1. Visão geral da Inteligência Artificial

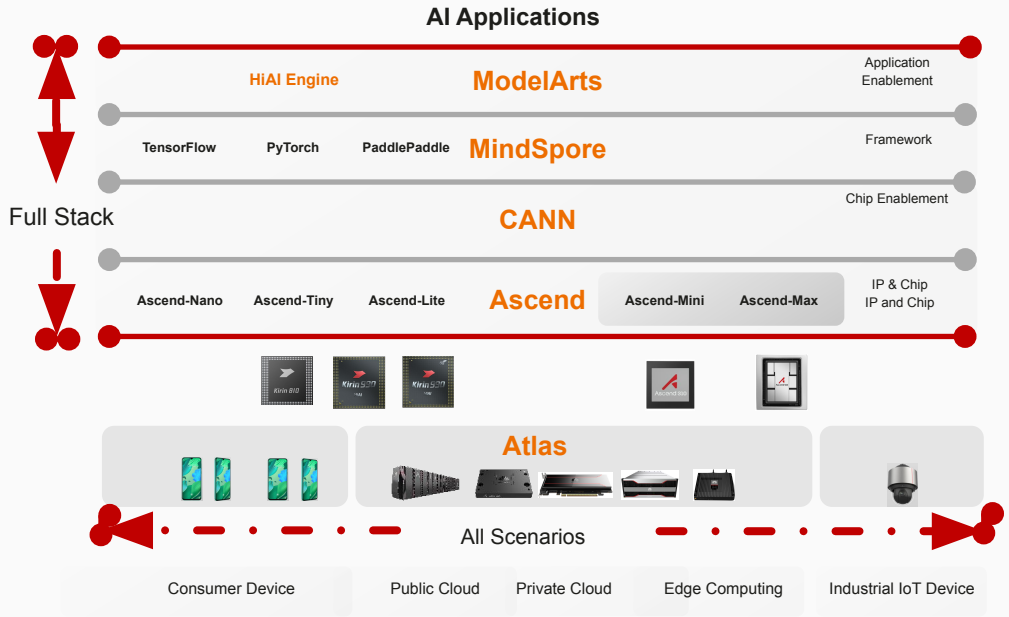
1.3 - Estratégias da Huawei para desenvolvimento em Inteligência Artificial e disputas envolvendo IA

Marciel Barros

Julho, 2020

1. AI Overview
2. Technical Fields and Application Fields of AI
3. **Huawei's AI Development Strategy**
4. AI Disputes
5. Future Prospects of AI

Huawei's Full-Stack, All-Scenario AI Portfolio



Application enablement: provides end-to-end services (ModelArts), layered APIs, and pre-integrated solutions.

MindSpore: supports the unified training and inference framework that is independent of the device, edge, and cloud.

CANN: a chip operator library and highly automated operator development tool.

Ascend: provides a series of NPU IPs and chips based on a unified, scalable architecture.

Atlas: enables an all-scenario AI infrastructure solution that is oriented to the device, edge, and cloud based on the Ascend series AI processors and various product forms.

Huawei's "all AI scenarios" indicate different deployment scenarios for AI, including public clouds, private clouds, edge computing in all forms, industrial IoT devices, and consumer devices.

Full Stack — ModelArts Full-Cycle AI Workflow

EI Intelligent Twins

EI Cognition Service AI Service

Data

AI data framework

Efficient filtering and semi-automated labeling, data preprocessing
Efficiency improved by 100 times

Algorithm development

Out-of-the-box development environment compatible with mainstream frameworks
MoXing library, simplifying model development
Built-in model algorithms, improving development efficiency

Training

Distributed training, shortening training period **from weeks to minutes**

Wizard-based AutoLearning, **code-free development, enabling model training from scratch**

Deployment

One-click deployment on device, edge, and cloud
All-scenario deployment
Inference on the Ascend AI processor

Market

AI sharing platform helps enterprises build internal and external **AI ecosystems**

AI applications

Visualized Workflow Management

Version management, **traceable** and worry-free development

ModelArts

AI data framework accelerates data processing by 100 folds.

Visualized workflow management makes development worry-free.

Distributed training shortens training from weeks to minutes.

One-click deployment on device, edge, and cloud supports various deployment scenarios.

Automatic learning enables you to start from scratch.

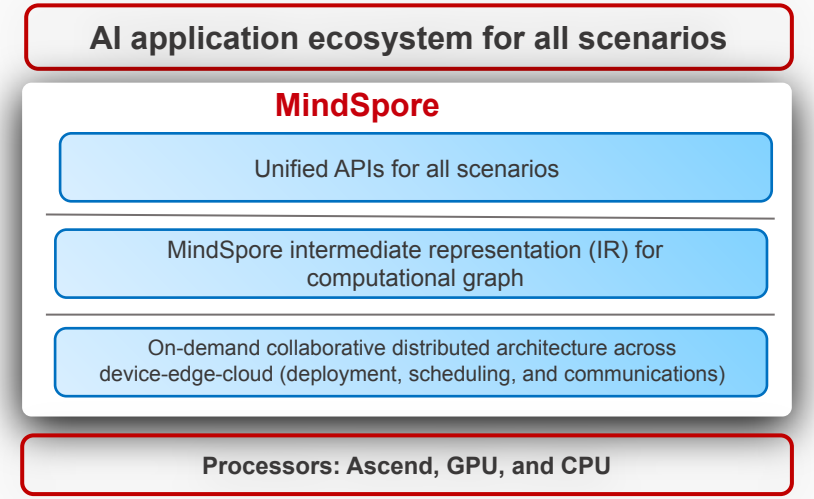
AI sharing platform builds internal and external AI ecosystems for enterprises.

HUAWEI

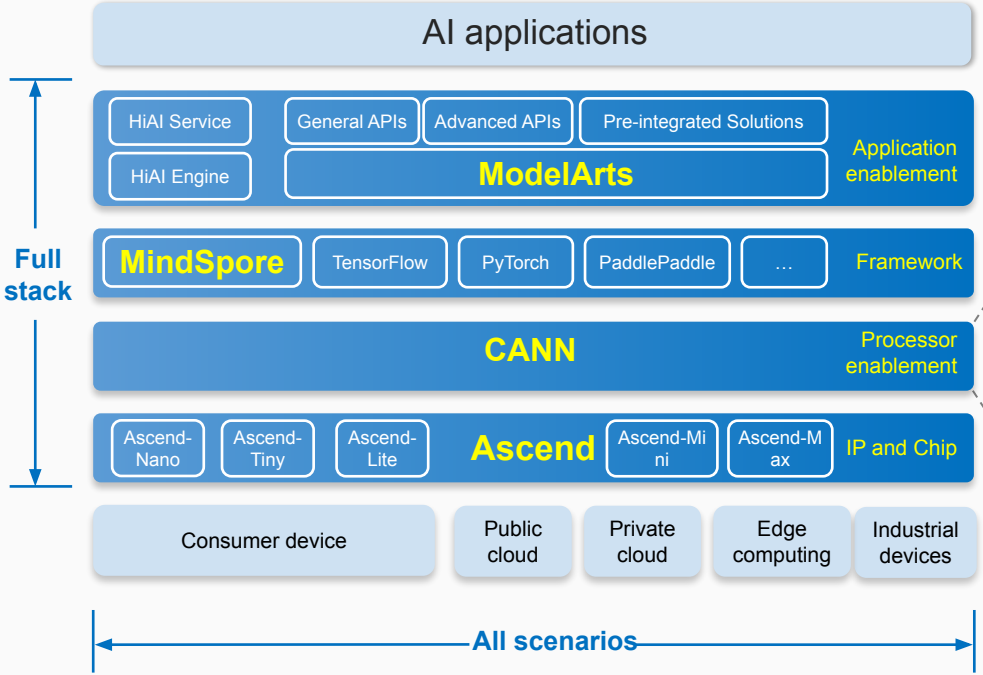
Full Stack — MindSpore (Huawei AI Computing Framework)

- MindSpore provides **automatic parallel capabilities**. With MindSpore, senior algorithm engineers and data scientists who focus on data modeling and problem solving can run algorithms on dozens or even thousands of AI computing nodes with only a few lines of description.
- The MindSpore framework supports both large-scale and small-scale deployment, adapting to independent deployment in all scenarios. In addition to the **Ascend AI** processors, MindSpore also supports other processors such as **GPUs** and CPUs.

MindSpore will be open-source on March 30, 2020



Full Stack — CANN

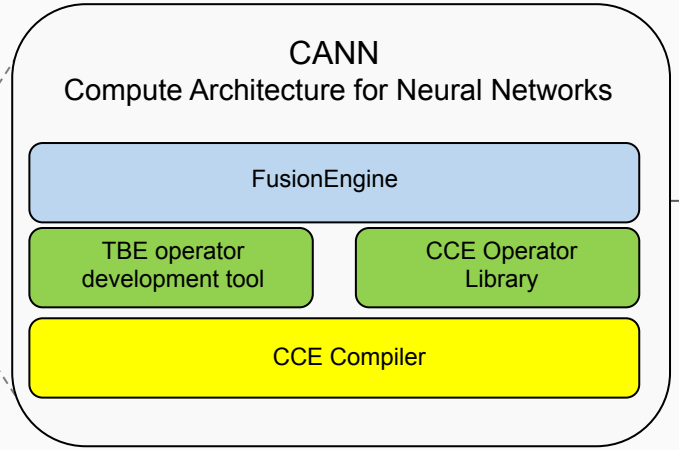


CANN:

A chip operators library and highly automated operator development toolkit

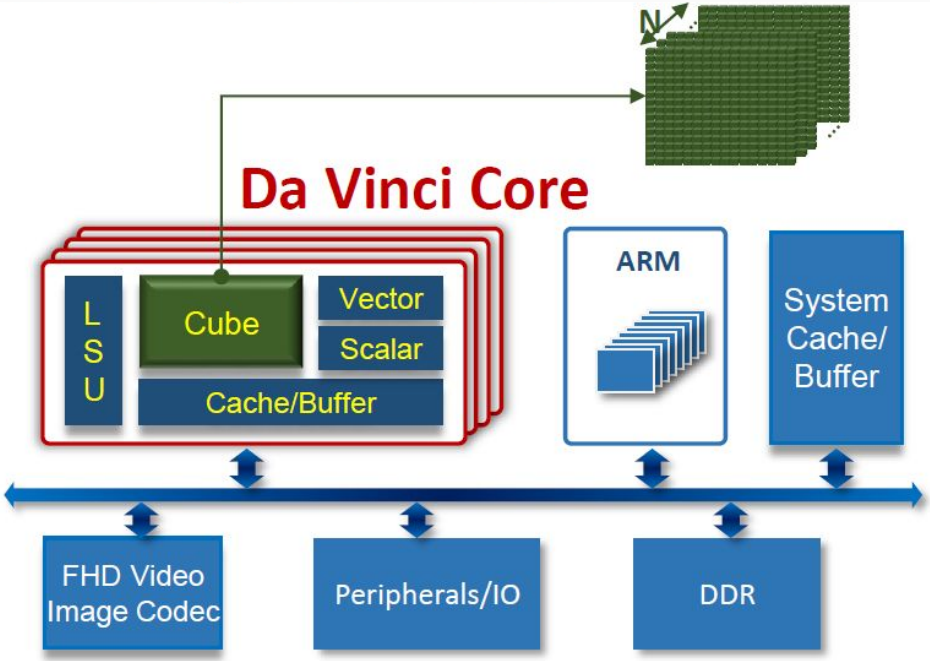
Optimal development efficiency, in-depth optimization of the common operator library, and abundant APIs

Operator convergence, best matching the performance of the Ascend chip



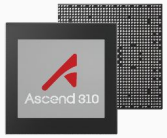
Full Stack — Ascend 310 AI Processor and Da Vinci Core

SPECIFICATIONS	Description
Architecture	AI co-processor
Performance	Up to 8T @FP16
	Up to 16T@INT8
Codec	16 Channel Decoder – H.264/265 1080P30 1 Channel Encoder
Memory Controller	LPDDR4X
Memory Bandwidth	2*64bit @3733MT/S
System Interface	PCIe3.0 /USB 3.0/GE
Package	15mm*15mm
Max Power	8Tops@4W, 16Tops@8W
Process	12nm FFC



Note: This is typical configuration, high performance and low power sku can be offered based on your requirement.

Ascend AI Processors: Infusing Superior Intelligence for Computing



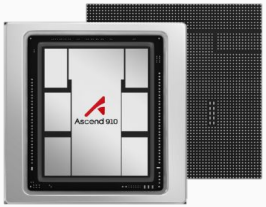
Ascend 310

AI SoC with ultimate energy efficiency

Ascend-Mini
Architecture: Da Vinci

Half-precision (FP16): 8 TFLOPS
Integer precision (INT8): 16 TOPS
16-channel full-HD video decoder: H.264/265
1-channel full-HD video encoder: H.264/265

Max. power: 8 W



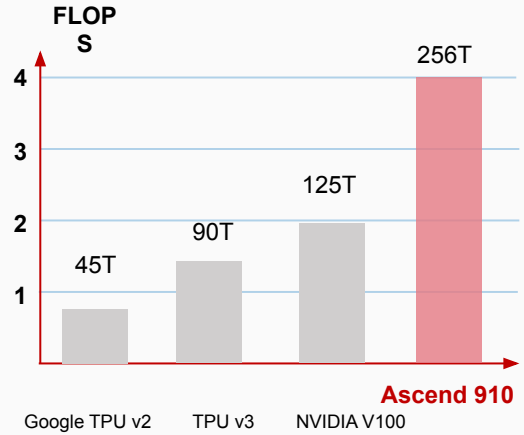
Ascend 910

Most powerful AI processor

Ascend-Max
Architecture: Da Vinci

Half-precision (FP16): 256 TFLOPS
Integer precision (INT8): 512 TOPS
128-channel full HD video decoder: H.264/265

Max. power: 310 W



HUAWEI

Atlas AI Computing Platform Portfolio

Internet, security, finance, transportation, power, etc.

Application Enablement

Atlas intelligent edge platform
Industry SDK/Container engine/Basic service repository

Atlas deep learning platform
Cluster management/Model management/Data pre-processing

Framework

MindSpore

TensorFlow/PyTorch/Caffe/MxNet

Framework Adapret

CANN

AscendCL

Graph engine for graph optimization

Operator/Acceleration/Communication libraries (BLAS, FFT, DNN, Rand, Solver, Sparse, HCCL)

Runtime

Driver

Common components

Unified O&M and configuration management subsystem

Safety subsystem

AXE toolchain (log/profiling/Mind Studio)

Chips & Hardware

Atlas 200
16 TOPS INT8

Atlas 200 developer kit

Atlas 300 inference accelerator card
64 TOPS INT8

Atlas 300 training card
256 TFLOPS FP16

Atlas 500
16 TOPS INT8

Atlas 800 AI inference server
512 TOPS INT8

Atlas 800 AI training server
2 PFLOPS FP16

Atlas 900
256–1024 PFLOPS FP16

Ascend 310

Da Vinci Architecture

Ascend 910

1. AI Overview
2. Technical Fields and Application Fields of AI
3. Huawei's AI Development Strategy
4. **AI Disputes**
5. Future Prospects of AI

Algorithmic Bias

- **Algorithmic biases are mainly caused by data biases.**
- When we use AI algorithms for decision-making, the algorithms may learn to discriminate an individual based on existing data including race and gender, and therefore create unfair outcomes, such as decisions that are discriminatory based on race, sex or other factors. Even if factors such as race or gender are excluded from the data, the algorithms can make discriminatory decisions based on information of names and addresses.

If we search with a name sounds like an African American, an advertisement for **a tool used to search criminal records** may be displayed. The advertisement, however, is not likely displayed in other cases.

Online advertisers tend to display advertisements of lower-priced goods to **female users**.

Google's image software once mistakenly labeled an image of black people as **"gorilla"**.

Privacy Issues

- The existing AI algorithms are all data-driven. In this case, we need a large amount of data to train models. We enjoy the convenience brought by AI every day while technology companies like Facebook, Google, Amazon, and Alibaba are obtaining an enormous amount of user data, which will reveal various aspects of our lives including politics, religions, and gender.

In principle, technology companies can record each click, each page scrolling, time of viewing any content, and browsing history when users access the Internet.

Technology companies can know our privacy including where are we, where we go, what we have done, education background, consumption capabilities, and personal preferences based on our ride-hailing records and consumption records.

Seeing = Believing?

- With the development of computer vision technologies, reliability of images and videos is decreasing. Fake images can be produced with technologies such as PS and generative adversarial networks (GAN), making it hard to identify whether images are true or not.
- Example:
  - A suspect provided fake evidence by forging an image in which the suspect is in a place where he has never been to or with someone he has never seen using PS technologies.
  - In advertisements for diet pills, people's appearances before and after weight loss can be changed with PS technologies to exaggerate the effect of the pills.
  - Lyrebird, a tool for simulating voice of human beings based on recording samples of minutes, may be used by criminals.
  - Household images released on rent and hotel booking platforms may be generated through GAN.

AI Development = Rising Unemployment?

- Looking back, human beings have always been seeking ways to improve efficiency, that is, obtain more with less resources. We used sharp stones to hunt and collect food more efficiently. We used steam engines to reduce the need for horses. Every step in achieving automation will change our life and work. In the era of AI, what jobs will be replaced by AI?
- The answer is repetitive jobs that involve little creativity and social interaction.

Jobs Most Likely to Be Replaced by AI	Jobs Most Unlikely to Be Replaced by AI
Courier	Writer
Taxi driver	Management personnel
Soldier	Software engineers
Accounting	HR manager
Telesales personnel	Designer
Customer service	Activity planner
...	...

Problems to Be Solved

- Are AI-created works protected by copyright laws?
- Who gives authority to robots?
- What rights shall be authorized to robots?
- ...

Thank You!

Next: 1.5 - Future Prospects of AI