CAICT 中国信通院

SD-WAN + AI: Service Paradigm Shift, Key Technologies and Evolution Path

China Academy of Information and Communications Technology（CAICT）

2026.9

Directory

About us

- The Computing-Network Integration Research Team at the Information and Communications Convergence Innovation Center of the China Academy of Information and Communications Technology (CAICT) focuses on **"three major trends and one key infrastructure." Computing-network integration spans multiple sectors, including the Internet, supercomputing, telecommunications, finance, and energy, and its technical framework integrates innovative capabilities across computing, communications, and security.**

Technical Standards Development

- Through **ITU-T SG13**, initiated research on standards for **IP resilience networks**.
- Through **IEEE** and **TC260**, initiated standards development for **computing collaboration** and **Zero Trust maturity models**.
- Through **TC3** and **TC8**, jointly initiated a series of standards covering **Computing Power Networks, Software-Defined Wide Area Networks (SD-WAN), and SASE**; submitted **10+ standards proposals** and contributed to the development of **20+ standards documents**.
- Through the **Computing-Network Integration Standards Working Group of the China Institute of Communications**, jointly initiated standards on **computing-network infrastructure, Zero Trust, edge computing, high-performance computing(HPC), and multimodal networks**; published **20+ standards documents** and promoted the development of **50+ standards**.

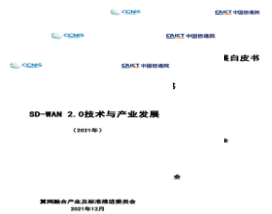
Industry Reports / White Papers

Commissioned by relevant organizations to organize and publish white papers and blue books:

- **3 white papers on Computing-Network Integration**
- **6 white papers on SD-WAN**
- **3 white papers on HPC Technology**
- **1 white paper on Zero Trust Technology**

Research reports:

- **2 reports on Networking of Computing Resources**
- **2 research reports on Computing-Enabled Networking**
- **2 research reports on Computing-Network Security**
- **2 research reports on Computing-Network Infrastructure**



Research Programs / Scientific Research Projects

- Led or participated in **10+ research projects** on key topics such as **"Eastern Data, Western Computing," "Computing Power Networks," "New Infrastructure," and Future Networks**, organized by the **Ministry of Industry and Information Technology (MIIT)** and relevant institutions.
- Undertook multiple research tasks under the **2018–2020 National Key R&D Program** of the **Ministry of Science and Technology (MOST)**, including projects related to **broadband communications and new-generation networks**.
- Participated in a series of ministerial-level research programs, including **MIIT high-quality development initiatives** and **MOST programs on mobile information networks.**

About us

Standards & Research as Dual Drivers

➢ Develop an SD-WAN/SASE Standards Framework

- Led the development of **10+ SD-WAN/SASE industry standards**, covering overall architecture, value-added services, interface protocols, QoS, and other areas.
- Led the development of **10+ SD-WAN/SASE association standards**, covering general technical requirements, IPv6 integration, maturity assessment, and other areas.

➢ Conduct In-Depth Industry Research

- Continuously published 5 white papers, including Global SD-WAN Technology and Industry Development and SD-WAN 2.0 Technology and Industry Development.

➢ Obtained 5 technology patents and 10+ software copyrights.

Build a Testing & Validation Platform

➢ Establish a Testing & Validation Platform

- Built an **SD-WAN/SASE testbed** and **independently developed 20+ SD-WAN/SASE technical validation capabilities**.

➢ Launch Industry Certification Programs

- **Initiated a series of industry certification programs, including SD-WAN Ready 1.0/2.0, SD-WAN Service Quality, SD-WAN + Information Technology Application Innovation, and SASE Capability Maturity, with participation from 100+ vendors.**

Deep Integration of Consulting & Case Studies

➢ Collect Industry Best Practices

- Continuously organized the **SD-WAN/SASE Best Practice Case Collection from 2020 to 2025**, with 100+ outstanding cases collected to date.

➢ Provide Network Planning & Consulting

- Supported China Yangtze Power in its **"136" Digital Transformation and Upgrade Project.**
- Supported China Oilfield Services Limited (COSL) in its **Digital Transformation Project for the Traditional Energy Industry**

Connect Industry Stakeholders

➢ Organize Industry Forums

- From 2019 to 2025, continuously organized the **SD-WAN Industry Development Forum and SD-WAN/SASE Technology Salons**.

➢ Publish the Industry Landscape

- From **2022 to 2025**, continuously organized industry-wide data collection for the **SD-WAN Industry Landscape** and published successive editions of the **SD-WAN Industry Landscape**.

SD-WAN + AI: Industry Demand Is Shifting from "Connectivity Upgrades" to "Intelligent Service Transformation"

CAICT 中国信通院

Connectivity
Always Connected
End-to-End Overlay/Underlay Interconnection

Awareness
Understand the Network
DPI / App-ID / User-ID Awareness

Orchestration
Intelligent Path Selection
PBR / FEC / Multi-Path Intelligent Path Selection

Assurance
Controllable
QoS / Security Micro-Segmentation Enforcement

Autonomy
Self-Driven
Intent-Driven Networking (IDN)

Service
Outcome-Based Delivery
AI-native SD-WANaaS

AI Requirements for SD-WAN: From "High Bandwidth" to "Task Awareness and Assured Experience"

High Bandwidth + Traffic Bursts	Concurrent agent workloads, model weight/data synchronization, and video/multimodal data make AI traffic increasingly **bursty and bandwidth-intensive**.
Low Latency + Predictability	User experience is affected by **TTFT (Time to First Token) and ITL (Inter-Token Latency)**. A single request may also involve multiple agents and tools in sequence.
Task Awareness	The network needs to understand whether traffic is associated with **inference, retrieval, synchronization, or control**, and select the appropriate path accordingly.
Multi-Cloud / Multi-Compute Coordination	AI computing resources are becoming distributed across **central clouds, regional clouds, edge nodes, and private AI clusters**.
Data Security	Prompts, enterprise knowledge, and model outputs may contain sensitive data, requiring coordinated **routing, access-control, and data-security policies**.

AI Requirements for SD-WAN: The Network Must Not Only Be AI-Assisted, but Also "Understand AI Workloads"

1. **Business Data Awareness & Network Modeling**
   Identify applications, users, devices, links, and task context, and establish relationships among tasks, paths, and resources.

2. **Task-Traffic Prediction & Network Service Reasoning**
   Move from real-time monitoring toward anomaly prediction, root-cause analysis, and capacity forecasting, reducing manual intervention after network failures occur.

3. **Natural-Language Network Operations**
   Translate natural-language intents such as "Prioritize video conferencing" or "Increase the priority of specific AI traffic" into executable network policies and configurations.

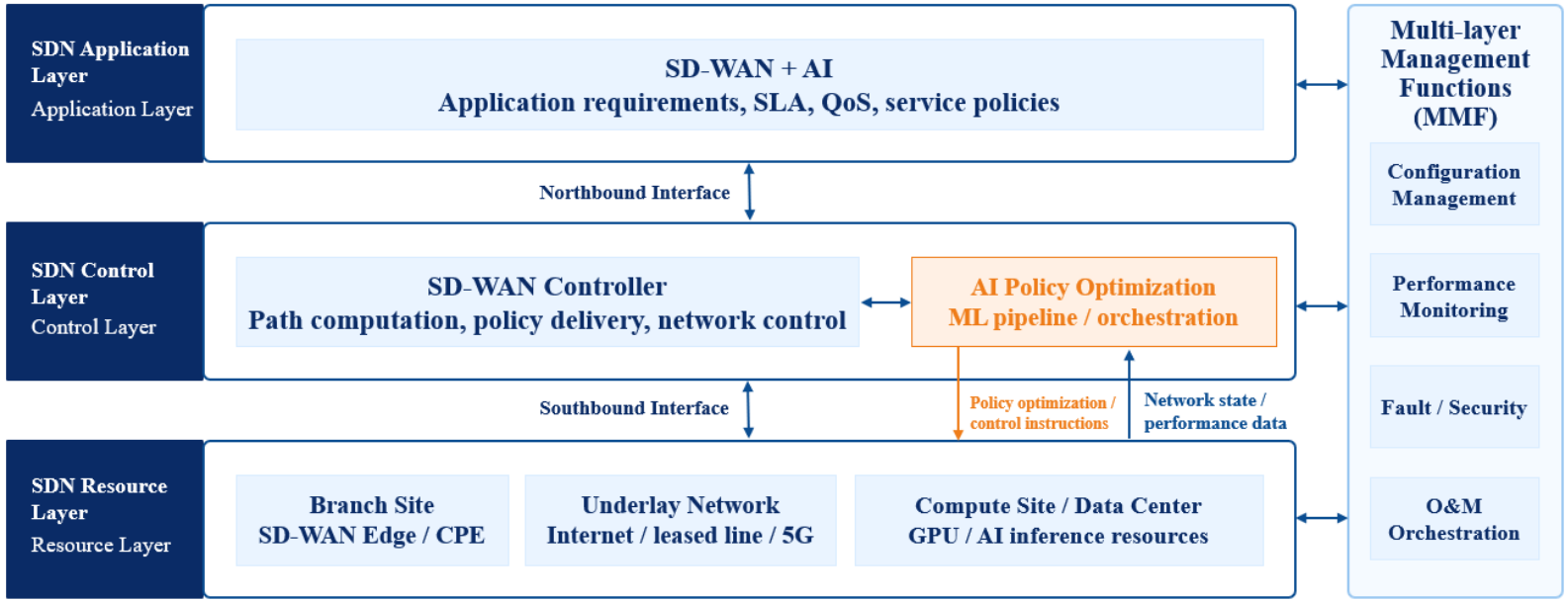
4. **AI-Driven Security Services**
   Identify AI traffic, Shadow AI, anomalous behavior, and data leakage risks, enabling coordinated network and security operations.

5. **Autonomous Operations & Orchestration**
   Under defined guardrails, approval mechanisms, and rollback conditions, enable SD-WAN intelligent controllers to automatically perform configuration, scheduling, validation, and closed-loop optimization.

SD-WAN + AI Functional Architecture: Building a Closed Loop of "Intent – Task – Resource – Policy"

CAICT 中国信通院

- **AI is redefining what SD-WAN is expected to do: evolving from connecting services to understanding services, assuring services, and autonomously operating services.**



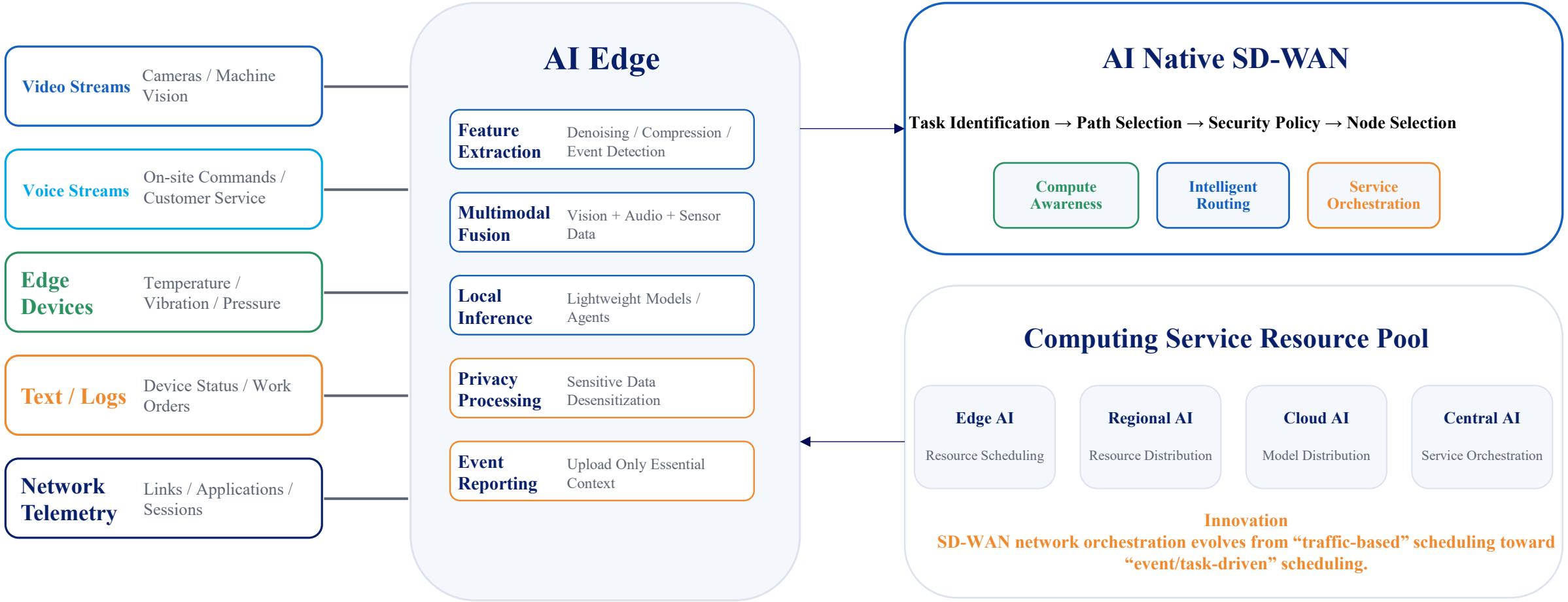
Deployable Closed-Loop Control

① Observe	Network / Application / Security / Task Monitoring
② Reason	Intent Understanding / Anomaly Detection / Root-Cause Analysis
③ Decide	Task SLA / Cost / Security Policy
④ Orchestrate	Routing / Computing / Security / Service Orchestration
⑤ Execute	Configuration Deployment / Link Scheduling / Policy Enforcement
⑥ Validate	Service Outcome Validation / Rollback / Learning

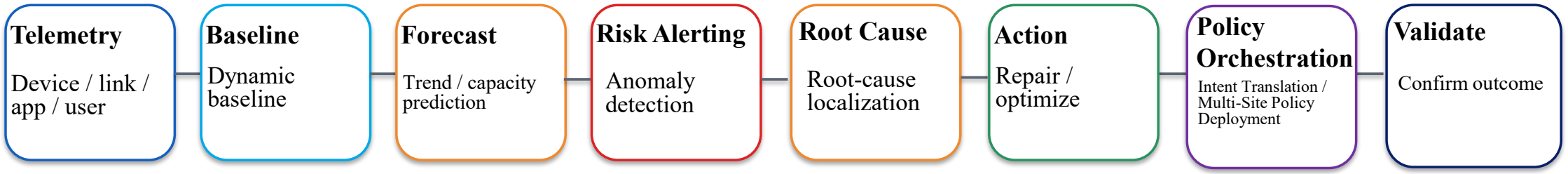
Key Architectural Principles

- SD-WAN is responsible for reliable connectivity and traffic execution
- AI is responsible for understanding, prediction, decision-making, and generation
- Business systems provide service intent and acceptance criteria

- **SD-WAN + AI Edge enables real-time awareness and lightweight inference for multimodal data close to the data source. Through a layered architecture of endpoint filtering – edge decision-making – cloud enhancement, and by leveraging SD-WAN orchestration across computing and networking resources, it achieves a dynamic balance between low-latency response and global intelligence.**



- **Session-level telemetry, dynamic baselines, and trend forecasting enable SD-WAN to shift from alert-driven to prediction-driven operations, allowing failures to be mitigated hours or even days in advance.**



Four-Dimensional Blast-Radius Assessment Model for SD-WAN Network Failures

- **Spatial Dimension:** Number of affected sites, links, and users
- **Topology Dimension:** Position and connectivity of the failed node within the Overlay topology
- **Service Dimension:** Number and criticality of affected business systems
- **Time Dimension:** Relationship between fault duration and the service SLA tolerance window

Link Degradation Risk

Based on Underlay and Overlay dual-layer telemetry, predict link degradation trends. Before VoIP quality is affected, trigger FEC-based adaptation or dynamic path switching, such as PBR, to proactively avoid high-loss or high-jitter links.

Capacity Risk

Based on application traffic profiles and growth curves, predict bandwidth saturation risk and coordinate SD-WAN intelligent path selection with computing-resource scheduling, such as 5G / broadband hybrid access, so that capacity expansion is triggered before congestion occurs.

Configuration Risk

Use network digital twins to simulate planned changes to SD-WAN Overlay policies before deployment. Validate new policies against higher-level intent, with support for automated rollback in case of failure.

Experience Risk

End-to-End Client-to-Cloud Path

Use Application-Aware Routing to steer traffic toward the optimal application path. If the current path becomes unavailable, the system can automatically reroute to an alternative path or a better-performing POP, instead of waiting for user complaints.

Security Risk

Based on SASE / SSE policy convergence, identify Shadow AI access behavior and coordinate SD-WAN security policies. If anomalous cross-site access or suspicious outbound traffic is detected, dynamically restrict access or reroute traffic to inspection nodes, forming a closed loop from detection to containment.

SD-WAN + AI Key Technology 3: Natural Language as the New Network Interface, Intent as the New Programming Model

CAICT 中国信通院

■ **Natural language expresses service intent, while LLMs translate intent into network policies—shifting SD-WAN from manual CLI configuration to intent-driven networking.**

Network Intent

Intent Parsing

Extract key service entities, including:
- **Application identity (App-ID)**
- **User identity (User-ID)**
- **Site / Region**
- **Time dimension** such as peak periods
- **Service criticality** such as mission-critical / general

Policy Decomposition

Map intent into SD-WAN-specific policy objects:
1.Application-aware routing: Identify VISION_QC application characteristics
2.SLA policies (SLA Profile): latency < 20 ms, jitter, packet-loss rate
3.Path control (PBR): Prefer MPLS underlay, fail over to 5G upon failure
4.Security boundary (SASE): ZTNA access control + DLP protection
5.QoS and queue scheduling: High-priority DSCP marking and priority queues

Conflict Validation

Use a **Network Digital Twin** for simulation and validation.
•**Topology feasibility:** whether an available path exists between sites/regions
•Whether bandwidth capacity is sufficient
•Whether global compliance constraints are violated, such as data sovereignty requirements
•Whether routing loops or tunnel flapping may be introduced

Configuration Generation

Generate **Intent API objects** for the SD-WAN controller, rather than traditional CLI scripts, with support for modern network programming interfaces such as:
SRv6 Policy / OpenConfig / NETCONF / YANG

Controller Orchestration & Execution

```
{
 "Intent_ID": "Mfg_Vision_QC_East_China",
 "App_Profile": "VISION_QC (DPI/App-ID)",
 "SLA_Profile": {"Latency": "<20ms", "Jitter": "<5ms", "Loss": "<0.1%"},
 "Overlay_Topology": {"Primary": "MPLS", "Backup": "5G", "Failover": "Auto"},
 "Path_Control": {"PBR": "EDGE->REGION", "FEC": "Adaptive", "Duplication":
"Off"},
 "QoS": {"DSCP": "EF", "Peak_Priority": "Enable"},
 "Security": {"ZTNA": "Enforce", "DLP": "Block_Cross_Region", "SASE_PoP":
"East_China"},
 "Compliance": {"Data_Residency": "CN-EAST", "Audit": "Enable"}
}
```

➔ **Policy Engine / API / Orchestrator**
➔ **SD-WAN Controller** — Global Orchestration
➔ **CPE / Edge** — Local Execution: Tunnel Establishment / PBR / FEC / QoS
➔**SASE PoP** — Security Policy Enforcement

Key Controls
Semantic Validation
+ Policy Conflict Check
+ Digital Twin Simulation
+ Approval & Canary Deployment
+ Validate & Rollback

SD-WAN + AI Key Technology 4: Proactive Intelligent Security — Unifying "AI Traffic Identification" with "Network Security Policies"

CAICT 中国信通院

- **With SD-WAN as the intelligent traffic steering hub and SASE/SSE as the deep inspection engine, L3–L4 network telemetry is deeply integrated with L7 AI application identification. This enables a shift from "detecting attacks" to "understanding AI-related business risks and intervening proactively," forming a closed-loop AI security defense framework of Identify → Steer → Mitigate → Close the Loop.**

Risk Type	Typical Risk	AI + SD-WAN Awareness	Proactive Defense Action
Shadow AI	Employees privately accessing external AI models	SD-WAN DPI identifies AI application characteristics and DNS requests; combined with SASE/SWG for user identity and behavior analysis.	**PBR Redirection:** Steer Shadow AI traffic to an enterprise-approved AI gateway, or block it directly through **Local Internet Breakout**
Data Leakage	Sensitive data contained in prompts or model outputs	SD-WAN flow-level telemetry (IPFIX) combined with CASB/DLP for LLM interaction auditing and sensitive-data tagging.	**Dynamic Micro-Segmentation:** Work with **ZTNA** to isolate network access for suspicious endpoints, restrict lateral movement, or block outbound API calls
Prompt Injection	Malicious input manipulating AI systems	Integrate with AI firewalls / WAFs for input/context risk detection; SD-WAN provides abnormal network session characteristics.	**Policy Isolation:** Push dedicated isolation policies for the AI service through the SD-WAN controller, restricting access to internal core data sources
Model / API Abuse	Abnormal calls / burst requests	SD-WAN behavioral baselines based on traffic frequency and packet-length features, combined with identity analysis (User-ID)	**QoS Rate Limiting & Blocking:** Use SD-WAN QoS policies to limit API request rates; directly block anomalous IPs through **CPE ACLs**
Network Attacks	DDoS / lateral movement	Correlate network and security telemetry (**NetFlow + security logs**) to detect abnormal lateral scanning	**Automated Scrubbing Response:** Coordinate the SD-WAN controller with the scrubbing center, using **BGP Flow Spec** or **PBR** to redirect malicious traffic for mitigation

Industry Direction: Network, SASE, AI Security, and AIOps converge on a unified context and execution plane, enabling coordinated NOC–SOC operations.

SD-WAN + AI Industry Applications: From Connectivity Infrastructure to Business Production Systems

■ **SD-WAN + AI is evolving into an intelligent execution plane for business production systems. Session-level telemetry, AI reasoning, and PBR/QoS/micro-segmentation embed network capabilities into industrial inspection, financial compliance, and live streaming-shifting from connectivity assurance to business outcomes.**

Industrial Manufacturing: From "Dumb Pipe" to Real-Time Executor for AI Quality Inspection

SD-WAN + AI Decision Layer:

• Identify VISION_QC video features via DPI.
• Monitor 4G, wired, and Wi-Fi link quality in real time.
• Predict link degradation and assess service SLA risk.

SD-WAN + AI Control Layer:

• Trigger PBR to prioritize quality alarms over low-latency paths to MES.
• Switch welding control instructions to the optimal path when links degrade.

Business Outcome: Continuous production-line operation, fewer unplanned outages, and a shift from post-event inspection to real-time intervention.

Financial Services: From Compliance Pipe to Zero Trust and AI Operations Orchestration

SD-WAN + AI Decision Layer:

• Collect network-wide session telemetry: latency, packet loss, and bandwidth.
• Use AI behavior baselines to detect abnormal API calls and off-hours sensitive-data access.

SD-WAN Execution Layer:

• Low risk: adjust QoS for core transaction systems.
• Medium risk: redirect anomalous traffic to an audit gateway via PBR.
• High risk: coordinate with SASE for micro-segmentation.
• Keep core data local; upload only desensitized indicators.

Business Outcome: Central training plus edge inference; core data remains local to meet financial compliance requirements.

Live Streaming / Media: From Best Effort to Deterministic Experience Assurance

SD-WAN + AI Decision Layer:

• Combine video features - bitrate, frame rate, GOP - with network metrics.
• Predict link degradation and trigger service-experience alerts in advance.

SD-WAN Execution Layer:

• Aggregate 5G, satellite, and wired links with technologies such as SpeedFusion.
• Prioritize live-streaming traffic through application-aware QoS.
• Execute seamless switching to support low-latency, low-loss delivery.

Business Outcome: Live streaming evolves from best-effort Internet delivery to business-grade deterministic experience assurance.

Industrial Manufacturing Case: Nestlé — Global Production Network Evolving Toward Observable, Predictable, and AI-Ready Operations

CAICT 中国信通院

Nestlé Connected Every Corner of Global Operations

1700+
Production / Office Sites
Global Operations Coverage

35–40%
Application Performance Improvement
Public case estimate, driven by lower latency

99.97%
Network Availability
Public case metric

80%
Reduction in Support Tickets / Major Incidents
Public statement in recent two years

Case Overview

- Enterprise network spans factories, warehouses, offices, data centers, and four public cloud environments.
- Uses Cisco SD-WAN multi-region Fabric, ThousandEyes, and Secure Routers to unify visibility, routing, and security.
- Smart factories connect sensors, IIoT, and OT, using real-time monitoring, analytics, and proactive alerts to reduce downtime.

Innovation Highlights

Network path optimization + end-to-end digital experience observability + AI-ready factory data loop

SD-WAN + AI Industry Mapping

Field Devices / OT	Multimodal Data	Edge Awareness
WAN Fabric	Application-Aware Routing	Task / Experience SLA
ThousandEyes	End-to-End Telemetry	Prediction & Root Cause
Secure Router	Security & Policy	AI-Ready Protection

Sources: Cisco, “Connecting Every Corner of Nestlé’s Global Operations”, published 2026. URL: https://www-cloud.cisco.com/site/us/en/about/case-studies-customer-stories/nestle.html

Financial Services Case: Generali China Life and Ping An Technology

CAICT 中国信通院

Generali China Life &Ping An Technology

Branch-Cloud connectivity and AI customer Service at Scale

100+
sites connected

99.99%
Network availability

200
Branches Delivered

2h→10min
Policy Issuance Time
Branches Delivered

Case Overview

- Enterprise network spans factories, warehouses, offices, data centers, and four public cloud environments.
- Uses Cisco SD-WAN multi-region Fabric, ThousandEyes, and Secure Routers to unify visibility, routing, and security.
- Smart factories connect sensors, IIoT, and OT, using real-time monitoring, analytics, and proactive alerts to reduce downtime.

1	**Business:**	secure branch-cloud connectivity + AI customer-service assurance
2	**Network:**	application-aware steering + centralized orchestration
3	**Security:**	financial-grade secure connectivity and compliance
4	**AI evolution:**	network modeling + real-time monitoring + predictive optimization
5	**Core value:**	from connectivity infrastructure to an intelligent execution layer

Sources: CITIC Telecom CPC / China Entercom public case on Generali China Life SD-WAN optimization; Huawei Enterprise public case on Ping An Fast AI with Huawei SD-WAN Solution.

SD-WAN + AI Evolution Path: From Traffic Networks to Task Networks and Service Networks

CAICT 中国信通院

■ **SD-WAN is becoming an intelligent intermediate layer that helps AI services understand intent, select resources, and assure outcomes.**

Phase 1

IP Traffic Scheduling

The smallest business unit of future network policies may no longer be only a five-tuple or application, but an AI task + inference phase + resource constraints.

Phase 2

Application Awareness

Application Experience & AIOps

Core KPI **Application SLA**

Industry Status **Mainstream**

Phase 3

AI Task Awareness

Inference Stage	What Happens on the Network	Key KPI	Network Focus	AI-Native Action
Prompt Ingress	Authentication / API / Routing	TTFT	Low latency, nearest access	Prefer the shortest / least-jitter path
Prefill	Large-input context processing	TTFT	Bandwidth, compute queue	Compute-near access + capacity coordination
Decode	Token-by-token generation	ITL / TPS	Stability, low jitter	Avoid path jitter / congestion
Agent Chain	Multi-round model / tool invocation	Task completion latency	End-to-end continuity	Cross-domain path + cache / RAG coordination
Response Delivery	Streaming output	User experience	Stability, low packet loss	Experience-aware policy

Core KPI **TTFT/ITL/TPS**

Industry Status **Accelerating Adoption**

SD-WAN will evolve from choosing a better path to selecting the optimal network + computing combination for an end-to-end inference task.

Phase 4

Computing-Network Coordination

Joint Scheduling of Network + Computing + Security

Core KPI **Business Outcome SLA**

Industry Status

Operators, cloud providers, and SD-WAN vendors can turn "Token Experience" into a new acceleration service and SLA entry point by incorporating TTFT, ITL, and TPS into network SLAs and building a closed loop from business outcomes to network actions.

Phase 5

Computing-Network Coordination

Agentic orchestration + NaaS

Core KPI **Outcome-/Consumption-Based Billing**

Industry Status **Frontier**

Essence of Evolution: scheduling objects evolve from packets/flows → applications → tasks → resource combinations → business outcomes.

CAICT 中国信通院

Thanks!