

AI SYSTEMS AND VISUAL INTELLIGENCE AT THE NETWORK EDGE:

TRENDS AND OPPORTUNITIES IN ASIA

Gary Brown, Director of AI Marketing, IoT Group, Intel  @garyisp
US-Asia Technology Management Center, Stanford Univ. October 25, 2018

TODAY WE'LL LOOK FROM SEMICONDUCTOR INDUSTRY POINT OF VIEW— THE CRITICAL INGREDIENTS IN THE EVOLUTION OF **AI AT THE EDGE**

- **BACKGROUND: AI SYSTEMS IN ASIA**
- **DEFINITION OF 'NETWORK EDGE'**
- **TRENDS, MARKETS**
- **ACCELERATING INTELLIGENT VISION AT THE EDGE**
- **CHALLENGES AHEAD**

2018: AI IS BOOMING EVERYWHERE, ESPECIALLY IN ASIA

NOVEMBER 2015 **THIRD ARROW**

Artificial Intelligence—Can Japan Lead the Way?

By Richard Jolley

There are plenty of examples of sci-fi writers imagining products that eventually become commonplace. Think video calling, touch-screen computers, and earbud headphones, the latter imagined by Ray Bradbury in his 1950s classic novel *Fahrenheit 451*.

On the face of it, the sci-fi genre seems to be a good diviner of things to come, as well as being a popular kind in literature. And we can imagine dedicated research and development departments of companies and governmental divisions reading the latest sci-fi in an effort to peer into the future and outmaneuver rivals.

Given recent concerns about the Japanese economy, it wouldn't be surprising if the reading department of its

It's clear the economy needs a boost on industrial production and a consecutive quarter of new

Moreover, Abenomics is struggling in the economy, with the third consecutive

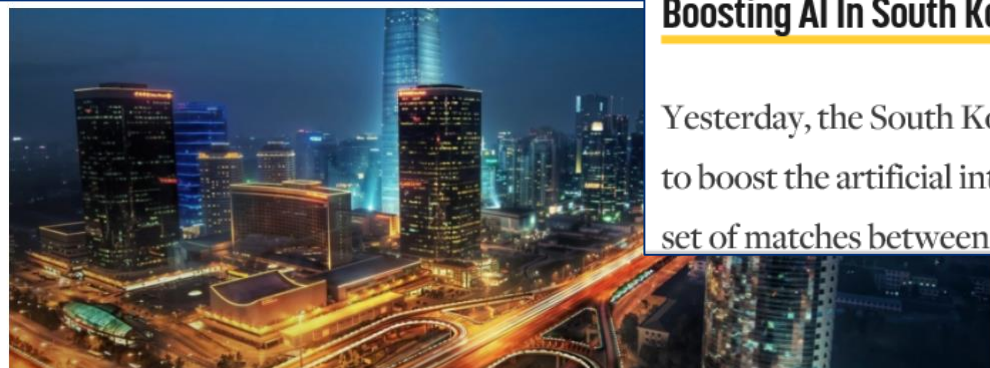
Indeed, a sci-fi ministry is being set up as the government sees tech



Sarah Marquardt March 19th 2016

Boosting AI In South Korea

Yesterday, the South Korean government announced plans to boost the artificial intelligence industry through a series of matches between Lee



Artificial Intelligence in India – Opportunities, Risks, and Future Potential

LAST UPDATED ON SEPTEMBER 16, 2018 BY RAGHAV BHARADWAJ



ARTIFICIAL INTELLIGENCE

South Korean Government Announces Nearly \$1 Billion in AI Funding

The government plans to support the establishment of a high-profile research center that will serve as a catalyst for the national research and development in AI

China announces goal of AI leadership by 2030

July 21, 2017 by Joe McDonald



In this April 26, 2016 photo, a visitor takes a photo of a LeEco LeSEE self-driving electric concept car at the Beijing International Automobile Exhibition in Beijing. China's government announced Thursday, July 21, 2017, a goal of ... more

China's government has announced a goal of becoming a global leader in artificial intelligence in just a decade, putting political muscle behind growing investment by Chinese companies in developing self-driving cars and other advances.

Ad closed by Google

Report this ad

Communist leaders see AI as key to making China an "economic power" in a Cabinet statement on Thursday. It calls for developing skills and resourcing educational resources to achieve "major breakthroughs" by 2025 and

**SOME BACKGROUND IN COMPUTER VISION
AND WHAT WE ARE DOING AT INTEL
TO ENABLE NEW **AI** SYSTEMS AT **THE EDGE****

COMPUTER VISION THEN AND NOW



.1960.

THEN

NOW



- Heuristic Algorithms > > > > ○ AI / Deep Learning
- Human Engineered > > > > ○ Real-time, Higher Accuracy
- Server Based > > > > ○ Edge Based

EDGE-TO-CLOUD: DATA IS KEY DRIVER

By 2019, **45%** of data will be stored, analyzed, and acted on **at the edge**

Transportation • Retail
Public Sector • Logistics • Smart Cities



Video • Healthcare • Manufacturing
Smart Buildings • Energy

DEVICES • THINGS

**EDGE
COMPUTE NODE**

DRIVERS FOR EDGE

**LATENCY
BANDWIDTH
SECURITY
CONNECTIVITY**

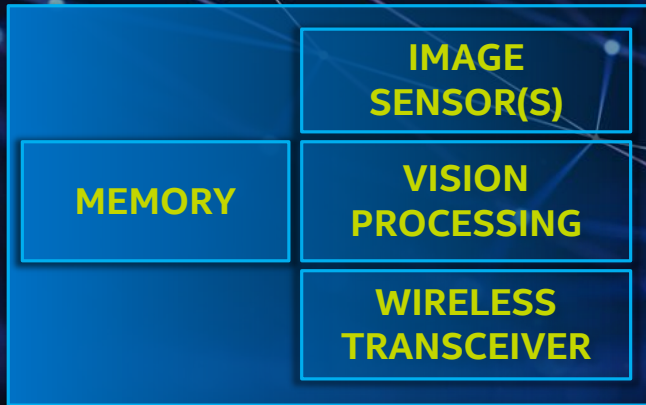
**NETWORK HUB
OR REGIONAL
DATA CENTER**

**CORE
NETWORK**

**CLOUD
DATA CENTER**

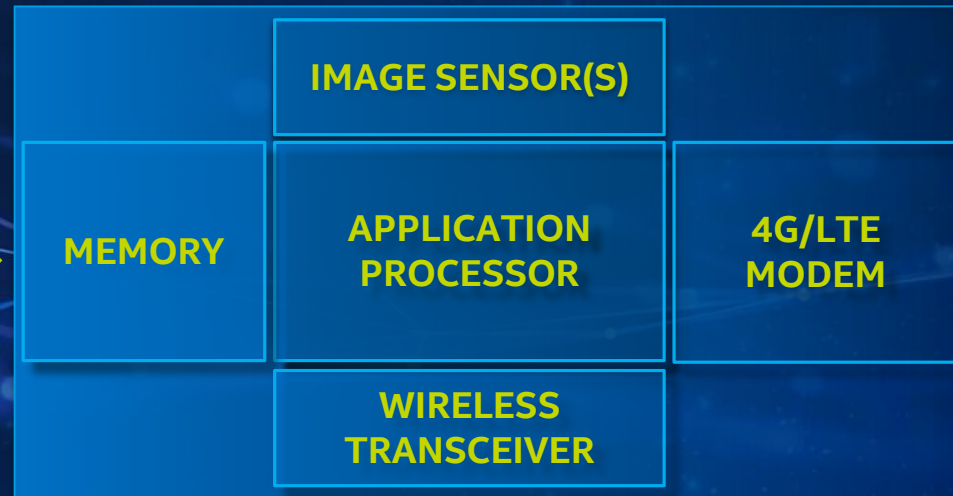
EDGE-TO-CLOUD: MOVING DATA CRITICALLY IMPACTS POWER

< 2 W



END POINT
DEVICE

5~20 W



EDGE COMPUTE NODE /
EDGE SERVER

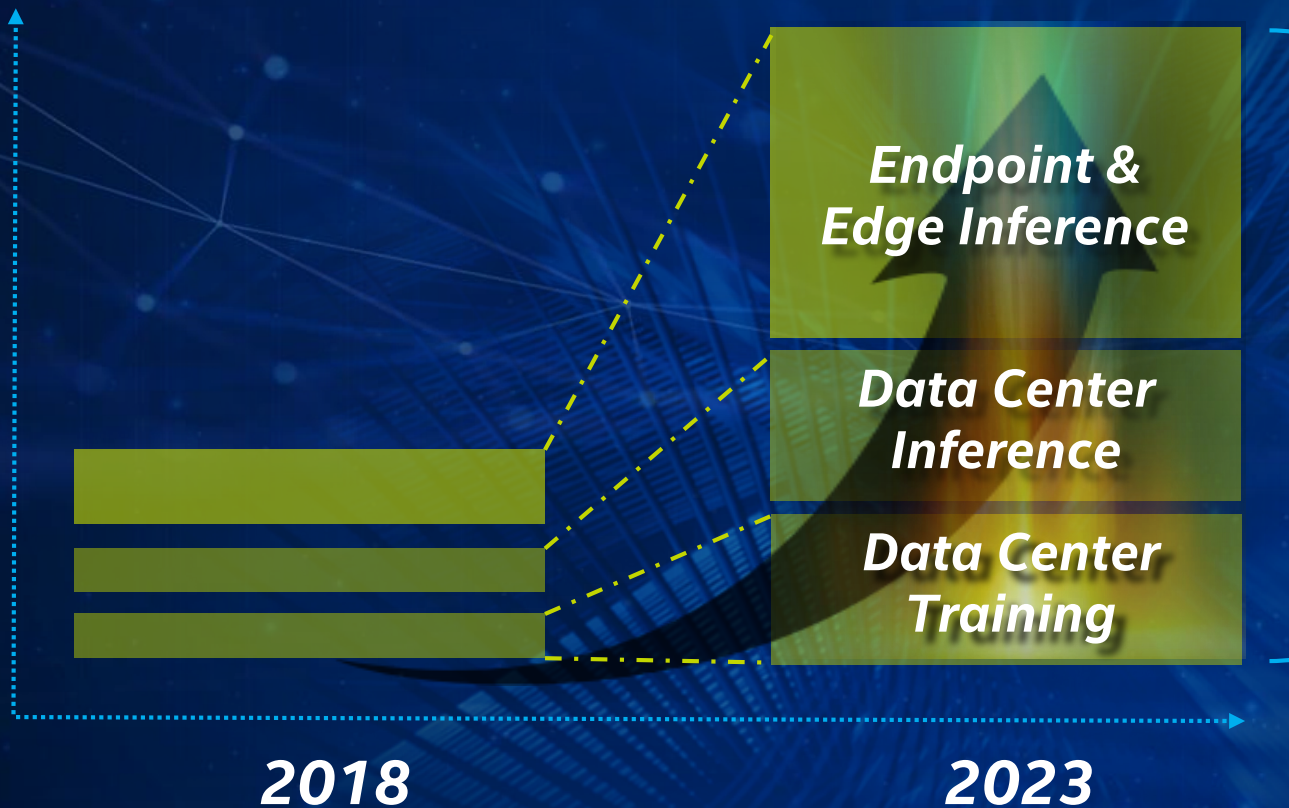
1000s W



CLOUD

THERE IS A GROWING OPPORTUNITY FOR AI AT THE EDGE

TAM



- **25%-35% CAGR** in AI for silicon industry
- **Growth larger at the Edge,** compared to Data center

Note: Endpoint and Edge excludes smartphones and client PCs
Source: Intel, IDC, Gartner

THE VIRTUOUS CYCLE DRIVING INNOVATION IN AI AT THE EDGE

EDGE PROCESSING REQUIREMENTS

- Huge Compute Requirements
- Low Latency
- Power Efficiency
- Local Data / Privacy

CHIP TECHNOLOGY IMPROVEMENTS

- Efficient Data Flow
- Better Algorithm Processing
- Efficient Memory Use
- Silicon Process Technology



EDGE AI CAPABILITIES ENABLE AUTONOMOUS FEATURES IN MANY APPLICATIONS



CITIES • STATE • FEDERAL

Public Safety & Surveillance
Traffic, Parking and LPR
Emergency Response



FINANCE • BANKING

People Counting Customer
(i.e. Gender, Wait Time)
ATM Facial Recognition



INDUSTRIAL

Machine Vision Asset Inspection
(i.e. Pipeline)
Augmented Reality



CASINO GAMING

Public Safety & Surveillance
Facial Recognition



TRANSPORTATION

Autonomous Vehicles
Public Safety (i.e. Bus/Rail)
Traffic & People Counting



HOME • RETAIL • SURVEILLANCE

Security & Surveillance
Responsive Retail Advertising
Digital Home Assistant



ROBOTICS

Manufacturing Automation
Industrial (i.e. Pipeline Welding)



DRONES

Emergency Response
Asset Inspection (i.e. Windmill)

INDUSTRIAL EXAMPLE: SMART FACTORY

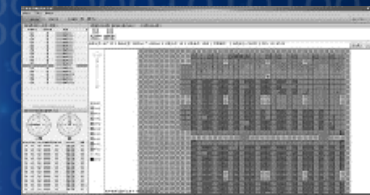
Smart Factory:

Will Generate an Estimate of 1PB data/day by 2020

Things & Edge Node

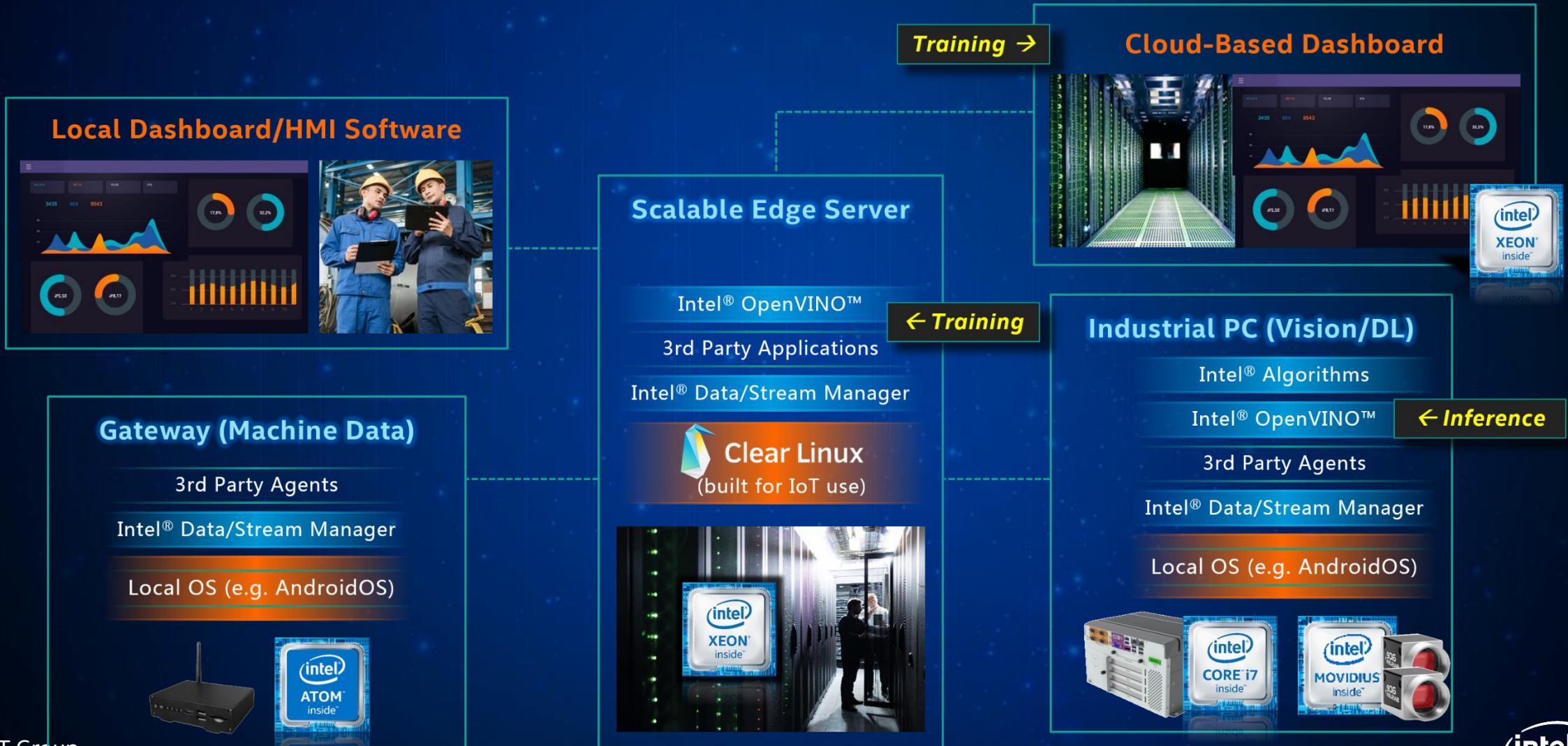
Edge Network & Core Network

Cloud



IN SMART FACTORY CONFIGURATION, FAST INSPECTION DATA IS ACTIONABLE

Vision+DL in industrial PC feeds into edge server



RETAIL EXAMPLE: SMART STORES

Retail Store:
Huge Data Processing Requirements

Things & Edge Node



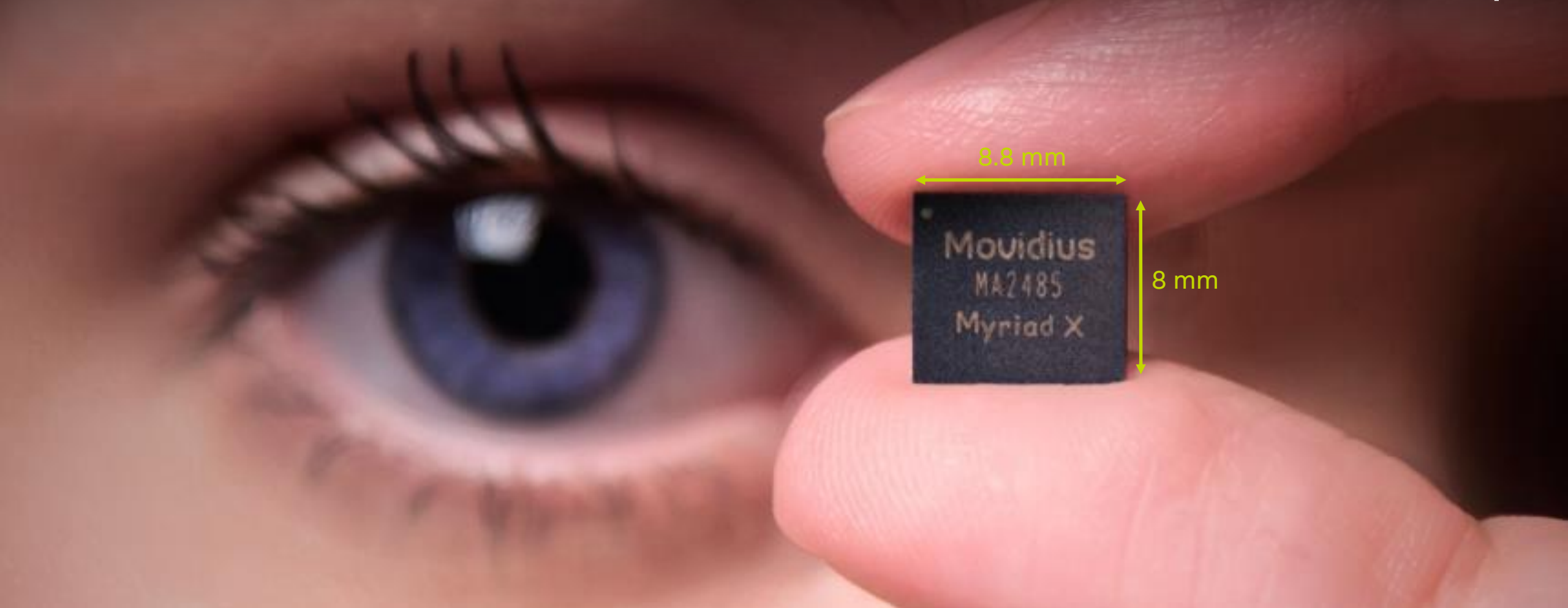
Edge Network & Core Network

- Item Tracking
- Mobile Payment
- Store Analytics
- Inventory Management
- Customer Tracking

Cloud



NEW INNOVATIVE CHIP ARCHITECTURES ENABLE COMPUTE-EFFICIENT
DEEP LEARNING INFERENCE **AT THE EDGE** WITH MORE PERFORMANCE/WATT/\$



**AI AT THE EDGE:
NOT JUST A TECHNICAL PROBLEM...**

HOW TO SCALE TO DEVELOPERS WORLDWIDE?*

***HINT: OPPORTUNITIES IN ASIA**

TO MAKE ENHANCED DEEP LEARNING AT THE EDGE MORE ACCESSIBLE, WE JUST ANNOUNCED: INTEL® VISION ACCELERATOR DESIGNS

Intel® Vision Accelerator Design with Intel® Movidius™ VPU

Intel® Vision Accelerator Design with Intel® Arria® 10 FPGA

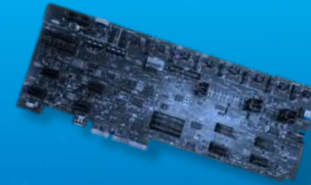
EXAMPLE CARD
BASED ON
VISION
ACCELERATOR
DESIGNS



1 Movidius
MA2485 VPU



2 Movidius
MA2485 VPUs



8 Movidius
MA2485 VPUs



Intel® Arria® 10 FPGA
1150GX/660GX

INTERFACE

M.2, Key E

miniPCIe

PCIe x4

PCIe x4

CURRENTLY
MANUFACTURED BY*



SOFTWARE
TOOLS

OPENVINO™ TOOLKIT

Develop NN Model; Deploy across Intel® CPU, GPU, VPU, FPGA; Leverage common algorithms

EXAMPLE PARTNER: KAGA ELECTRONICS (JAPAN)

加賀電子グループ

加賀電子株式会社

ENGLISH お問い合わせ

会社情報 | 事業紹介 | 投資家情報 | 環境・CSR | ニュースリリース | **製品情報** | 採用情報 | 加賀電子ストーリー

TOP > 製品情報 > 製品トピックス > ディープラーニング推論アクセラレータボード

製品トピックス ディープラーニング推論アクセラレータボード

エッジコンピューティングとディープラーニング

IoTの普及に伴い各種センサーが生み出すデータ量は、通信インフラやストレージインフラの進化を超える勢いで増え続けており、また、IoTサービスの質の面からも処理の即時性が求められる中、センサーデータの収集・分析にクラウドを使わないデータの地産地消を可能にするエッジコンピューティングが注目されています。

当社のディープラーニング推論アクセラレータボードは、既存のエッジサーバーやゲートウェイにアドオンする事で、デ



製品トピックス

ディープラーニング推論
アクセラレータボード

高機能HEMSゲートウェイ

CTUS-01

M2M通信ビジネス

LED照明ビジネス

eビジネス

TAYANI旧製品情報



EXAMPLE PARTNER: AAEON (TAIWAN)



Search

[Compare](#) [Inquiry](#) [Partners](#) [Global](#)

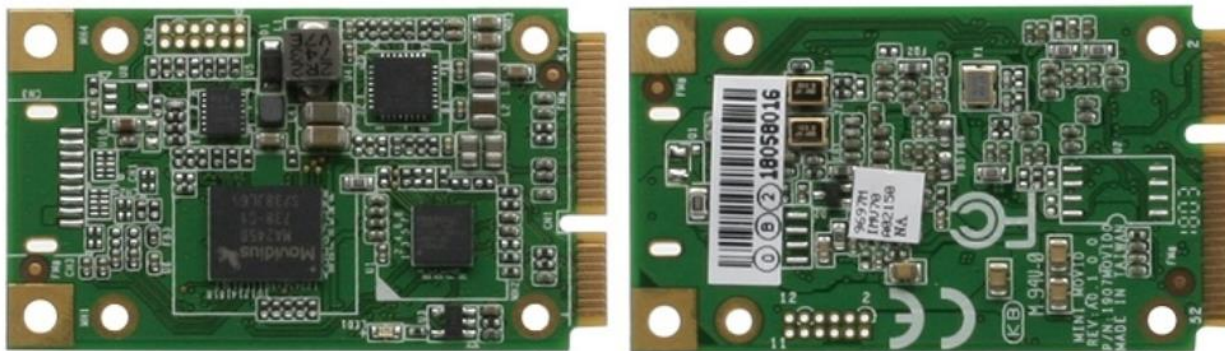
[eShop](#)

[Products](#) [Services](#) [Technologies](#) [Applications](#) [Support](#) [News & Events](#) [Resources](#) [About us](#) [Contact](#) [eShop](#)

AI Core – Artificial Intelligence On The Edge

2018-03-06

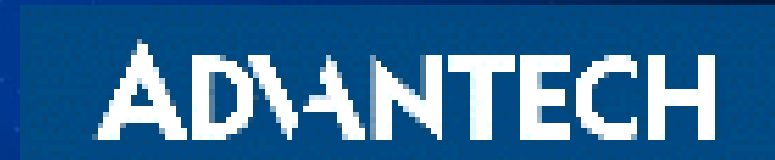
The first embedded ultra-compact Artificial Intelligence processing card for on the edge computing



UP Bridge the Gap – a brand of AAEON Europe – is proud to launch AI Core: the first embedded ultra-compact Artificial Intelligence processing card for on the edge computing.

[Get a Quote](#)

OTHER EXAMPLE PARTNER ODM COMPANIES FOR INTEL® VISION ACCELERATOR DESIGNS



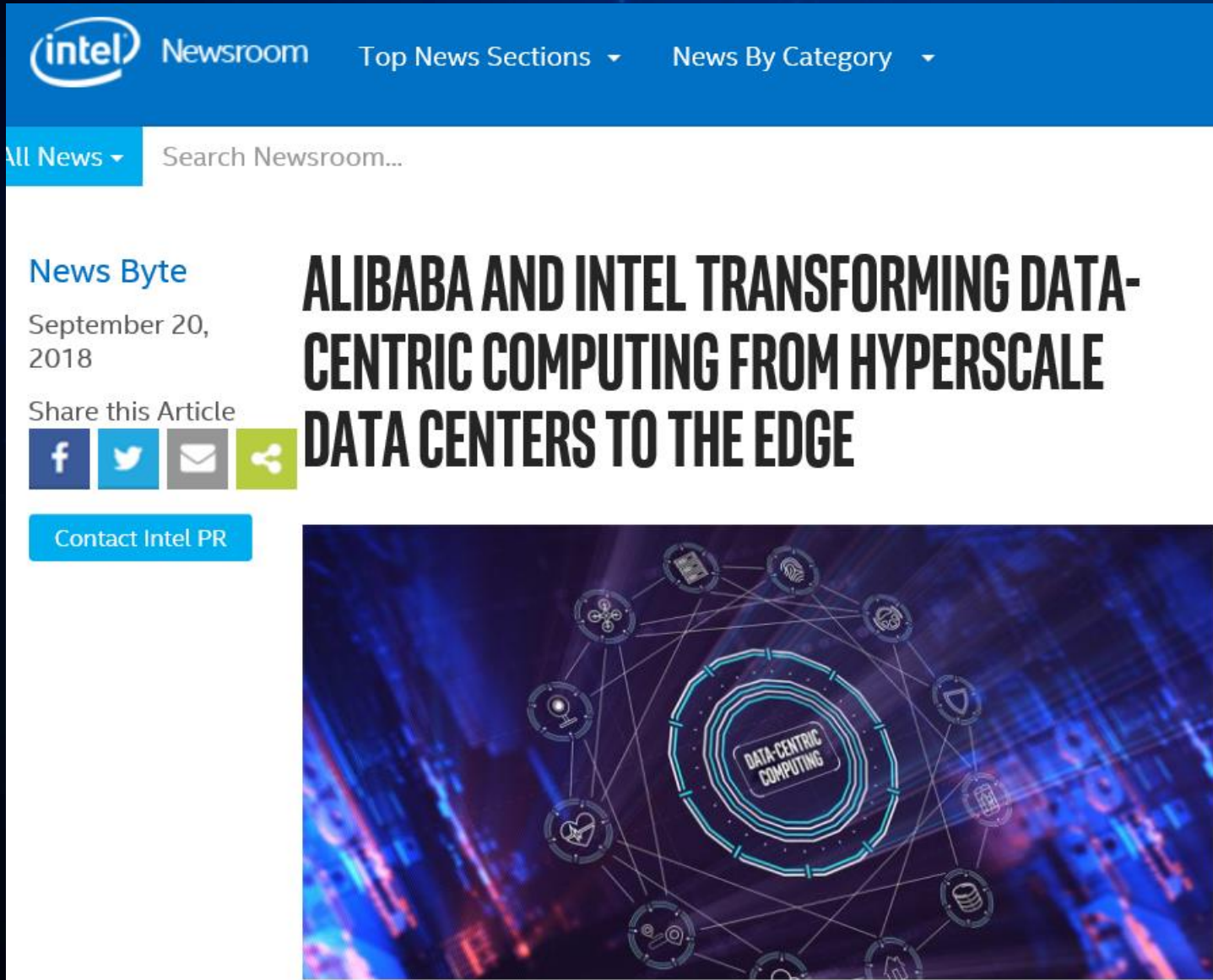
MANY OTHER PARTNERS TO BRING SCALE TO **AI AT THE EDGE**



INTEL® AI: IN PRODUCTION

A LEADING ECOSYSTEM FOR AI
AT THE EDGE DEPLOYMENTS

EXAMPLE PARTNERSHIP: ALIBABA (CHINA)



The screenshot shows the Intel Newsroom website. At the top is the Intel logo and navigation links for 'Newsroom', 'Top News Sections', and 'News By Category'. Below the navigation is a search bar and a 'All News' dropdown. The main article is titled 'ALIBABA AND INTEL TRANSFORMING DATA-CENTRIC COMPUTING FROM HYPERSCALE DATA CENTERS TO THE EDGE' and is dated September 20, 2018. It includes social media sharing icons for Facebook, Twitter, Email, and a general share icon, along with a 'Contact Intel PR' button. The article's featured image depicts a network of nodes connected by lines, with a central circular graphic containing the text 'DATA-CENTRIC COMPUTING'.

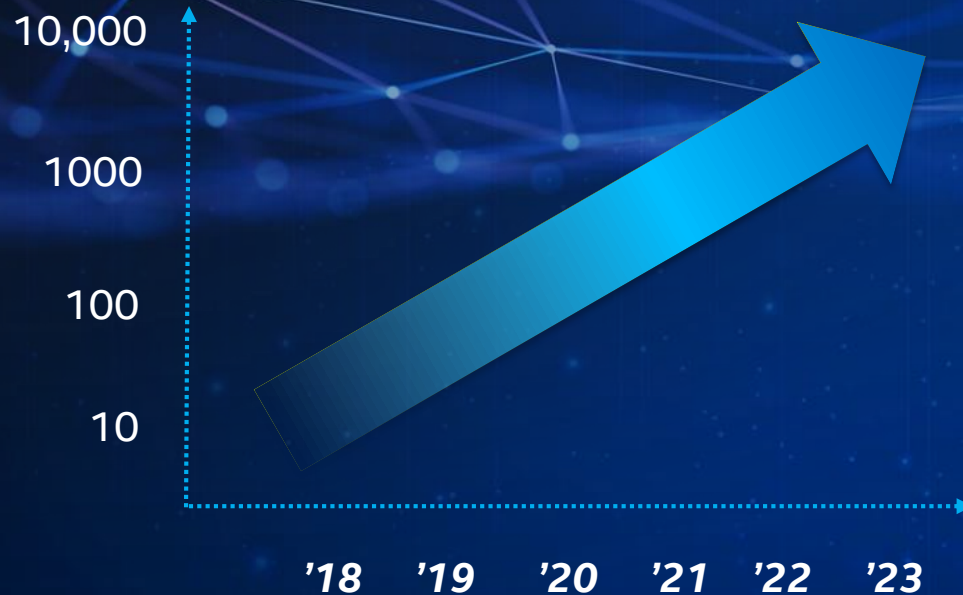
Intel and Alibaba Group are:

- Launching of a Joint Edge Computing Platform to accelerate edge computing development
- Establishing the Apsara Stack Industry Alliance targeting on-premises enterprise cloud environments
- Deploying latest Intel technology in Alibaba to prepare for the 11/11 shopping festival
- Bringing volumetric content to the Olympic Games Tokyo 2020 via OBS Cloud
- Accelerating the commercialization intelligent roads

SILICON INDUSTRY HAS CHALLENGES TO OVERCOME TO MAKE AUTONOMY **AT THE EDGE** POSSIBLE

AI PERFORMANCE FOR POWER HELD AT 5~10 W

DEEP NEURAL NETWORK PERFORMANCE
*Inferences Per Second for ResNet50, Batch Size = 1



SEMI INDUSTRY PROGRESS GOALS:

REDUCED PRECISION NEURAL NETWORKS

Improved performance moving from fp16 to binary weights, for example

NEURAL NETWORK COMPRESSION / SPARSITY

Reducing compute requirement for NN structure, and taking advantage of zeros in matrix computation

EFFICIENT MEMORY

Power efficient memory access with improved bandwidth

ACCELERATORS

Balancing fixed function performance with flexibility/programmability

COMPILER INNOVATION

Tools making best use of new hardware

THE **FUTURE** IS IN OUR HANDS

Movidius
MA2485
Myriad X

*I360

