AI Edge

How traffic patterns change in the age of AI

Johan Ervenius
Systems Engineer, Arista Networks

AI Acronyms

Training

Inference

LLM

Generative AI

Agentic AI

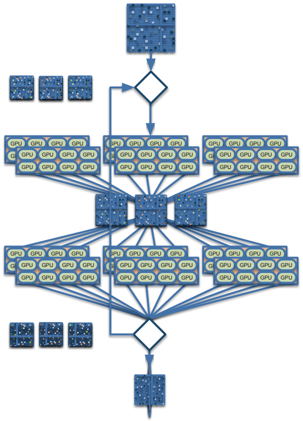
NLP

Neural Networks

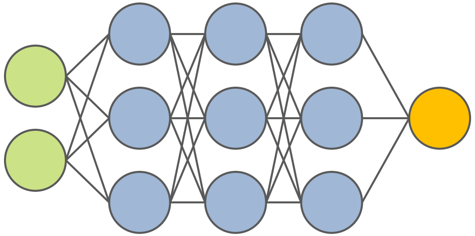
Computer Vision

AI Training vs Trained

Training

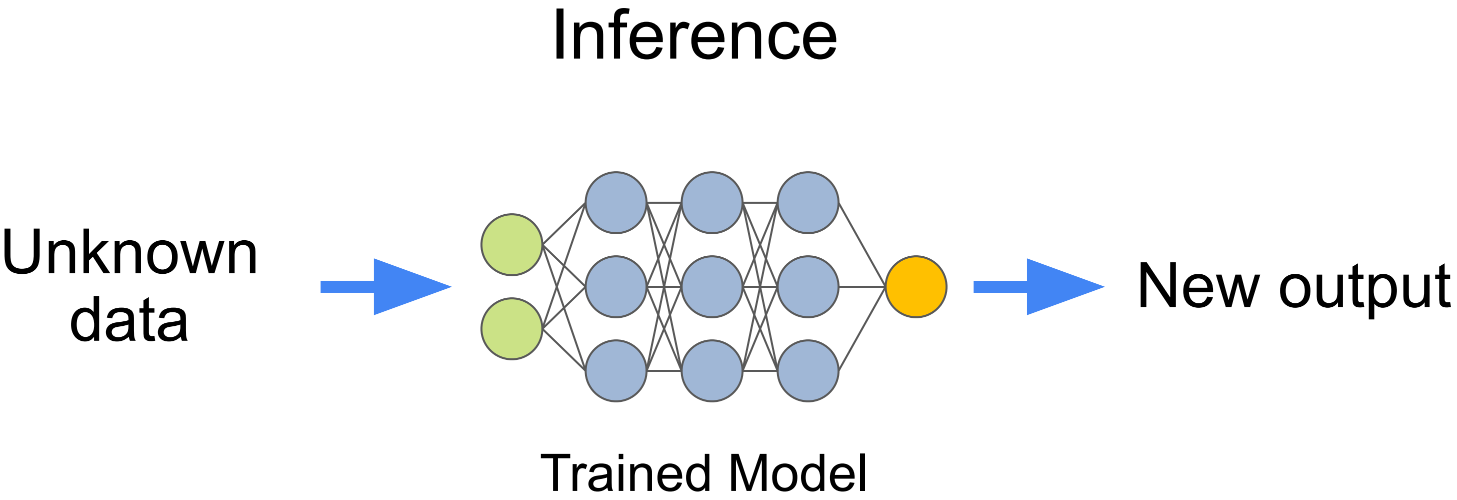


Trained
Model



Within the Data Center

Use of the Trained model



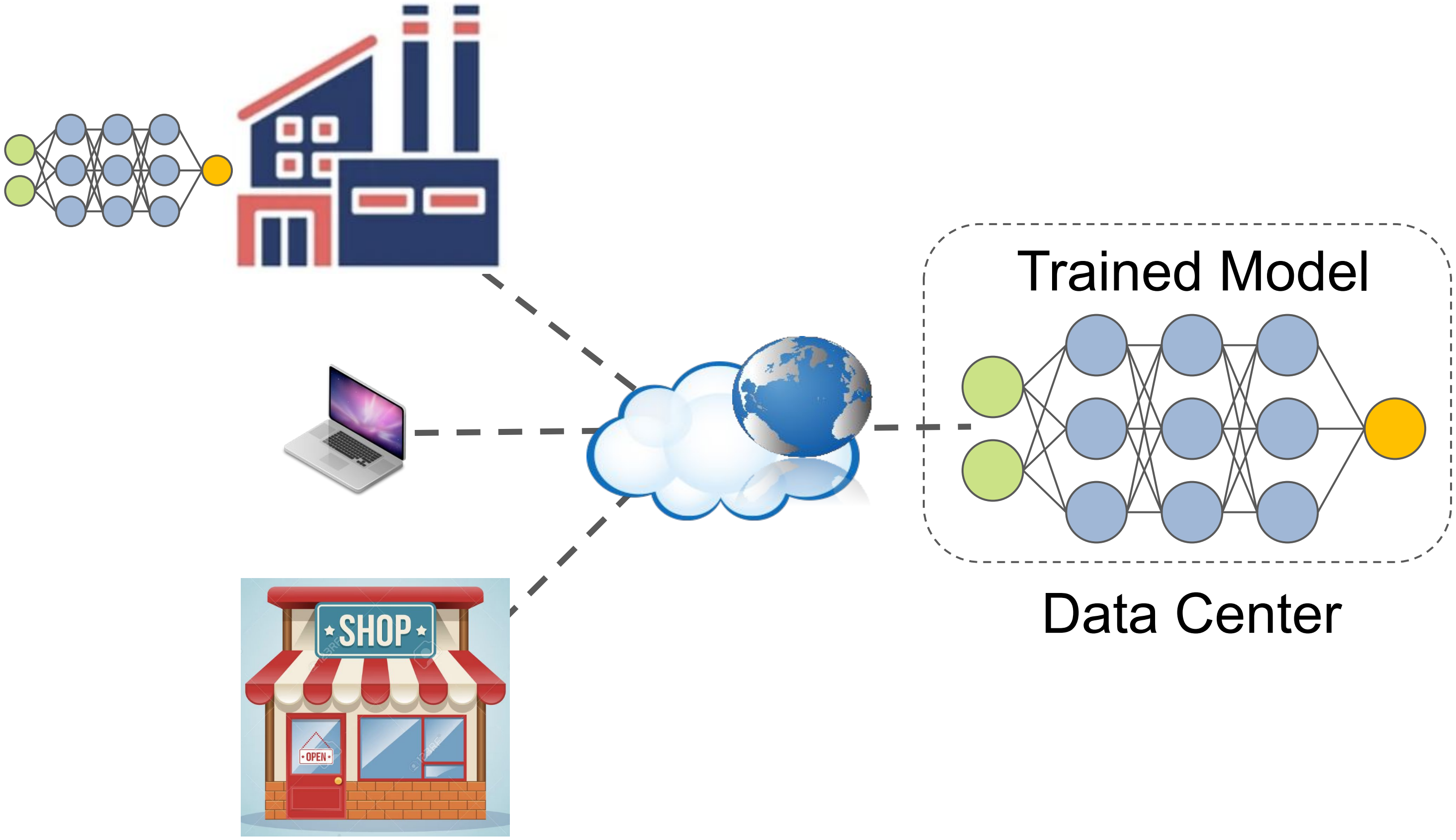
Inference

Unknown data

New output

Trained Model

Online Inference



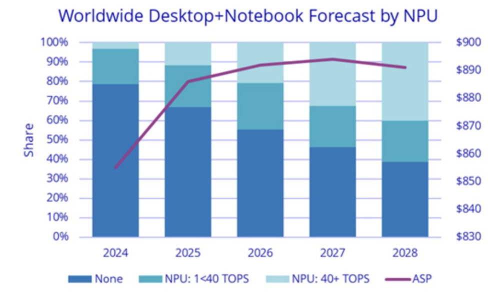
Trained Model

Data Center

NPUs Everywhere Enabling Pervasive AI At The Edge

Laptop/PC Edge

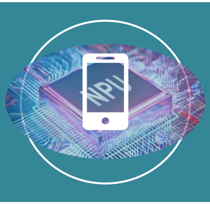
IDC: "…we see AI PCs (with NPUs) ramping up to **nearly two out of every three shipped in 2028**."



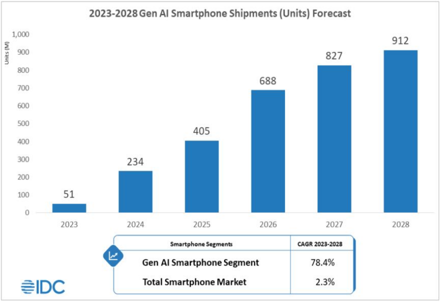
Source: IDC 2024

Increased Bandwidth & Power

Mobile Edge



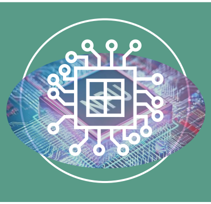
By 2028, worldwide Gen AI smartphone shipments will reach 912 million units, resulting in a compound annual growth rate (CAGR) of 78.4% for 2023-2028.



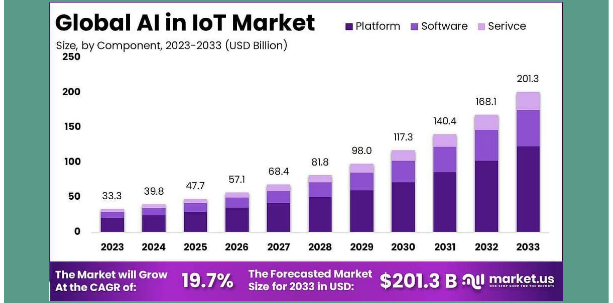
Smartphone Segments	CAGR 2023-2028
Gen AI Smartphone Segment	78.4%
Total Smartphone Market	2.3%

AI Application Performance at Edge

IoT Edge



By 2028, worldwide Gen AI smartphone shipments will reach 912 million units, resulting in a compound annual growth rate (CAGR) of 78.4% for 2023-2028.



Increased exposure to perimeter breaches / data loss

ARISTA

AI in Retail

AI Applications

Customer Experience

Video analytics

Inventory Optimization

Network Requirements

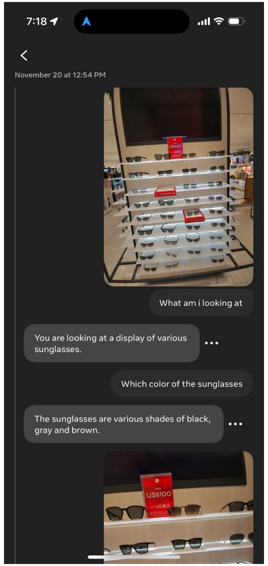
Traffic Asymmetry
10:1 upload to download

Increased Bandwidth Requirements

Latency Sensitive

Bursty Traffic

AI Traffic Is Now Driving New Requirements



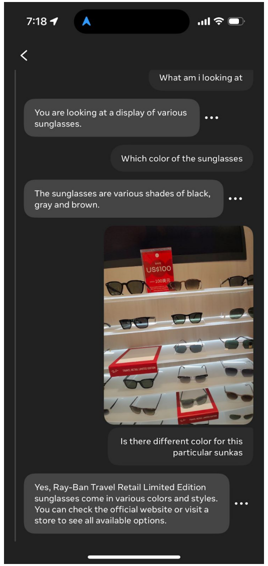
November 20 at 12:54 PM

What am i looking at

You are looking at a display of various sunglasses.

Which color of the sunglasses

The sunglasses are various shades of black, gray and brown.



What am i looking at

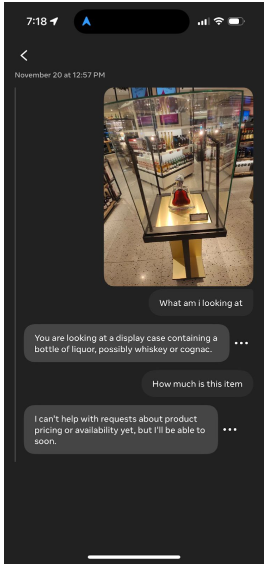
You are looking at a display of various sunglasses.

Which color of the sunglasses

The sunglasses are various shades of black, gray and brown.

Is there different color for this particular sunkas

Yes, Ray-Ban Travel Retail Limited Edition sunglasses come in various colors and styles. You can check the official website or visit a store to see all available options.



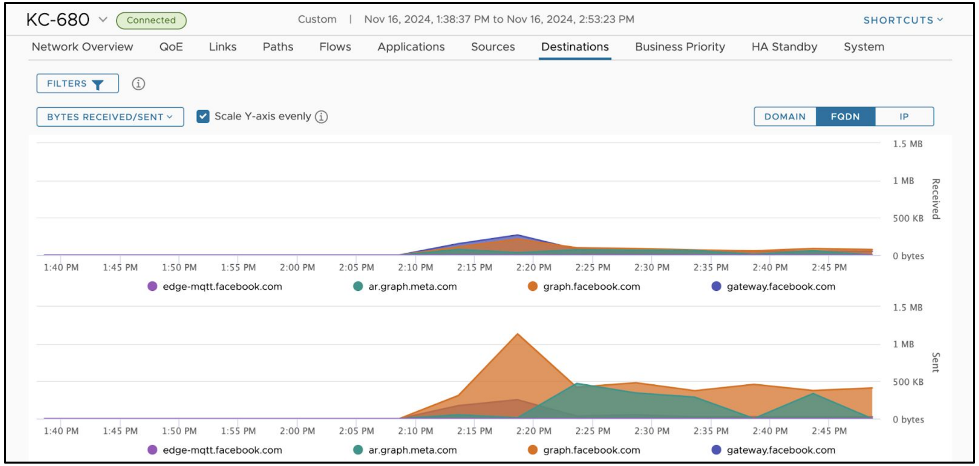
November 20 at 12:57 PM

What am i looking at

You are looking at a display case containing a bottle of liquor, possibly whiskey or cognac.

How much is this item

I can't help with requests about product pricing or availability yet, but I'll be able to soon.



KC-680 Connected

Custom | Nov 16, 2024, 1:38:37 PM to Nov 16, 2024, 2:53:23 PM

BYTES RECEIVED/SENT

☑ Scale Y-axis evenly

FQDN

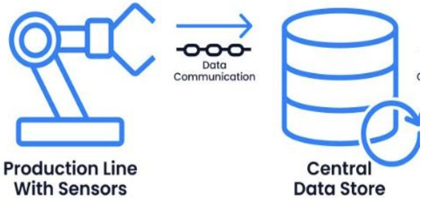
AI in Manufacturing

AI Applications

AI analyzes sensor data from machinery to predict equipment failures

AI algorithms identify defects in real-time

AI predicts demand, optimizes inventory levels



Network Requirements

Traffic Asymmetry
Increased upload

Increased Bandwidth Requirements

Latency Sensitive

Bursty Traffic

AI in Healtcare

AI Applications

Robotic Surgery

Remote patient monitoring

Supporting with diagnostics

Clinical support

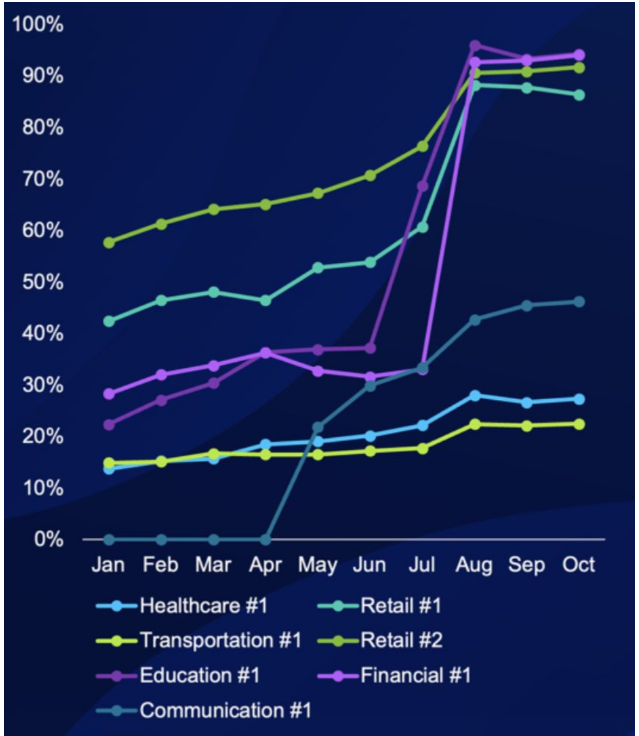
Network Requirements

Traffic Asymmetry Increased upload to download

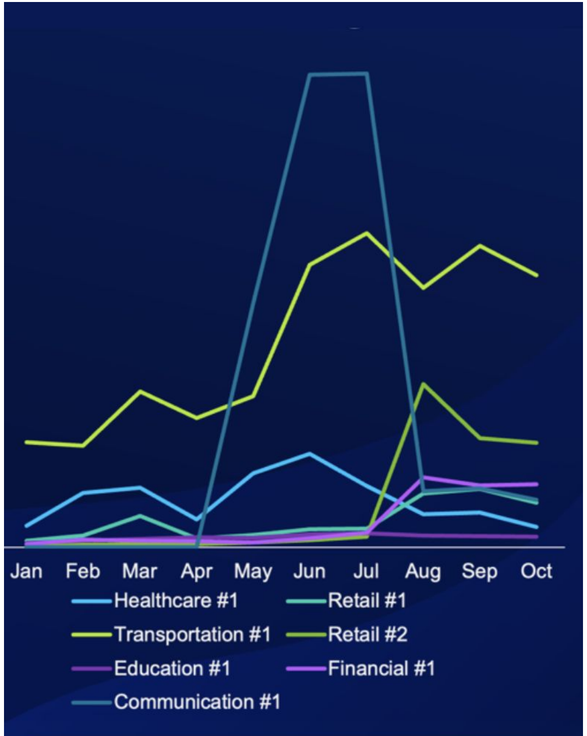
High Bandwidth Requirements

Very Latency Sensitive

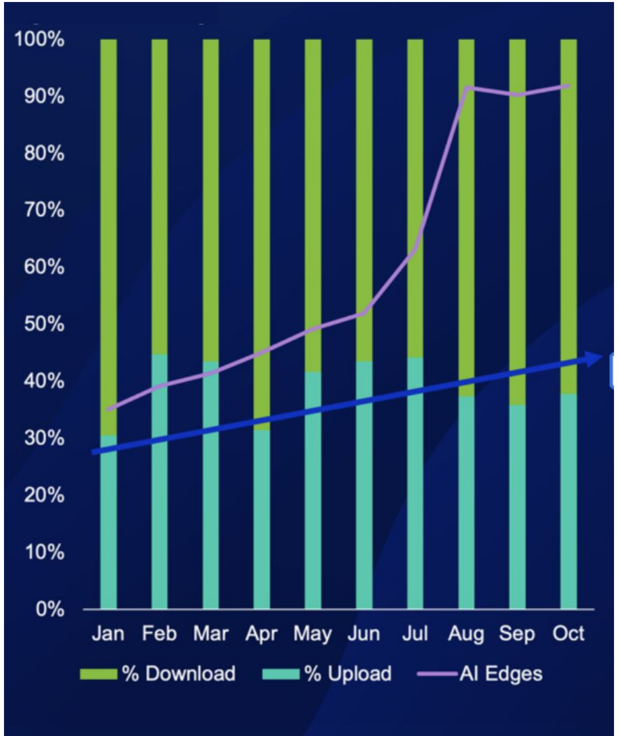
Generative AI Consumption has Increased Significantly



%Edge with GenAI Traffic



GenAI volume/site trending up



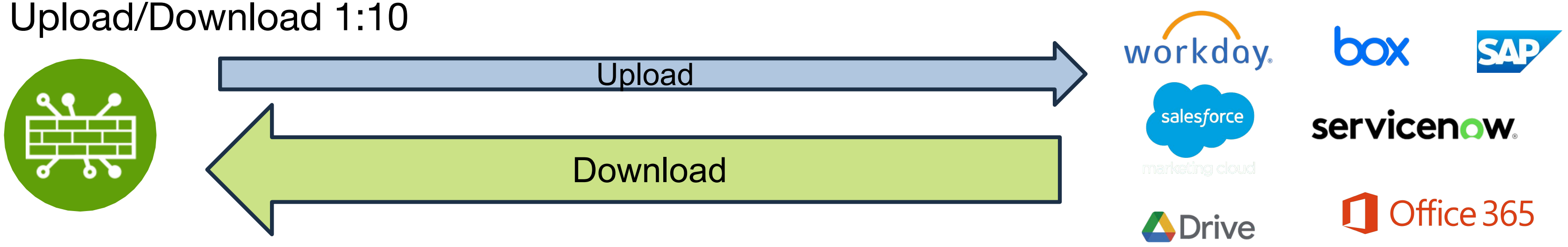
Increase in upload traffic %

ARISTA

AI Network Bandwidth Impact

Before

Upload/Download 1:10

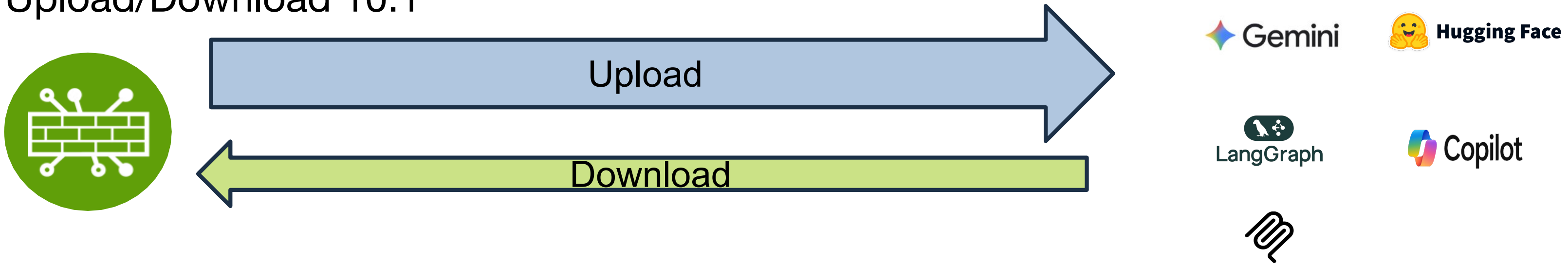


Upload

Download

2026

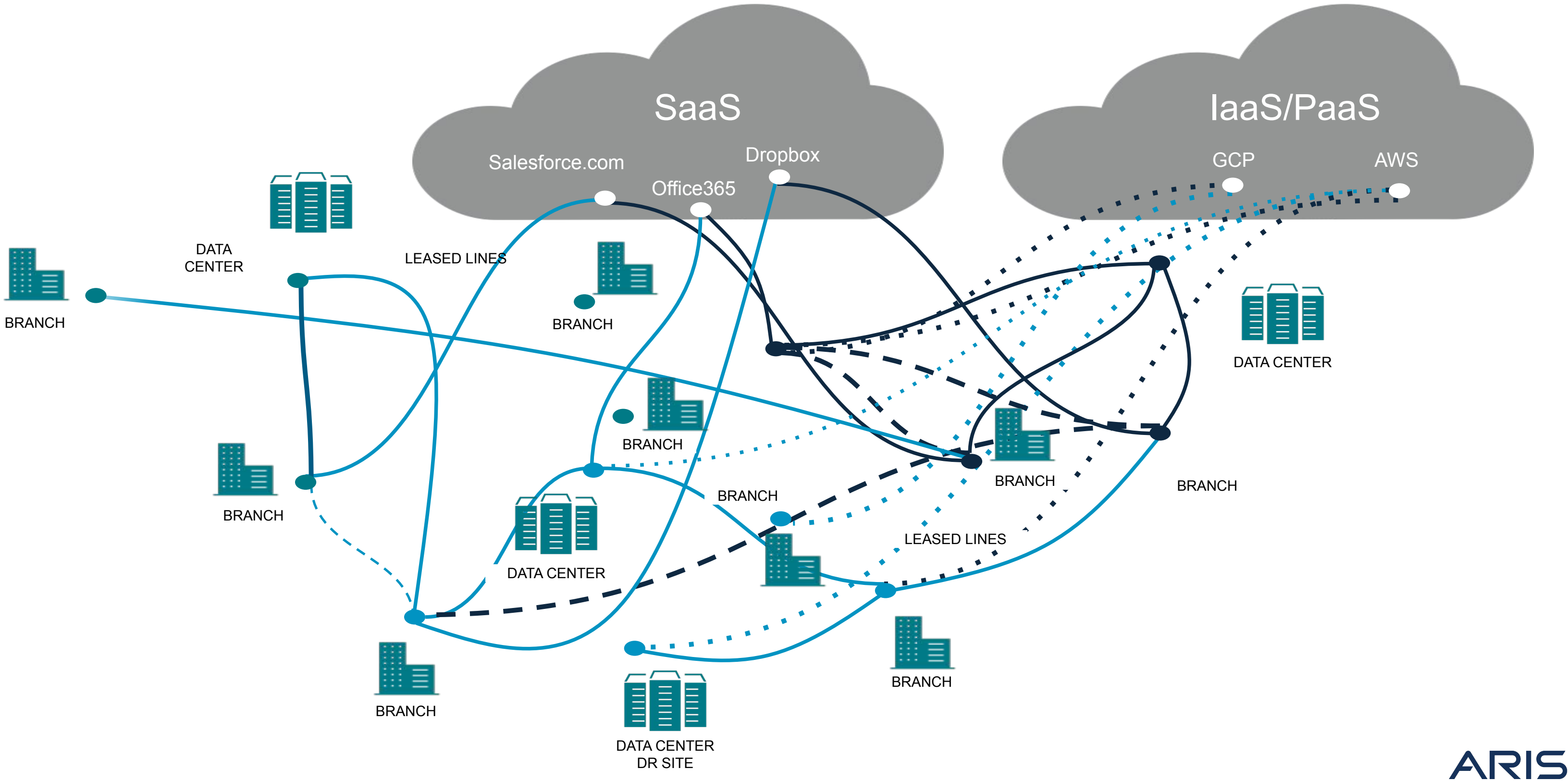
Upload/Download 10:1



Upload

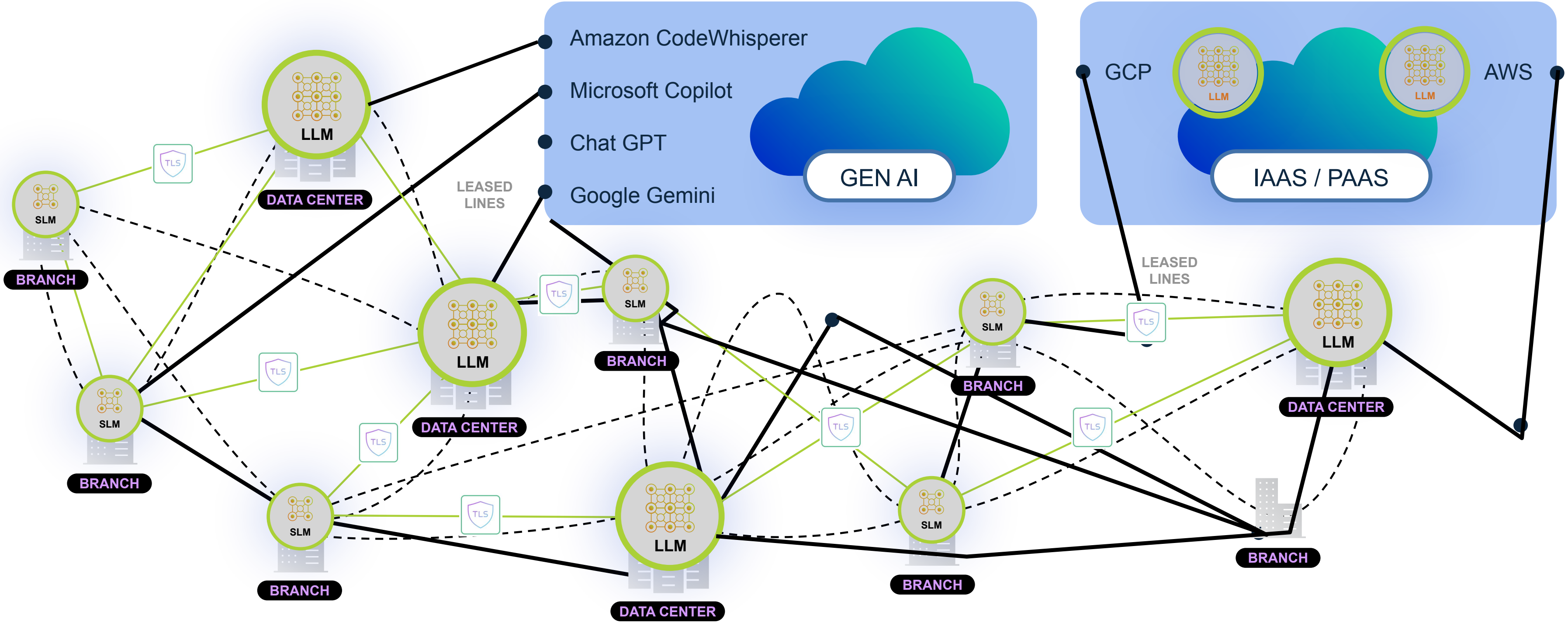
Download

Enterprise WAN is getting increasingly complex



ARISTA

Agentic AI interactions are on the rise



ARISTA

AI Edge Application Optimization

What is on the network?

Deep Application Recognition

What paths are available?

Secure Overlay

How are the paths performing?

Link Qualification

What is the best path for the application?

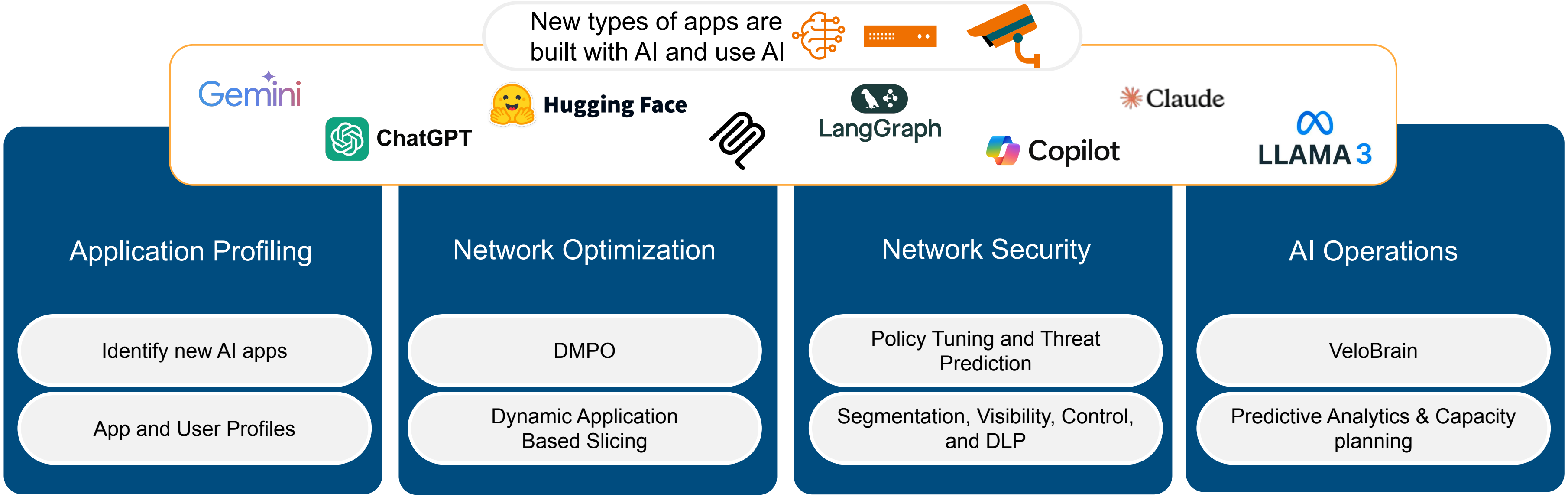
Application Steering

Reacting to changed link conditions

On-Demand Remediation & Aggregation

Deliver an AI-Ready SD-WAN Architecture

You need a networking architecture designed to optimize performance, security, and scalability for distributed AI workloads at the edge - VeloRAIN from Arista



New types of apps are built with AI and use AI

Application Profiling

- Identify new AI apps
- App and User Profiles

Network Optimization

- DMPO
- Dynamic Application Based Slicing

Network Security

- Policy Tuning and Threat Prediction
- Segmentation, Visibility, Control, and DLP

AI Operations

- VeloBrain
- Predictive Analytics & Capacity planning

Thank you!