



Atlas: Opening the door to AI with massive computing power

With constant breakthroughs in core technology, artificial intelligence (AI) is already integrated into work and daily life. For example, news and shopping apps now push recommendations for things that they think we want to see or buy. Mapping apps help plan time-saving routes based on our preferences, and even the voices used for navigation apps are synthesized by AI. The basis of all these advances is powerful computing power. But, it's only when AI computing power is affordable, effective, and reliable that these applications can be supported.

By Gao Fan

Use cases for AI

AI will trigger an industrial revolution that will impact all industries, reconstructing industry and enterprises with an entirely new model. AI particularly suits the

following three types of scenarios:

High-volume, repetitive scenarios:

For example, recognizing images and reviewing documents. AI suits highly repetitive tasks with clearly defined goals and huge amounts of data. In one use case in a city transportation

department, only 13 percent of more than 30,000 traffic violation photos taken a day could be identified by staff, because of time and stamina limits. After using an AI traffic brain, enforcement rates significantly improved, and traffic violations were reduced by over 60 percent.



Specialist experience scenarios: Many industries suffer from a lack of experts, including the healthcare sector. AI-assisted medical image analysis and screening can reduce dependence on specialists. And in manufacturing, ordinary designers can harness AI-assisted design to complete specialist design tasks.

Multi-domain coordination scenarios: These are the most complex scenarios and include urban smart transportation systems and modern manufacturing. Traffic light signal control, for example, involves multiple variables and requires fast, accurate analysis and decision-making, which is difficult for the human brain to do.

Massive computing power

Making AI affordable, effective, and reliable will first require sufficiently powerful computing power. Second, we will need to deploy cloud, edge, and devices, and ensure efficient coordination based on scenario. In recent years, we've seen AI applications continue to grow in breadth and depth, from safe cities to smart retail. But, AI computing power powered by centralized processing in data centers can no longer satisfy the requirements of new high-bandwidth, low-latency scenarios.

Huawei is accelerating its research in AI, and continues to embrace the data processing

challenges of the AI era with a full-stack AI solution. A key part of Huawei's full-stack AI solution, the Atlas intelligent computing platform, is designed to provide computing power for AI products. It harnesses key technologies such as intelligent heterogeneity, device-edge-cloud hierarchical deployment, and edge-cloud synergy to enhance computing power for AI.

To meet new application scenarios, the Atlas computing platform also comes in various product forms, including a module, AI accelerator card, AI edge station, and appliance. These are all built with Huawei's Ascend chipsets and major heterogeneous computing components, and form a complete AI solution.

Huawei's Atlas intelligent computing platform is formed of the Atlas 200 AI accelerator module for devices, the Atlas 300 AI accelerator card for data centers, the Atlas 500 AI edge station for the network edge, and a one-stop AI platform, the Atlas 800 AI appliance, positioned for enterprise.

Atlas 200 AI accelerator module: Just half the size of a credit card, the Atlas 200 consumes only 10 watts of power, supports 16-channel real-time HD video analytics, and can be deployed in devices such as cameras and drones.

Atlas 300 AI accelerator card: With a HHL

The various Atlas intelligent computing platform products suit different application scenarios, enabling hierarchical deployment to cloud, edge, and device.

PCIe standard card form factor, the Atlas 300 is designed for data center and edge servers. It supports multiple data precisions and delivers up to 64 TOPS of INT8 on a single card, a threefold lead over competing products, and provides superior computing power for deep learning and inference.

Atlas 500 AI edge station: This industry-leading edge product integrates AI processing capabilities and supports 16-channel HD video processing in the size of a set-top box (STB), delivering a fourfold performance increase over existing products. The AI edge station suits a broad range of applications, including transportation, healthcare, unattended retail, and smart manufacturing.

Atlas 800 AI appliance: Based on the standard framework and programming environment, this product provides an optimized AI environment and pre-installed underlying software library. It's ready to work in less than two hours out of the box. In addition, the AI appliance integrates management software for cluster management and task scheduling, and system-level performance monitoring capabilities, greatly reducing the entry requirements for enterprise AI application.

The various Atlas intelligent computing platform products suit different application scenarios, enabling hierarchical deployment to cloud, edge, and device. In the case of traffic light control systems, current systems aren't flexible enough to automatically adjust traffic light cycles based on intersection traffic flow. In a joint innovation project, Huawei and a research institute developed traffic monitoring algorithms and signal adjustment algorithms

respectively, and deployed Atlas 500 smart stations at intersections to adjust traffic light cycles in real time. The project involved:

Edge deployment: Three-to-four traffic lights were connected to each small station, which were installed in street cabinets or poles close to each intersection.

Real-time offline processing: Real-time offline adjustment of traffic light cycles through edge stations greatly improved processing efficiency.

Edge-cloud synergy: The cloud handles unified management of the edge station hardware, pushing algorithms and firmware upgrades and monitoring their status. Structured data processed by edge stations is returned to the cloud, with the cloud controlling traffic light switching over a broader area.

AI technology is constantly progressing, and the number of AI-driven industry solutions is also growing. As we stride toward a fully connected, intelligent world, innovation is essential and computing power will create new possibilities. Huawei's intelligent Atlas platform provides enhanced computing power to help customers integrate AI capabilities into all business processes and bring the computing power required by AI from the data center to the network edge and devices. Atlas also enables connectivity and collaboration to support enterprise intelligence in more application scenarios. As we strive to drive AI adoption and applications, Atlas will continue to make AI even faster, facilitating the advancement of industry verticals. 