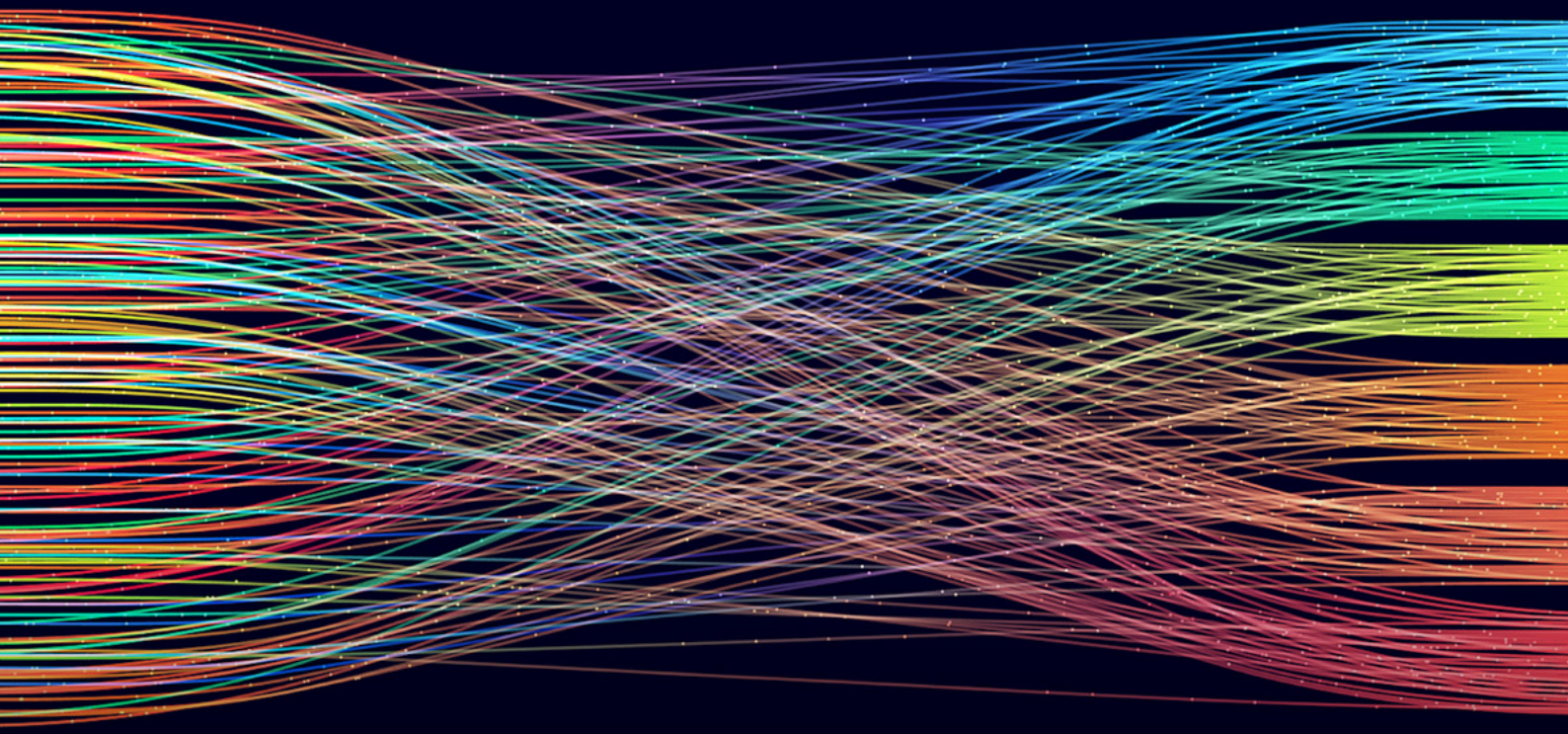

FROM DATA TO DECISIONS: HOW GENERATIVE AI AND DPI ARE SHAPING THE FUTURE OF NETWORK MANAGEMENT

Driving highly-realistic, context-aware GenAI outputs with real-time traffic analytics from R&S®PACE 2 and R&S®vPACE

ROHDE&SCHWARZ

Make ideas real

CONTENT

1. INTRODUCTION

Generative artificial intelligence (GenAI) is transforming everything we do. By leveraging advanced computing technologies to generate original content that emulates human creativity, GenAI is speeding up complex and time-consuming tasks and improving decision-making across various sectors. In network management, the use of GenAI is rapidly redefining how networks are designed, operated, monitored, and secured—with machine-driven intelligence in the form of GenAI being incorporated across key functionalities. This enables vendors and network owners to simplify network management, drive performance, and strengthen the network against security threats[1].

This whitepaper

This whitepaper discusses the rise of GenAI, explaining different types of GenAI models[2] and their applications. By extracting information from millions of structured and unstructured data sources, these models establish critical domain knowledge that would have required thousands of person-hours allocated towards learning, analyzing and memorizing data, indefinitely. As part of the discussion, the whitepaper explores how increasingly popular large language models (LLMs) are paving the way for intelligent conversations in natural languages between humans and machines. Also highlighted is how the combination of GenAI-driven domain knowledge and intelligent conversations enables users to acquire greater insights into network activities and behavior, and invoke intelligent and contextual responses to network events.

To evaluate the full potential of GenAI in network management, this whitepaper explains how GenAI-driven generative outputs are applied across key classes of network functions in various stages—design, operations, and maintenance. From there, it establishes the role of network intelligence as a vital element in a GenAI-based network function, specifically across the following core GenAI processes: model training and fine-tuning, data referencing and model testing. The use of network intelligence in managing and optimizing GenAI workloads, and mitigating GenAI-related threats, is also assessed.

DPI for GenAI

As a primary source of network intelligence, real-time traffic analytics is crucial in powering GenAI-based network functions. This is explored in detail by introducing deep packet inspection (DPI), an advanced technology for detecting and analyzing traffic flows, and how it supports the deployment of GenAI in networks. Highlighting DPI's fine-grained traffic awareness, as well as its performance, flexibility, and efficiency attributes, the solution segment uncovers how DPI-driven network intelligence enhances GenAI models and outputs, optimizes network functions and drives solution monetization. The discussion also focuses on how next-gen DPI, through advanced machine learning (ML) and deep learning (DL)-based techniques, delivers full visibility into networks by identifying every flow, even when encrypted. To further illustrate the importance of DPI, this whitepaper presents two use cases—GenAI-based load balancers and CASBs—and how DPI-driven traffic analytics drives their functionalities and impact on networks.

By leveraging advanced computing technologies to generate original content that emulates human creativity, GenAI is speeding up complex and time-consuming tasks and improving decision-making across various sectors.

1) IBM - How a solid generative AI strategy can improve telecom network operations
2) TechTarget - Generative models: VAEs, GANs, diffusion, transformers, NeRFs

2. GENAI MODELS

Over the past decade, AI implementations in networks have been geared primarily towards creating autonomous, self-learning and self-managing networks. By leveraging ML and DL techniques, AI enables networks to seamlessly identify network events (e.g., a routine device handover or an unusual file transfer) and predict the future (e.g., an impending traffic burst or a denial-of-service attack). AI allows networks to learn and form an implicit understanding of their elements, thus creating built-in intelligence that can be used to drive automated decisions.

GenAI is an extension of traditional AI. GenAI uses neural networks, the foundational architecture for DL, to learn from large volumes of unstructured and structured data. This enables a GenAI model to generate highly-realistic responses which closely reflect the attributes of real data. In recent years, continuous advancements in GenAI have resulted in the introduction of specialized GenAI models, each with the ability to ingest, process and create different forms of data, for example text, imagery, and video.

2.1. GenAI models for text-based outputs

One of the most popular classes of GenAI models in network management is LLMs, which train on text-based data. Extensive training over long periods enables LLMs, which use the DL-based transformer architecture, to acquire knowledge of both natural languages (e.g., English, Spanish) and other text-based languages such as programming code. By observing patterns in text, these models also acquire general knowledge of random topics. Additionally, an LLM can be trained in a specific domain to acquire knowledge of that domain. This can be knowledge of an industry, an organization, a product, a task, or a content format.

Table 1 lists examples of LLMs and the type of knowledge they offer. In network management, LLMs can be used to generate text-based outputs such as network audit reports, incident summaries, performance KPIs, automation workflows, tool configurations, troubleshooting guides, compliance assessments, and risk assessments.

Type of LLMs	Examples
General purpose LLMs	GPT-4[1] (OpenAI), Gemini[2] (Google), Claude[3] (Anthropic), LLaMA[4] (Meta), Grok-3[5] (xAI), Qwen[6] (Alibaba)
Domain-specific LLMs	Codex, CodeBERT, DevBERT, PolyCoder, DialoGPT (IT) FinBERT, FinanceGPT, BloombergGPT (Finance) LegalGPT, ChatLAW, LegalBERT (Legal) ChemBERTa, MedPaLM, PharmacyGPT (Healthcare/Chemistry)

Table 1: LLM-based GenAI models for text-based outputs

> Extensive training over long periods enable LLMs, which use the DL-based transformer architecture, to acquire knowledge of both natural languages and other text-based languages.

1) GPT-4
2) Gemini
3) Claude
4) LLaMA
5) Grok-3
6) Qwen

2.2. GenAI models for non-text data

Non-text-based GenAI models use model architectures such as the Generative Adversarial Network (GAN), Variational Autoencoders, Deep Convolutional GAN, StyleGAN, WaveGAN and Latent Diffusion Models (LDM) to train on non-text data such as numerical values, images, videos, audio clips, 3D renderings, maps, and graphs. This enables these models to acquire knowledge of non-text languages (e.g., mathematical language, visual languages, audio/acoustic languages, and 3D/spatial languages), and generate realistic outputs in these formats. Similar to LLMs, these models acquire domain knowledge based on the areas they are trained in. They may also incorporate text-based prompting, resulting in hybrid models. In network management, these models can be used to generate outputs such as network efficiency metrics, signal strength distribution maps, optimized routing path illustrations, fault detection heat maps, capacity planning models, virtual reality-based network walkthroughs, and device failure prediction models. Table 2 summarizes various non-text-based GenAI models.

Type of models	Format	Examples of LLMs	Training input	Prompt input	Output
General purpose models	Pure	**Pix2Pix, CycleGAN, DeepDream, and GauGAN**	Image	Image	Image
		Magenta	Music	Music	Music
	Hybrid	**DALL·E, Stable Diffusion, and MidJourney**	Text + Image	Text	Image
		MuseNet or Riffusion	Text + Music	Text	Music
Domain-specific models	Pure	**AlphaFold (Biomedical)**	Sequence + Structure	Structure	3D structure
		Magenta Studio (music)	MIDI	MIDI	MIDI / audio
	Hybrid	**Synthesia (Synthetic video)**	Text + Video	Text	Video
		GauGAN (AI art/design)	Image	Sketch	Image

Table 2: GenAI models for non-text outputs

Models that combine generative and retrieval capabilities

Apart from pure generative models, networks can benefit from extended GenAI models such as Retrieval-Augmented Generation (RAG) models, which combine generative capabilities with retrieval capabilities to produce outputs that merge synthetic content with real data. For example, a RAG model could generate a set of configurations for a caching engine based on applications currently consuming the most bandwidth.

3. KNOWLEDGE CREATION IN NETWORKS

Knowledge of languages and domains enables networks to undertake various complex tasks, including tasks that were previously driven entirely by human intelligence. For instance, language knowledge enables a network to understand human instructions and respond intelligently through various modalities—such as text (e.g., chatbots), images (e.g., visual assistants displaying product catalogs), audio (e.g., voice assistants), and video (e.g., animated avatars that can talk and move). On the human end, natural language interfaces allow authorized personnel, even those without programming expertise, to effortlessly steer the network—from changing network configurations, running a health check, or modifying user privileges. For expert users, GenAI's ability to generate complex commands in machine language enables them to improve network control and performance.

Similarly, domain knowledge enables machines to understand a subject area and generate highly contextual and meaningful responses to prompts. This can correspond to any domain layer—whether it is the entire network (e.g., overall costs, total users), a specific segment (e.g., edge, cloud), a particular solution (e.g., SWG, routers), a task (e.g., application monitoring, device detection), an entity (e.g., customers, clients), or an asset (e.g., smart meters, laptops). For network administrators, readily available domain knowledge enhances network decisions and related processes, across both manual and automated streams.

Table 3 shows how various text and non-text data sources are used to enrich different types of GenAI models with domain knowledge that is relevant to modern IP networks.

Domain layer	Example	LLMs		Non-text-based GenAI models	
		Learning sources	Acquired knowledge	Learning sources	Acquired knowledge
Industry vertical	Telecom	Industry report from analyst firms	Global digital content usage trends	Industry Growth Charts from Statistics App	Telecom Industry Traffic Growth
Organization	Mobile Operator X	Press releases	Operator X subscriber base	Network planning diagrams	Operator X network topology
Solution / Product	OSS	BrandX OSS system architecture documentation	BrandX OSS solution hosting stack	BrandX OSS tutorial videos	Step-by-step activation of automation module
Task	Bandwidth optimization rules	Application usage reports	Bandwidth-intensive applications in the network	Throughput graphs and latency heat maps	Congestion nodes in the core network
Content format	Weekly summary report	Corporate content style guidelines	Language style for network performance reports	Digital image libraries	Approved brand colors and imagery type

Table 3: GenAI-driven domain knowledge from text and non-text sources

Combining knowledge of languages and domains unleashes GenAI-driven interactive and generative capabilities. It enables inputs in natural languages (prompts / queries) to be translated into realistic outputs (responses) that are intelligent, contextual and easily understood. Diagram 1 summarizes the information pipeline that supports intelligent interactions in different languages and the creation of realistic content across various domain layers.

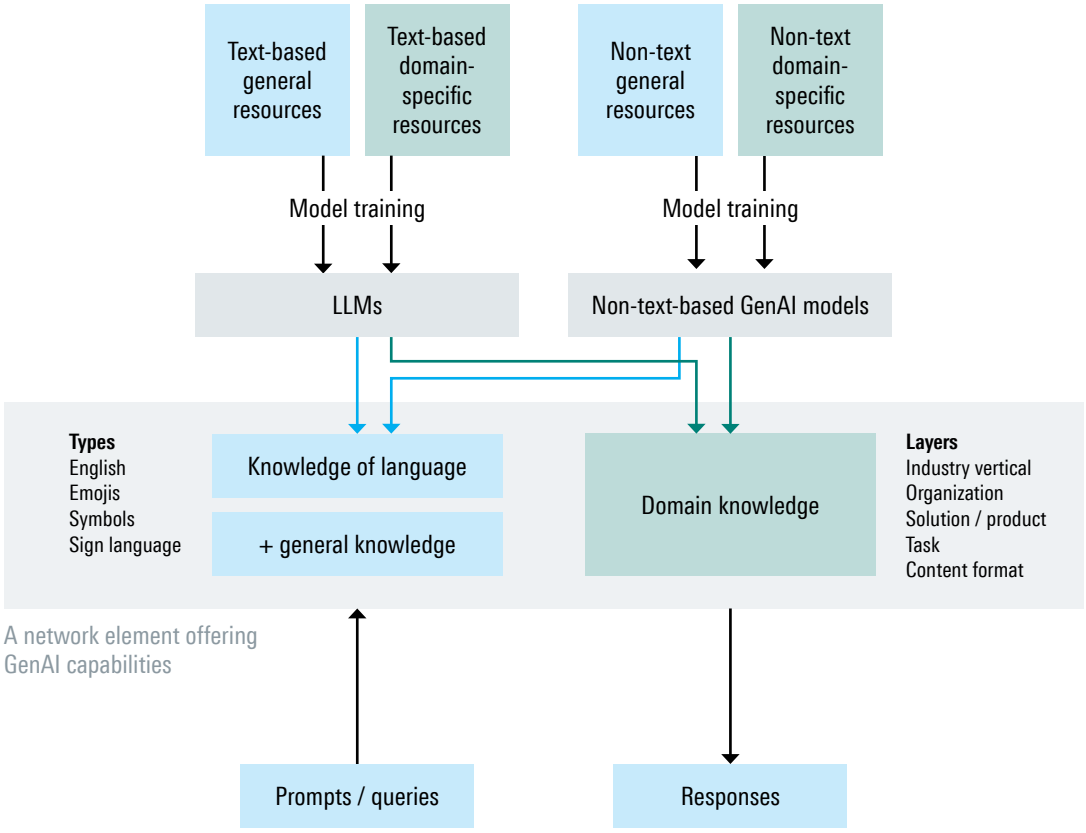

Diagram 1: GenAI information pipeline

GenAI's interactive and generative capabilities can be applied to a diverse set of networking tasks. The following are some examples:

1. **An LLM-based NetOps model** can generate CLI commands within a network management system. For example, given a prompt like "configure border gateway protocol (BGP) for failover with ASN 78112," the model generates the appropriate CLI commands based on the target device.
2. **A StyleGAN-based GenAI model** generates synthetic network map visuals for training classifiers. When paired with a language model, it can respond to prompts like "identify single points of failure in the multi-cloud routing architecture" by analyzing SDN topologies and highlighting risky connections.
3. **A WaveGAN-based GenAI model** generates optimized topologies for mmWave-based Flying Ad hoc Networks (FANETs). Coupled with a language model, it can respond to prompts like "optimize the UAV network for high-throughput and low-latency" by creating efficient mesh topologies for UAVs[1].

1) A WaveGAN Approach for mmWave-Based FANET Topology Optimization

4. TRAFFIC ANALYTICS FOR GENAI-BASED NETWORK FUNCTIONS

Current deployment of GenAI in network management targets network functions. These functions are individual components in the network that are designed to execute specific tasks associated with processing, forwarding, monitoring and securing traffic flows in a given network. According to a recent report on GenAI in network management[1], functional areas that see the highest rate of GenAI adoption are network analytics, service assurance, network monitoring and network automation. These areas correspond to devices or software such as flow analyzers, SLA monitors, IP probes, and automation engines.

GenAI improves a network function's effectiveness (by responding contextually and accurately to network events) and efficiency (by simplifying resource-intensive tasks), leveraging natural language conversations and domain knowledge parsed from a vast number of data sources. For example, GenAI can simulate, in real-time, failure scenarios during a mobile traffic surge in a stadium and generate recommendations such as temporary edge caching and re-routing. The adoption of GenAI improves the quality of network decisions and removes bottlenecks that arise from excessive dependency on human intervention—leading to greater network responsiveness, performance, security, and efficiency. It also enhances organizational outcomes, particularly employee experience, by taking over mundane, time-consuming, and repetitive tasks.

4.1. The critical role of network data

At the core of the policies, processes and decisions executed by a network function is network data. A primary source of network data is **traffic analytics**. By analyzing traffic flows, network administrators are able to gather real-time and historical insights into network processes, functions, devices, infrastructure, users, applications and security threats. Traditionally, specific metrics—for example, bandwidth consumption patterns and application speeds—are used to invoke network actions in real-time, as well as to plan and improve the network over the long term.

With the incorporation of GenAI in network management, the demand for traffic analytics is expected to surge rapidly. From training and fine-tuning GenAI models to testing their effectiveness and ensuring their security, vast amounts of traffic data have now become necessary in powering network functions that are programmed to deliver meaningful and highly-contextual generative outputs. For example, a compression engine that once required basic application classification data now needs terabytes of training data from millions of Internet sessions in order to generate realistic and relevant traffic compression ratios. Beyond supporting GenAI information pipelines, traffic analytics will also play a crucial role in monitoring, assessing and optimizing GenAI workloads, and in managing any disruptions in the networks that can affect the performance and security of GenAI workloads and processes.

This growing importance of traffic analytics in GenAI implementations has started pushing network administrators to re-evaluate existing network intelligence tools and systems, which were designed primarily to cater to traditional workloads. There is, therefore, a growing urgency across today's networks to scale up their analytical capabilities, not only to support continuous cycles of GenAI training and inferencing, but also to enrich existing network monitoring dashboards overseeing GenAI-based network functions and their impact on the network.

4.2. Deep packet inspection

DPI is a traffic detection technology that is widely deployed in IP networks to deliver real-time traffic intelligence. GenAI-based network functions, whose policies and actions are designed and executed based on deep knowledge of the network, benefit tremendously by leveraging DPI's advanced traffic analytics.

ipoque, a leading player in the network analytics space, offers an advanced suite of DPI engines for networking vendors. The suite comprises ipoque's renowned DPI software, R&S®PACE 2 and its vector packet processing (VPP)-native counterpart, R&S®vPACE.

1) Advancing Network Management with Generative AI: The Role of DPI-Driven Traffic Intelligence, Rohde&Schwarz

The engines combine behavioral, statistical and heuristic analyses, and a weekly updated signature library containing thousands of signatures for real-time classification of protocols, applications and services—as illustrated in Diagram 2. This is merged with metadata extraction, which logs various packet attributes such as source and destination IP addresses, port numbers, protocol types, packet sizes, timestamps, time-to-live (TTL), DSCP values, TCP flags, sequence numbers, and fragmentation indicators.

In combination with metadata extraction, traffic classification enables ipoque to establish a broad range of traffic metrics—such as throughput, speed, flow duration, connection stability, device types, encryption rates, inter-packet arrival times, connection counts, port usage distribution, retransmissions, bytes per flow, average payload sizes and round-trip times. Additionally, R&S®PACE 2 and R&S®vPACE can detect traffic flows that are malicious, suspicious, or anomalous, enabling network administrators to identify adverse events such as security attacks, unauthorized access attempts, and resource abuse.

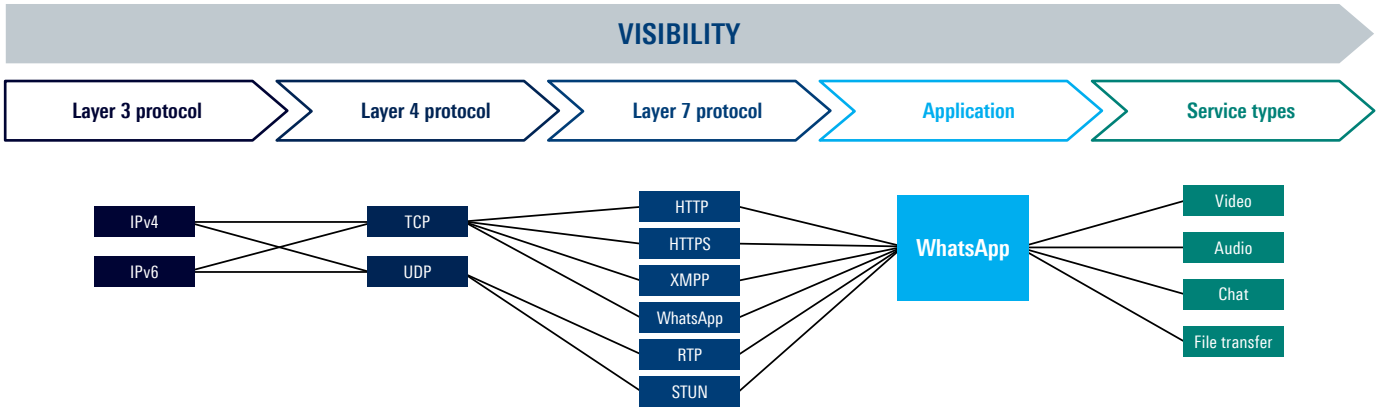

Diagram 2: Classification of traffic by R&S®PACE 2 and R&S®vPACE

4.3. How DPI supports GenAI-based network functions

R&S®PACE 2 and R&S®vPACE offer excellent support for GenAI-based network functions, by ensuring the following:

1. Scalability, given unlimited capacity to process traffic flows
2. Super-fast performance with unparalleled speeds
3. Compatibility across traditional, virtualized, and cloud-native environments, with both DPI engines available for deployment as VNFs or CNFs.
4. Lean form factor featuring a small memory footprint and a customizable minimum build configuration
5. Encrypted traffic intelligence (ETI) which uses a combination of ML and DL techniques, high-dimensional data analysis and advanced caching to deliver visibility into flows that are encrypted, obfuscated and anonymized. This covers latest protocols such as TLS 1.3, ESNI, QUIC and DoX.
6. A high degree of customizability for filtering frequency, traffic coverage and reporting formats
7. Support for new application signatures
8. Seamless integration into IPFIX reporting environments
9. Support for VPP environments, which enables users to cater to network functions with higher computing requirements, e.g., 5G UPFs and CNFs

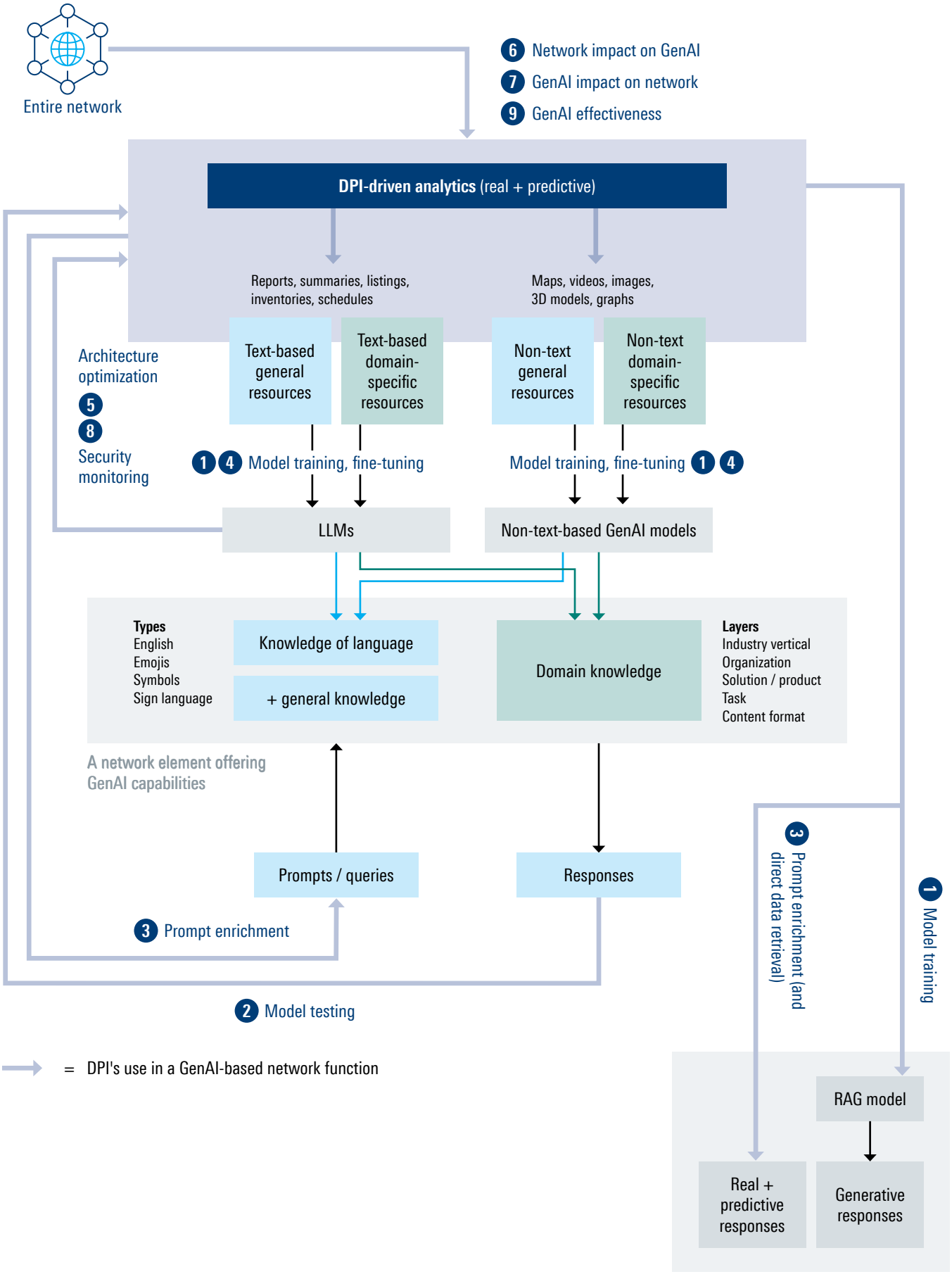

Diagram 3: The flow of DPI-driven traffic analytics in GenAI-based network management

Diagram 3 provides an illustration of how DPI-driven traffic analytics from ipoque are incorporated into GenAI information pipelines and implementation processes in a network function. Diagram 4 summarizes these processes, which are explained in detail below:

1 Model training

- R&S®PACE 2 and R&S®vPACE contribute directly to GenAI model training and fine-tuning by providing high-volume data points spanning various network layers, processes, and components. Data from these DPI engines can be integrated directly with the data lakes used for GenAI model training.
- Both DPI engines also contribute indirectly to GenAI model training as outputs from these engines are channeled into various reporting and analytical systems, and incorporated into hundreds of information sources that are subsequently used for GenAI training. This includes text-based sources such as flow log summaries and error reports, as well as non-text-based sources such as time-series graphs of video traffic, Sankey diagrams illustrating traffic distribution by protocol and spatial heat maps highlighting latency distribution across the network.
- The following summarizes key features of ipoque's DPI suite that ensure the quality of its analytics as a source for GenAI training data:
- Advanced traffic classification leveraging thousands of signatures, for detailed labeling of traffic data
- Advanced metadata extraction, ensuring a vast number of data points
- Unlimited traffic capture for comprehensive analysis of all flows
- Highest detection accuracy rates in the industry with virtually zero false positives, providing analytics that truly reflect the state of the network
- Highly reliable inputs for generating predictive traffic data that is used to train GenAI models used in predictive analysis
- Highly customized analytics that align with different network functions, tasks, and environments. This is supported by the following features:
- Custom signatures: New signatures can be added any time to include specific protocols, applications, and services that are pertinent to a network function
- Flexible data capture: Reporting levels can be adjusted to balance data filtering costs with the depth of information required

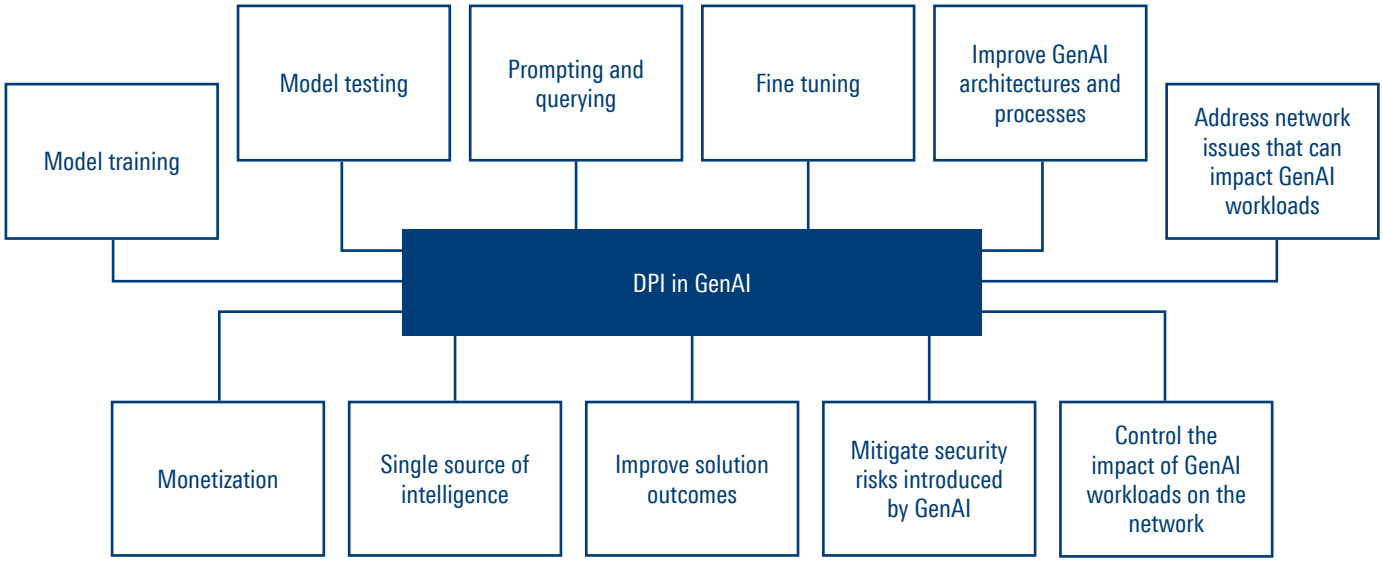

Diagram 4: GenAI processes that leverage DPI-driven traffic analytics

2 Model testing (GenAI effectiveness)

To ascertain whether GenAI outputs are realistic, they must be tested against real data. R&S®PACE 2 and R&S®vPACE provide accurate real-time and historical traffic analytics that enable network administrators to ascertain the quality of their models, while providing insights into potential gaps in GenAI's generative capabilities. For example, simulated traffic flows can be compared to actual traffic flows over a period of time to identify major deviations. Similarly, synthetic projections of network events by a GenAI model can be compared to AI-based predictions of real data.

3 Prompting and querying

Real data is necessary to enrich prompts that are engineered to generate synthetic responses to actual events. For example, an OSS engine in a 5G network requires access to current traffic conditions before generating configurations for a new URLLC network slice. By capturing traffic insights at line speed, ipoque's DPI engines can provide not only periodic analysis of traffic, but also real-time data streams that help networks institute instantaneous responses. This is particularly critical in supporting Agentic AI implementations and automated network actions that draw their instructions from GenAI outputs. The DPI engines also extremely useful for RAG-based GenAI models that use real data references in their responses. An example of this is a synthetic response that discusses phishing attacks as a potential new vulnerability in TLS 1.3, but uses real CVE and patch references.

4 Fine tuning

Fine-tuning of GenAI models allows generic models to be customized to the host solution and the network where they will be deployed. The process thus requires intensive training using the customer's network data. ipoque's DPI suite, by offering a high degree of reporting customizability and seamless integration in any computing environment, enables vendors to capture the exact data that is required to shape domain knowledge that is relevant for a network function and the traffic it processes. Using ipoque's DPI solution accelerates the fine-tuning process as sufficient and reliable data is available in a short time.

5 Improve GenAI architectures and processes

The ability of R&S®PACE 2 and R&S®vPACE to detect and classify traffic by application and service enables network administrators not only to isolate traffic flows related to GenAI, but also break these down by various performance and security attributes. For example, both DPI engines can identify performance metrics for GenAI backend networks by monitoring traffic across components such as inference servers, vector databases, API gateways, service meshes, load balancers, and inter-service links within data centers or cloud environments. The engines can also evaluate GenAI frontend networks involving edge servers, CDN nodes, access gateways, and user devices. This enables network administrators to:

- Monitor the performance of GenAI workloads and corresponding user experiences, for continuous improvement and optimization
- Identify, diagnose and remediate issues relating to GenAI computing stacks, databases and communication links
- Provide real-time diagnostic data for GenAI-based network functions for built-in observability and self-healing capabilities

6 Address network issues that can impact GenAI workloads

Comprehensive, end-to-end DPI-driven visibility helps administrators managing a network function uncover external events—such as cyber-attacks, issues with other network functions, congestion, shortage of network resources, and unauthorized changes in network configurations—that are affecting the performance of GenAI-based network functions. This information helps them manage their SLAs and handle service disruptions more effectively.

7 Control the impact of GenAI workloads on the network

Computing-intensive GenAI workloads can result in resource competition, leading to gradual degradation in network latency and a slowdown in the response times of other applications in the network. By correlating real-time GenAI transactions with performance metrics of other components—applications, connected functions, shared orchestration platforms and shared infrastructure—, DPI helps network administrators improve resource allocation to support the rapid growth of GenAI workloads.

8 Mitigate security risks introduced by GenAI

GenAI workloads can expand a network's attack surface and introduce new security vulnerabilities. This encompasses risks related to models (e.g., model poisoning, adversarial attacks on model inputs, data leakage through model inference), databases (e.g., unauthorized access to sensitive training data, data corruption, database breaches exposing model weights), and users (e.g., phishing attacks targeting API endpoints, credential theft, misuse of AI-generated content). R&S®PACE 2 and R&S®vPACE enable networking vendors to instantaneously detect suspicious, malicious, or anomalous traffic flows associated with GenAI workloads and components, enabling them to avert potential attacks and institute security responses in real-time. The engines also help network administrators improve the security posture of their network functions and address emerging threat vectors more effectively.

9 Improve solution outcomes

Gaps in a network function's policies, configurations, or processes can undermine network outcomes, even when a GenAI model produces accurate outputs. To address this, comprehensive and real-time traffic insights from ipoque can be integrated into solution dashboards. These insights help network administrators identify correlations between the actions performed by a network function (e.g., traffic caching, traffic distribution) and resulting network outcomes (e.g., performance, efficiency, security). By revealing such correlations, DPI-driven traffic analytics enable networking vendors to detect and address weaknesses in their policies, configurations, and operational workflows. Table 4 illustrates how this leads to improved functions.

10. Single source of intelligence

DPI's advanced traffic analysis provides extensive traffic capture (including encrypted traffic) and reporting granularity that align with the data requirements of virtually any network function or process. By using DPI-driven analytics as a unified source of traffic intelligence, networks can achieve greater harmony and consistency in policies and decisions, across both traditional and GenAI-driven functionalities. This approach also eliminates redundancy that stems from duplicated and siloed analytics systems across different network functions.

11. Monetization

DPI-driven analytics can enhance solution capabilities. For instance, traffic analytics from ipoque can be used to create digital twins on which GenAI outputs can be tested. Users can also run their own exploratory analyses to better understand network behavior, before executing GenAI-driven decisions. Apart from these, DPI data can be used in automated prompting, with its metrics acting as a trigger. All these advanced capabilities can be offered as premium features, creating new revenues for solution vendors.

GenAI Outputs	Policy gaps revealed by DPI-driven analytics	Corrective actions
GenAI generates accurate bandwidth allocation policy using data on available bandwidth and users subscription tiers	**DPI**: Poor streaming quality persists across YouTube and Netflix (This is traced to a lack of application prioritization)	Incorporate application prioritization into bandwidth allocation policies
GenAI generates accurate Wi-Fi access point distribution design using data on customer space and endpoint density	**DPI:** WiFi speeds on some access points are lower than expected (This is traced to physical obstructions impacting signal quality)	Factor in physical obstructions into access point distribution

Table 4: Uncovering gaps in policy, configuration and processes

4.4. Why choose a commercial DPI solution

DPI technology can be commercial, in-house or open source. Commercial solutions, built on proprietary technologies, offer broader and deeper analytics to meet the demands of GenAI training and fine-tuning. This is particularly important for networking functions that execute real-time, data-driven decisions. Apart from covering almost all latest applications and protocols in today's networks, commercial DPI solutions ensure scalability, adaptability to existing traffic reporting environments and a high level of vendor support, due to the expertise and experience that is accumulated from deployments in various environments and client setups. This enhances the quality of their traffic insights, and thus improves the accuracy of GenAI models.

In contrast, traffic analytics sourced from in-house DPI tools are typically confined to networking use cases and specific environments served by a vendor, limiting the ability of GenAI models to tackle situations that involve complex, multi-domain and novel scenarios. Similar concerns affect open-source DPI, as they often lack continuous updates and support.

The GenAI paradox

Access to GenAI modifies user behavior, leading to rapid growth in inference activity. This growth can be attributed to several factors. First, GenAI enables network administrators aiming to introduce new features and capabilities to experiment more intensively with their networks. Second, it allows for new levels of analysis and investigation, helping uncover deeper linkages within the network. Third, it pushes the introduction of GenAI-native workflows aimed at continuously improving network operations and control. Paradoxically, this surge in inferencing increases the need for high-quality data, further driving demand for technologies such as DPI.

5. DPI-DRIVEN GENAI-BASED NETWORK FUNCTIONS

Deploying DPI-driven analytics for GenAI—particularly for model training and fine-tuning—differs substantially from its traditional use in traffic management, due to the sheer volume of data required and the diversity of traffic analytics specific to each network function and sub-function. DPI's analytics must also correspond to the formats and content types of both the information sources ingested by GenAI models and the generative outputs expected during inferencing.

DPI provides the versatility and robustness that is required to support data requirements of GenAI-based network functions. Its comprehensive, granular, and highly customizable analysis can be adapted fully to any networking task. The following illustrates how DPI's analytics can be customized to each category of network functions and their key processes.

Network planning and design

Sample tasks

- Network goals
- Compliance requirements
- Network design
- Resource planning and provisioning
- Vendor selection
- Deployment timeline and milestone setting
- Redundancy planning
- Disaster recovery
- Asset management
  - Lifecycle management
  - Auditing

DPI-based traffic analytics

- Application types (e.g., YouTube, Microsoft Teams, Netflix, WhatsApp)
- Protocol types (e.g., TCP, UDP, QUIC, HTTP/3, SIP, DNS)
- Service types (e.g., VoIP, video streaming, online gaming, file sharing)
- Cloud / server identifiers (e.g., AWS EC2, Azure Front Door, Google Cloud Run, Akamai Edge)
- Device and equipment IDs (e.g., MAC addresses, DHCP hostnames, SNMP sysDescr, OUI)
- Device OS fingerprinting (e.g., Windows 11, Android 13, iOS 17, Linux Ubuntu)
- Destination URLs (e.g., cdn.netflix.com, api.dropbox.com, play.google.com)
- Host URLs (e.g., facebook.com, news.google.com)
- Geolocation data (e.g., New York, Tokyo, São Paulo, Frankfurt)
- Connection types (e.g., 5G NR, LTE, Wi-Fi 6)
- Encryption protocols (e.g., TLS 1.3, HTTPS, QUIC)

Network orchestration and traffic management

Sample tasks

- Policy and configuration
- Resource allocation
- Task execution (e.g., packet forwarding, routing, NAT, VLAN tagging and stripping, encapsulation and decapsulation, fragmentation and reassembly, compressing, caching, multicast replication, checksum verification, and error correction)
- Results monitoring
- Load, health and performance monitoring
- Load balancing and auto-scaling
- Documentation and reporting

DPI-based traffic analytics

- Network throughput (e.g., 850 Mbps between core routers during peak hours)
- Mobile network throughput (e.g., 100 Mbps downlink speed and 50 Mbps uplink speed on an LTE network)
- Bandwidth utilization (e.g., 97% utilization on the upstream link)
- Active connections (e.g., 140,000 concurrent TCP sessions on a firewall)
- Active user sessions (e.g., 49,000 logged-in mobile users on LTE)
- Session ID and subscriber mapping (e.g., Session ID xyx### mapped to IMSI 39082727377373732)
- Session duration (e.g., 12-minute SIP voice call)
- Flow start / stop timestamps and packet count (e.g., Start: 12:01:22, Stop: 12:04:31, 1,242 packets)
- NAT mapping info (e.g., 10.0.1.25 : 5000 translated to 203.0.113.15 : 61234)

Network automation

Sample tasks

- Automation framework and policy definition
- Design and modeling
- Automation configurations
- Script / code integration
- Automation testing
- Execution of automation
- Monitoring
- Error handling
- Documentation and reporting
- Optimization

DPI-based traffic analytics

- Total throughput (e.g., Gbps per link, interface, or application)
- Per-flow throughput (e.g., 120 Mbps average for a video streaming flow)
- One-way latency between network functions (e.g., 4 ms between router and firewall)
- Round-trip time (RTT) (e.g., 30 ms RTT from user device to application server)
- Path selection convergence time (e.g., 200 ms to reconverge routing after link failure)
- Flow setup time (e.g., 90 ms to establish a new HTTPS session)
- Concurrent flows handled (e.g., 50,000 concurrent flows on a load balancer)
- Throughput per function (e.g., 950 Mbps through a caching proxy)
- Traffic load per function (e.g., firewall processing 750 Mbps out of 1 Gbps capacity)
- Queue depth (e.g., average of 20 packets queued per interface)
- Queue wait time (e.g., 8 ms average wait on the ingress interface)
- Service function chain latency (e.g., 10 ms total across router / firewall / caching engine)
- Hop count reduction (e.g., reduced from 6 to 3 hops due to automated path optimization)
- Automated remediation time (e.g., 29 seconds to reroute traffic after a failure)
- Dynamic path selection time (e.g., 145 ms to select a lower-latency path)
- Network function uptime (e.g., 99.8% uptime of a caching engine)
- Dropped packets per network function (e.g., 3% of packets dropped by the router due to buffer overflow)
- Session counts per network function (e.g., 20,000 active sessions handled by the NAT gateway)

Network monitoring

Sample tasks

- Telemetry system setup
- Trigger and alert configurations
- Monitoring and measurement
- KPI analysis
- Periodic reporting
- Anomaly detection
- Fault management
- Troubleshooting
- Diagnosis
- Root cause correlation
- Remediation

DPI-based traffic analytics

- Packet loss (e.g., 2% packet loss detected during video streaming)
- Packet duplication (e.g., 0.1% of packets are duplicated for file transfers)
- Packet size distribution (e.g., 60% of packets above 1200 bytes)
- Packet round trip time (e.g., 100 ms round-trip time for ping packets)
- Packet errors (e.g., 0.5% of packets have checksum errors in TCP traffic)
- Flow session count (e.g., 1,000 active sessions handled by a load balancer)
- Application failure rate (e.g., 1% failure rate for login requests)
- Retransmission rate per application (e.g., 1.2% TCP retransmissions on Teams traffic)
- New application detection (e.g., new HTTP-based application detected on port 8080)
- TCP connection establishment time (e.g., 135 ms average TCP handshake time for HTTPS traffic)
- Protocol usage distribution (e.g., 65% of traffic is HTTP, 25% is HTTPS, 10% is DNS)
- Protocol errors (e.g., 2% of DNS queries result in timeouts)
- TCP retransmission count (e.g., 0.2% of TCP connections experience retransmissions)
- Packet forwarding rate (e.g., 1.2 million packets/sec handled by a core router)
- Network utilization (e.g., 85% utilization on the main data center link)
- Network latency (e.g., 50 ms network latency between data center and user devices)
- Function health (e.g., 100% uptime across critical routers in the network)
- VNF processing latency (e.g., 8 ms added by firewall VNF)
- Packet drop rate per function (e.g., 2% of packets dropped by NAT device during overload)
- Mobile handover latency (e.g., 62 ms handover time)
- Mobile data session drops (e.g., 0.7% drop rate for mobile data sessions in urban areas)
- Cloud storage latency (e.g., 45 ms latency for object storage access in AWS S3)
- Cloud network path latency (e.g., 145 ms average latency between cloud data centers in different regions)
- Cloud virtual private network (VPN) throughput (e.g., 640 Mbps throughput over cloud VPN connection)

Common modalities for generative outputs in GenAI-based network functions

Text | Maps | Charts | Graphs | Topologies | Diagrams | Models | Tables | Plots

Decision trees | Word clouds | Flowcharts | 3D renderings | Infographics | Videos

Service assurance and optimization

Sample tasks

- SLAs
- QoS policies
- Traffic classification
- Policy enforcement
- QoS monitoring
- SLA monitoring
- AI/ML-driven resource optimization
- Congestion handling
- Redundancy management
- Predictive analytics
- Reporting
- Feedback and fine-tuning

DPI-based traffic analytics

- Traffic latency (e.g., average one-way latency of 115 ms between branch and data center)
- Jitter (e.g., 6 ms packet delay variance in a VoIP stream)
- Packet Loss Rate (e.g., 0.3% loss in a video stream)
- Network Availability (e.g., 99.99% uptime availability over a month)
- Error Rate (e.g., 0.001% frame error rate)
- QoS Tags by packets (e.g., DSCP EF for Zoom video conferencing traffic)
- Application-Specific Latency (e.g., 17 ms for SAP, 39 ms for Office 365)
- Application speeds (e.g., 135 Mbps average download speed for video streaming services)
- Application Response Time (e.g., 190 ms average response time for an e-commerce website)
- TLS handshake time per application (e.g., 155 ms TLS handshake time for Zoom)
- Segment Performance Deviation (e.g., cloud segment adds 17 ms latency to end-to-end path)
- Radio Access Latency (e.g., 6 ms for 5G NR latency)
- QoS Flow Bit Rate (e.g., Guaranteed Bit Rate of 10 Mbps for URLLC)
- Cloud Latency (e.g., 18 ms latency to Azure services hosted in the West Europe region)
- Function Failover Time (e.g., 250 ms to switch to redundant firewall)
- Function Utilization Efficiency (e.g., only 65% of router's potential bandwidth being used)
- Traffic Rebalancing Efficiency (e.g., 85% optimal path selection under dynamic load)
- Congestion notification rate (e.g., Explicit Congestion Notification bits marked on 13% of packets)
- Network Congestion (e.g., 5% of network paths experiencing congestion during peak hours)

Common data types generated via GenAI in the context of network management

Rules	Policies	Configurations	Thresholds	Parameters	Templates	Standards		
Profiles	Reports	Metrics	KPIs	Strategies	Alerts	Logs	Protocols	Schedules
Definitions	Conditions	Notifications	Workflows	Models	Templates	Adjustments		
Tests	Filters	Plans	Diagnostics	Inventories	Audits	Summaries	Guidelines	

Network analytics

Sample tasks

- Data collection and processing
- Integration into analytics ecosystem
- Data storage
- Data protection and security
- Data compliance
- Real-time data analysis
- Trend analysis
- Predictive analysis

DPI-based traffic analytics

- Encrypted traffic / time (e.g., TLS 1.3 traffic volume increases 20% in 2024)
- QoS metrics trends / time (e.g., rising jitter in video conferencing over several weeks)
- Bandwidth usage per application / time (e.g., 37% of the total used by YouTube last week)
- Application session drops / time (e.g., frequent drops in VPN sessions in 2023)
- Application usage trends / time (e.g., Google Drive usage increased 12% in 2024)
- New applications detected over time (e.g., Edits was detected in May 2025)
- Traffic per device type / time (e.g., smartphones consuming 65% of peak-time bandwidth in 2024)
- User behavior / time (e.g., average user traffic up 30% over past 6 months)
- Failed sessions / time (e.g., 4% TCP handshake failures last month)
- VPN usage patterns (e.g., VPN traffic usage peaks between 9 AM and 5 PM)
- WiFi usage patterns (e.g., Wi-Fi users increased by 25% in the last quarter)
- Function error rate (e.g., 1% error rate on the core switch over the last week)
- API breakdown (e.g., authentication APIs accounting for 30% of all cloud API calls in Q1 2025)
- Function filtering performance (e.g., an SWG filtered 1,000 URLs per minute)

Network security

Sample tasks

- Framework and policy design
- Risk and vulnerability assessment
- Setup and integration
- Threat intelligence
- Threat monitoring and detection
- Alerting
- Remediation
- Reporting
- Impact assessment
- Security auditing

DPI-based traffic analytics

- Anomalous user behavior (e.g., Sudden data upload spike from a user at 12 am)
- Anomalous application behavior (e.g., WhatsApp sending traffic on non-standard ports)
- Types of malicious traffic (e.g., C2 communication)
- SYN flood patterns (e.g., High rate of SYN packets to port 80 from a single IP)
- Data exfiltration patterns (e.g., Repeated outbound DNS TXT record usage)
- Abuse of non-standard ports (e.g., SSH traffic on port 8080)
- DNS tunneling detection (e.g., large volume of long subdomain DNS queries)
- App mimicry attempts (e.g., custom app mimicking Dropbox traffic patterns)
- Share of encrypted, obfuscated, and anonymized traffic (e.g., 65% of traffic identified as VPN or Tor)
- Share of encrypted threats (e.g., Malware delivered over HTTPS)
- Irregular usage patterns (e.g., High bandwidth usage from IoT device during off-hours)
- File transfer size (e.g., 5GB file sent over FTP)
- File transfer protocols (e.g., FTP, SFTP, SMB, HTTP)
- Bad websites and applications attempts (e.g., malicious-domain(.)tld)
- Infected devices (e.g., a laptop sending repeated DNS requests to known malware domains)
- Tethering traffic (e.g., phone IP routing multiple distinct device traffic types)
- TLS handshake metadata (e.g., JA3 fingerprint indicating known malware signature)
- High-risk geographical traffic patterns (e.g., frequent connections to IPs in sanctioned regions)

6. GENAI USE CASES

To further examine how DPI-driven traffic insights augment GenAI capabilities in network functions, this white-paper discusses its use in two networking use cases: a GenAI-based load balancer and a GenAI based cloud access service broker (CASB).

6.1. GenAI-based load balancer

A GenAI-based load balancer enables users to draw upon its generative capabilities to distribute traffic to servers in a subsystem. The use of DPI-based analytics enhances model training and fine-tuning, enabling the load balancer to acquire domain knowledge, which includes an understanding of load balancing tasks, policies, techniques, and actions, as well as knowledge of the network and the traffic it handles. Table 5 shows how DPI-driven traffic metrics enhance a load balancer's generative and reference outputs.

Load balancer processes	DPI-derived analytics	GenAI outputs
Distribution of traffic	► Traffic volume ► Session count	Optimal algorithms for distributing traffic to servers in the pool (e.g., round robin, least connections, IP hash)
Traffic decryption	► Encryption protocol ► Share of encrypted traffic	Rejection policies for traffic encrypted with legacy or weak protocols (e.g., immediate blocking of SSLv2, SSLv3, RC4, DES)
Subsystem monitoring	► Server response time ► Server throughput	Server health thresholds to determine the eligibility of a server to receive traffic (e.g., server response time threshold: 300ms; server throughput usage threshold: 80%)
Session persistence	► Session affinity duration ► Session error rates	Techniques to enforce session persistence (e.g., SSL Session ID, cookie-based)
QoS management	► Application latency ► Packet loss rate	List of priority applications and corresponding rules for prioritization (e.g., video streaming applications are classified under priority apps and directed to the highest performing server)

Table 5: Use of DPI in model training and fine-tuning in a GenAI-based load balancer

DPI-driven analytics also benefits GenAI-based load balancers in the following ways:

- **GenAI model testing**: Benchmark a load balancer's GenAI outputs against outputs by human users (e.g., using DPI data, session persistence patterns resulting from GenAI recommendations are compared against those from manual configurations)
- **Prompting and querying**: Enrich prompt information (e.g., using current traffic profile data from DPI to generate rules for throttling low-priority applications)
- **Environment assessment**: Assess how network conditions impact GenAI processes in a load balancer (e.g., using DPI data to detect TCP connection resets and out-of-order packets within links to cloud storage networks, helping pinpoint intermittent connectivity issues that can affect GenAI database access)
- **GenAI externalities**: Determine how GenAI workloads affect other network processes (e.g., using DPI data on packet loss, connection timeouts and latency to detect instances of bandwidth contention that correlate with increases in GenAI inferencing traffic)
- **Solution architecture**: Assess if the GenAI architecture behind a load balancer is optimized (e.g, using DPI-derived speed and latency metrics to evaluate how auto-scaling of inference servers improves load balancer performance)

- **Network outcomes**: Assess how the combination of a load balancer's policies, configurations and GenAI outputs influence network outcomes (e.g., using DPI-based traffic analytics such as latency variance to identify alleviation of congestion in southbound systems)
- **Security risks**: Identify GenAI-load workloads that are under attack (e.g., using anomalous traffic patterns tracked by DPI to detect resource exhaustion attacks)

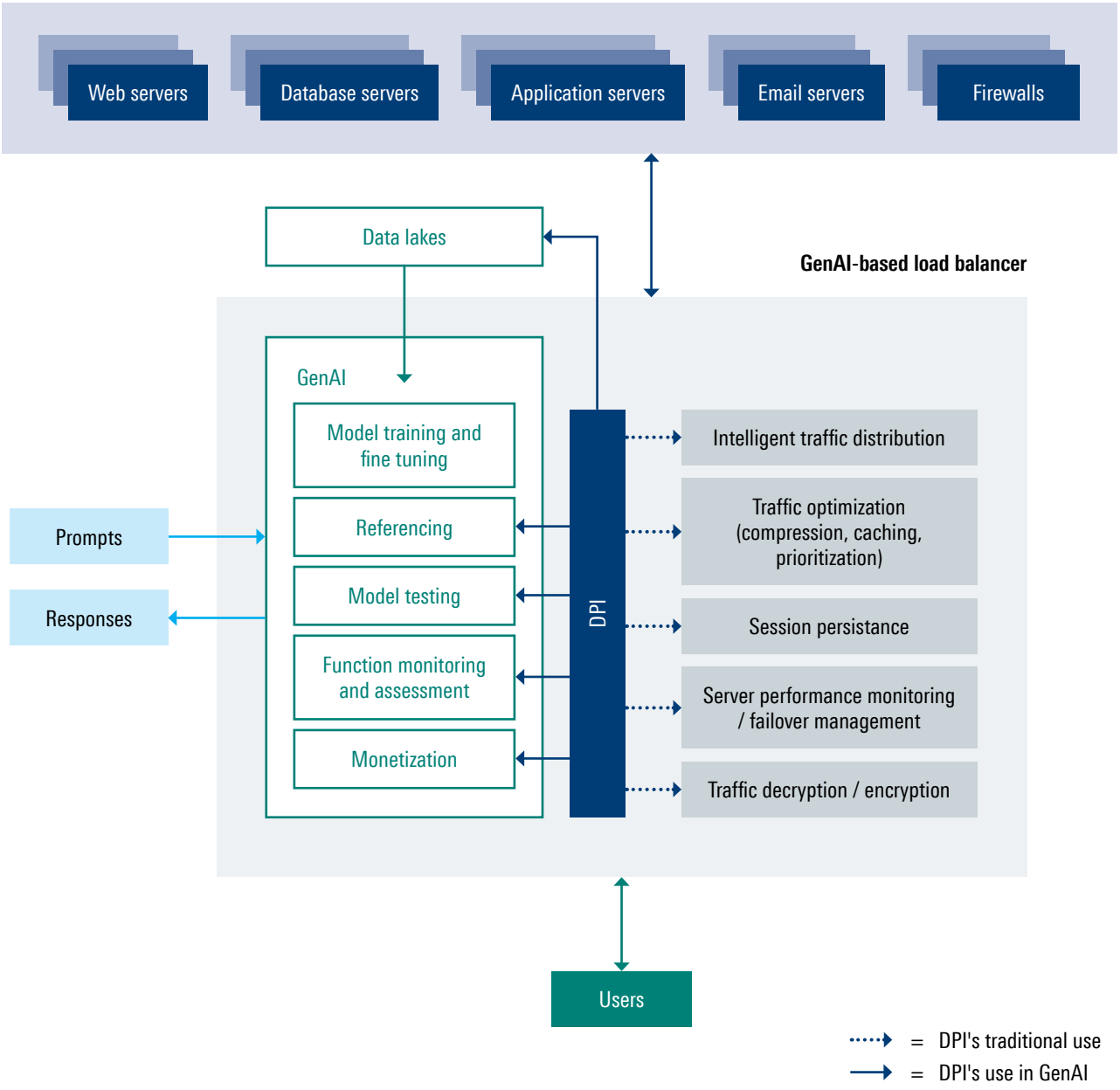

Diagram 6: Utilization of DPI-driven traffic insights in a GenAI-based load balancer

6.2. GenAI-based CASB

By integrating GenAI capabilities into a CASB, enterprises gain access to contextual and realistic data that can be used to filter and control traffic flows to their cloud platforms and applications. Leveraging DPI-based traffic analytics for training and fine-tuning, a CASB's GenAI model accelerates its learning and acquires accurate insights into the workings of the CASB, the enterprise's cloud application portfolio, commonly encountered security threats, and the current state of the network. Table 6 lists examples of GenAI outputs from a CASB that benefit from DPI's fine-grained analytics.

CASB processes	DPI-derived analytics	GenAI outputs
Traffic monitoring	► Application names ► Service types	Rate of filtering for latency-sensitive traffic (e.g., a random sampling method that selects 1 out of 10 flows is used to filter financial trading traffic)
Access control	► Number of sessions ► User IPs	Access and session policies that adjust dynamically to user behavior (e.g,. enforcing two-factor authentication and sending an SMS notification for sessions exceeding 2 hours)
Data loss prevention	► Large data transfers ► Frequency of access	Thresholds for transactions involving data transfers, based on application classes (e.g., maximum file size of 20GB or a limit of 5 transfers per user per day for cloud storage applications)
Threat detection	► Irregular traffic patterns ► Device type	Alerting configurations for suspicious user behavior (e.g., SOC notification for unrecognized devices or login attempts exceeding 3 times within 30 minutes)
Data compliance	► Database-related sessions ► Encrypted data transfers	Listing of high risk applications that require intensive monitoring and analysis (e.g., all applications handling credit card and banking information)

Table 6: Use of DPI in model training and fine-tuning in a GenAI-based CASB

DPI data also benefits GenAI-based CASBs in the following ways:

- **Model testing**: Compare CASB performance before and after the implementation of GenAI (e.g., using DPI data to determine the improvements in latency for key enterprise cloud applications)
- **Prompting and querying**: Provide contextual input for GenAI engines (e.g., leveraging real session data from DPI to prompt a CASB to extract summaries of remote user behavior)
- **Environment assessment**: Evaluate how broader network issues affect GenAI processes in a CASB (e.g., using DPI's data on throughput and latency to detect overuse of shared resources—such as inference GPUs—in multi-tenant environments, which can slow down model execution)
- **GenAI externalities**: Measure how GenAI processes across a vendor-hosted CASB impact an enterprise network (e.g., using DPI-based analytics to identify an increase in session counts and API request spikes between the CASB and enterprise devices / servers, and the effect on enterprise bandwidth, latency, and network speeds)
- **Solution architecture**: Assess whether GenAI processes and workflows are optimized to ensure CASB core functionalities are not affected (e.g., DPI analytics on vector database query volumes and response times can be used to scale database clusters, improve prefetching mechanisms, optimize regional data access, and enhance indexing strategies)
- **Network outcomes**: Evaluate if the mix of policies, configurations and GenAI models has improved threat detection (e.g., using DPI to detect suspicious incidents such as access from anonymized IPs, frequent file renaming, inconsistent TLS fingerprinting and large volumes of base64-encoded payloads over HTTPS).

In this cases, anonymized IPs can be identified via IP reputation services, geolocation data, and inspecting TLS handshake metadata for anomalies, helping surface unauthorized access. Similarly, frequent file renaming can indicate staging or obfuscation prior to exfiltration, especially when upload bursts or unusual file types are involved. For TLS fingerprinting, frequent changes in JA3/JA3S hashes may indicate attempts to bypass controls or emulate multiple applications. Finally, large volumes of base64-encoded payloads can be detected by analyzing entropy, payload size, and transmission bursts, allowing network administrators to flag potential covert transfers or even malicious prompt injection into GenAI APIs.

- **Security risks**: Securing a CASB from threats targeting GenAI workloads (e.g., DPI data on repeated requests to model APIs, requests containing malicious code, or incessant querying from a single IP address can be used to detect model poisoning and other malicious activity)

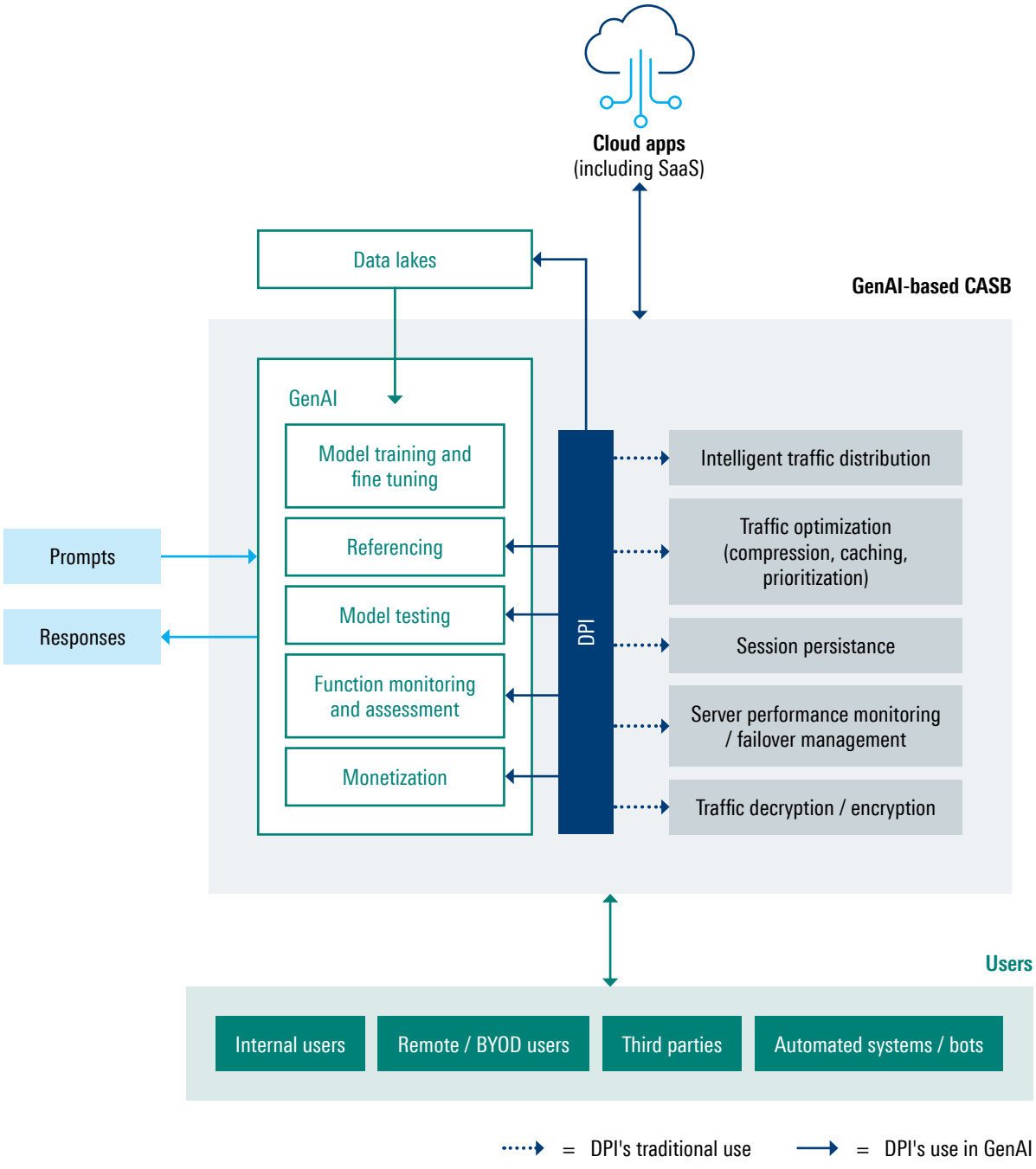

Diagram 7: Utilization of DPI-driven traffic insights in a GenAI-based CASB

7. CONCLUSION

Network intelligence is central to GenAI implementations in network management. DPI, as an advanced technology to extract and deliver traffic analytics, contributes to GenAI-based network functions by furnishing comprehensive and granular traffic inputs, primarily for model training and fine-tuning. Real-time analytical outputs from next-gen DPI engines such as R&S®PACE 2 and R&S®vPACE can be incorporated directly into GenAI training funnels, or adapted into various derivative outputs that are subsequently added to these funnels.

By training on DPI-based information sources, GenAI-based network functions gain access to GenAI-derived domain knowledge that is contextual to their functionalities and tasks. This knowledge, when coupled with GenAI-driven natural language interactions, enables network administrators to generate realistic responses to complex situations, speeding up decision-making and improving the quality of these decisions.

Beyond model training and fine-tuning, network administrators can leverage DPI's fine-grained traffic insights to:

1. Test the accuracy of a GenAI model by validating its responses against real data
2. Enhance GenAI prompting and querying
3. Improve GenAI architecture and processes
4. Observe the interrelationships between GenAI workloads and other components in the network
5. Improve the security of GenAI assets and traffic flows
6. Monitor and diagnose issues within GenAI's complex and distributed architectures
7. Measure the effectiveness of GenAI-based network functions in delivering network outcomes
8. Create shared traffic intelligence for different network functions, including traditional and GenAI-based functions, for greater consistency in network policies and actions
9. Improve monetization with add-on services based on robust and reliable analytics

The combination of advanced traffic classification techniques, encrypted traffic intelligence, and various customization features provided by R&S®PACE 2 and R&S®vPACE delivers the versatility that is needed to cater to key functional areas—network planning, traffic management, automation, monitoring, optimization, analytics, and security. This is illustrated in this whitepaper with various corresponding traffic metrics that relate specifically to distinct tasks under each area.

The use of DPI-powered analytics for GenAI builds on its established role in network functions, where it already fuels various data-driven decisions. As in the cases of load balancers and CASBs, it is clear that DPI forms a crucial information backbone that supports both GenAI and traditional processes.

The age of GenAI is just beginning

While the adoption of GenAI in network management is still in its early stages—centered primarily around LLM-based models and natural language chat interfaces—network vendors are already embarking on their next GenAI frontiers. This includes incorporating search capabilities, adopting RAG models and rolling out AgenticAI. These advancements will significantly increase the demand for next-gen DPI, pushing its deployment across more network functions.

At every level, DPI is essential not only for unlocking the full potential of GenAI in managing modern networks, but also for future-proofing these networks against the surge in traffic and complexity that GenAI itself is creating—in an era propelled by AI and its limitless potential.

ipoque

ipoque, a Rohde&Schwarz company, is a global leader in the world of network analytics software. We leverage our deep domain expertise to create customized software solutions that empower our customers to transform data into intelligence. As a subsidiary of Rohde&Schwarz, we take advantage of potential synergies.

Rohde&Schwarz

Rohde&Schwarz is striving for a safer and connected world with its Test & Measurement, Technology Systems and Networks & Cybersecurity Divisions. For over 90 years, the global technology group has pushed technical boundaries with developments in cutting-edge technologies. The company's leading-edge products and solutions empower industrial, regulatory and government customers to attain technological and digital sovereignty. The privately owned, Munich based company can act independently, long-term and sustainably. Rohde&Schwarz generated a net revenue of EUR 3.16 billion in the 2024/2025 fiscal year (July to June). On June 30, 2025, Rohde&Schwarz had more than 15,000 employees worldwide.

Rohde&Schwarz GmbH&Co. KG
www.rohde-schwarz.com

ipoque GmbH
Augustusplatz 9 | 04109 Leipzig, Germany
Info: + 49 (0)341 59403 0
Email: info.ipoque@rohde-schwarz.com
www.ipoque.com

Version 01.00 | November 2025
White paper | From data to decisions: How generative AI and DPI are shaping the future of network management