Deep Learning Chipsets 2020 Report

Part of the AI & Intelligent Automation Service Area Package

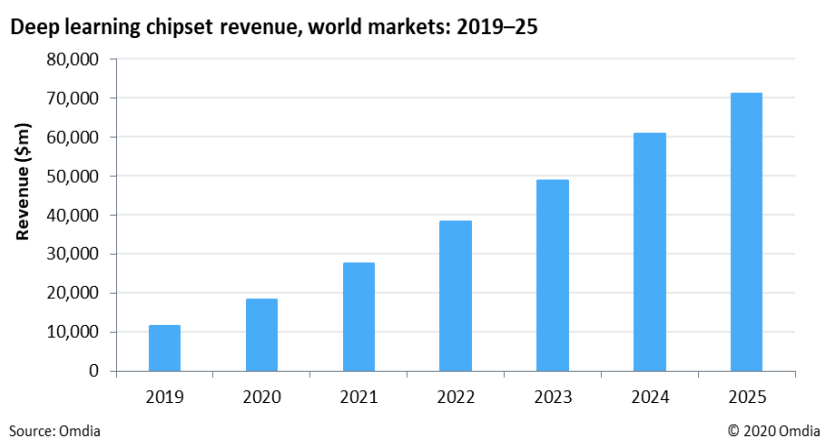
Anand Joshi, Principal Analyst

Published On July 1, 2020

CPUs, GPUs, FPGAs, ASICs, SoC Accelerators, and Other Chipsets for AI Training and Inference Applications: Global Market Analysis and Forecasts

The need for artificial intelligence (AI) acceleration is widely recognized as of 2020. AI acceleration chipsets have become a standard feature requirement for device manufacturers within the enterprise (data center) and edge markets. As a result, the volume and revenue of AI chipsets have increased drastically in the last two years. NVIDIA's latest A100 offers petaOPS of compute performance under certain compute conditions, marking a tremendous jump from the petaOPS server DGX-1 introduced just two short years ago.

Deep learning (DL) is slowly moving past its hype cycle as proof-of-concept (PoC) AI applications developed in the past two years go into production. AI chipset customers have become more sophisticated in terms of chipset needs for AI application acceleration and are asking for specific benchmarks when talking to vendors. Customers' needs for chipsets are coming to the forefront, forcing chipset companies to rethink the applicability of their technology. All prominent chip companies, such as Intel, NVIDIA, and Qualcomm, have invested heavily in AI. Cloud companies have started rolling out graphics processing units (GPUs), field programmable gate arrays (FPGAs), and application-specific integrated circuits (ASICs), giving developers a choice for AI acceleration. Omdia forecasts that global revenue for DL chipsets will increase from $11.4bn in 2019 to $71.2bn by 2025.



Report Coverage

KEY ISSUES ADDRESSED

- What chipset types are being used for deep learning (DL) today, and how will they change through 2025 and beyond?
- What are the power consumption and compute capacity profiles of chipsets used for DL applications?
- What is the market opportunity for DL chipsets in enterprise environments versus edge devices?
- Which market sectors and industries will drive demand for DL chipsets?
- What is the state of technology development for DL chipsets, and which companies are driving innovation?
- What are some of the emerging architectures for DL chipsets?
- What are the key performance matrices for DL chipsets?
- What are some of the use cases for DL chipsets in different application markets?
- What has changed in the DL chipset market in the last two years?
- How are startups faring in the DL chipset market?

KEY MARKET FORECASTS

- Deep learning chipset revenue, world markets: 2019–25
- Deep learning chipset year-on-year revenue growth rates, world markets: 2020–25
- Deep learning chipset revenue by chipset type, world markets: 2019–25
- Deep learning chipset revenue, inference vs. training, world markets: 2019–25
- Deep learning chipset revenue by compute capacity, world markets: 2019–25
- Deep learning chipset revenue by power consumption, world markets: 2019–25
- Deep learning chipset revenue by market sector, world markets: 2019–25

COVERAGE

Chipset Types

- Central processing unit (CPU)
- Graphics processing unit (GPU)
- Field-programmable gate array (FPGA)
- Application-specific integrated circuit (ASIC)
- System-on-chip (SoC) accelerator

Chipset Market Sectors

- Enterprise
- Edge

Segments

- Training
- Inference

Power Consumption

- High (>100 W)
- Medium (5-100 W)
- Low (<5 W)

Compute Capacity

- High (>50 TOPS)
- Medium (5–50 TOPS)
- Low (<5 TOPS)

Regions

- North America
- Europe
- Asia Pacific
- Latin America
- Middle East & Africa

APPLICABLE TO

- Semiconductor and component manufacturers
- OEM companies building devices using AI chipsets
- Cloud companies using AI chipsets
- Service providers and systems integrators
- End-user organizations deploying deep learning systems
- Industry associations
- Government agencies
- Investor community

CONNECT WITH US

CUSTOMER SUCCESS
customersuccess@omdia.com

SALES
US: +1 (212) 652 5335
APAC: +61 (0)396 016 700
EMEA: +44 (0)7771 980 316

@OmdiaHQ | omdia.com

ABOUT OMDIA

Omdia is a global technology research powerhouse, established following the merger of the research division of Informa Tech (Ovum, Heavy Reading, and Tractica) and the acquired Omdia technology research portfolio.* We combine the expertise of more than 400 analysts across the entire technology spectrum, covering 150 markets. We publish over 3,000 research reports annually, reaching more than 14,000 subscribers, and cover thousands of technology, media, and telecommunications companies. Our exhaustive intelligence and deep technology expertise enable us to uncover actionable insights that help our customers connect the dots in today's constantly evolving technology environment and empower them to improve their businesses–today and tomorrow.

*The majority of Omdia technology research products and solutions were acquired by Informa in August 2019 and are now part of Omdia.

Table of Contents

Executive summary
Introduction
2020 report update
Key findings
Market forecasts

Market issues
Use of AI in the market
- AI acceleration within enterprises
- AI acceleration at the edge

Market segmentation
- Segmentation by architecture (chipset type)
- Segmentation based on training vs. inference
- Segmentation based on compute capacity
Segmentation based on power consumption
Segmentation based on market sector: Enterprise and edge market

Market drivers
- Popularity of AI and increasing complexity
- Multiple AI pipelines
- Complexity of training
- Growth in enterprise applications
- Desire to minimize production costs
- Latency and throughput requirements for inference
- Computer vision
- Speech applications for embedded devices

Market barriers and challenges
- Capital needs for chip development
- Availability of expertise
- Long development cycle and rapidly changing market

Applications and use cases
- Enterprise applications and use cases
- Edge applications and use cases
- Other

Regional differences
Startup activity in deep learning chipsets
- Many acquisitions
- Casualties

Technology issues
Evolution of neural networks since 2012 and the need for hardware acceleration
- Computation needs per forward pass (inference)
- Compute needs for training
- A neural network zoo

AI inference workloads
- Recommendation engine
- Image and video
- Audio and speech
- Text/natural language processing
- Search

Translating neural network needs to chipset requirements
- Processing elements and arithmetic logic units
- Memory
- On-chip connectivity
- Chip-to-chip connectivity

Chipset architectures for deep learning
- Central processing units
- Graphics processing units
- Field-programmable gate arrays
- Application-specific integrated circuits
- System-on-chip accelerators

Emerging AI acceleration architectures
- Optical computing
- Analog computing
- Processing in memory
- Neuromorphic

Technology parameters for chipsets
Benchmark
Data formats used in AI chipsets
Deep learning development frameworks

Key industry players
Amazon
AMD
ARM
Cerebras Systems
CEVA
Esperanto Technologies
Facebook
Google
Graphcore
Groq
Gyrfalcon Technologies
Habana Labs (acquired by Intel)
Huawei
Intel
Kalray
MediaTek
Movidius (Intel)
Mobileye (Intel)
NVIDIA
Qualcomm
SambaNova
Thinci (now Blaize)
Xilinx
Deep learning chipset and IP companies

Market forecasts
Forecast methodology and assumptions
Omdia coverage of AI chipsets
Overall market
Revenue by chipset type
Revenue by training vs. inference

Revenue by compute capacity
Revenue by power consumption
Average selling price by chipset type

Revenue by market sector
Central processing units
Graphics processing units
Application-specific integrated circuits
Field-programmable gate arrays
System-on-chip accelerators
Conclusions

CONNECT WITH US

CUSTOMER SUCCESS
customersuccess@omdia.com

SALES
US: +1 (212) 652 5335
APAC: +61 (0)396 016 700
EMEA: +44 (0)7771 980 316

@OmdiaHQ | omdia.com

ABOUT OMDIA

Omdia is a global technology research powerhouse, established following the merger of the research division of Informa Tech (Ovum, Heavy Reading, and Tractica) and the acquired Omdia technology research portfolio.* We combine the expertise of more than 400 analysts across the entire technology spectrum, covering 150 markets. We publish over 3,000 research reports annually, reaching more than 14,000 subscribers, and cover thousands of technology, media, and telecommunications companies. Our exhaustive intelligence and deep technology expertise enable us to uncover actionable insights that help our customers connect the dots in today's constantly evolving technology environment and empower them to improve their businesses–today and tomorrow.

*The majority of Omdia technology research products and solutions were acquired by Informa in August 2019 and are now part of Omdia.

Tables

Deep learning chipset revenue by chipset type, world markets: 2019–25

Deep learning chipset revenue, enterprise vs. edge, world markets: 2019–25

Deep learning chipset revenue growth rates, world markets: 2020–25

Deep learning chipset revenue by power consumption, world markets: 2019–25

Deep learning chipset revenue by compute capacity, world markets: 2019–25

Deep learning chipset revenue, inference vs. training, world markets: 2019–25

Deep learning edge chipset revenue by chipset type, world markets: 2019–25

Deep learning edge chipset shipments by chipset type, world markets: 2019–25

Deep learning edge chipset revenue growth rates, world markets: 2020–25

Deep learning edge chipset revenue by power consumption, world markets: 2019–25

Deep learning edge chipset revenue by compute capacity, world markets: 2019–25

Deep learning edge chipset revenue for inference vs. training, world markets: 2019–25

Deep learning enterprise chipset revenue by chipset type, world markets: 2019–25

Deep learning enterprise chipset shipments by chipset type, world markets: 2019–25

Deep learning enterprise chipset revenue growth rates, world markets: 2020–25

Deep learning enterprise chipset revenue by power consumption, world markets: 2019–25

Deep learning enterprise chipset revenue by compute capacity, world markets: 2019–25

Deep learning enterprise chipset revenue for inference vs. training, world markets: 2019–25

Deep learning edge chipset ASPs by chipset type, mobile, HMDs, drones, and machine vision (non-PC), world markets: 2019–25

Deep learning edge chipset ASPs by chipset type, edge servers, world markets: 2019–25

Deep learning edge chipset ASPs by chipset type, PCs/tablets, world markets: 2019–25

Deep learning edge chipset ASPs by chipset type, cameras, world markets: 2019–25

Deep learning edge chipset ASPs by chipset type, smart speakers, world markets: 2019–25

Deep learning edge chipset ASPs by chipset type, automotive, world markets: 2019–25

Deep learning edge chipset ASPs by chipset type, robots, world markets: 2019–25

Deep learning enterprise training chipset ASPs by chipset type, world markets: 2019–25

Deep learning enterprise inference chipset ASPs by chipset type, world markets: 2019–25

Deep learning CPU chipset revenue by market sector, world markets: 2019–25

Deep learning CPU enterprise chipset revenue, training vs. inference, world markets: 2019–25

Deep learning CPU edge chipset revenue, training vs. inference, world markets: 2019–25

Deep learning CPU enterprise chipset revenue by power consumption, world markets: 2019–25

Deep learning GPU chipset revenue by market sector, world markets: 2019–25

Deep learning GPU enterprise chipset revenue, training vs. inference, world markets: 2019–25

Deep learning GPU edge chipset revenue, training vs. inference, world markets: 2019–25

Deep learning GPU enterprise chipset revenue by power consumption, world markets: 2019–25

Deep learning GPU enterprise chipset revenue by compute capacity, world markets: 2019–25

Deep learning GPU edge chipset revenue by power consumption, world markets: 2019–25

Deep learning GPU edge chipset revenue by compute capacity, world markets: 2019–25

Deep learning FPGA chipset revenue by market sector, world markets: 2019–25

Deep learning FPGA enterprise chipset revenue, training vs. inference, world markets: 2019–25

Deep learning FPGA edge chipset revenue, training vs. inference, world markets: 2019–25

Deep learning FPGA enterprise chipset revenue by power consumption, world markets: 2019–25

Deep learning FPGA enterprise chipset revenue by compute capacity, world markets: 2019–25

Deep learning FPGA edge chipset revenue by power consumption, world markets: 2019–25

Deep learning FPGA edge chipset revenue by compute capacity, world markets: 2019–25

Deep learning ASIC chipset revenue by market sector, world markets: 2019–25

Deep learning ASIC enterprise chipset revenue, training vs. inference, world markets: 2019–25

Deep learning ASIC edge chipset revenue, training vs. inference, world markets: 2019–25

Deep learning ASIC enterprise chipset revenue by power consumption, world markets: 2019–25

Deep learning ASIC enterprise chipset revenue by compute capacity, world markets: 2019–25

Deep learning ASIC edge chipset revenue by power consumption, world markets: 2019–25

Deep learning ASIC edge chipset revenue by compute capacity, world markets: 2019–25

Deep learning SoC accelerator chipset revenue by market sector, world markets: 2019–25

Deep learning SoC accelerator enterprise chipset revenue, training vs. inference, world markets: 2019–25

CONNECT WITH US

CUSTOMER SUCCESS
customersuccess@omdia.com

SALES
US: +1 (212) 652 5335
APAC: +61 (0)396 016 700
EMEA: +44 (0)7771 980 316

@OmdiaHQ | omdia.com

ABOUT OMDIA

Omdia is a global technology research powerhouse, established following the merger of the research division of Informa Tech (Ovum, Heavy Reading, and Tractica) and the acquired Omdia technology research portfolio.* We combine the expertise of more than 400 analysts across the entire technology spectrum, covering 150 markets. We publish over 3,000 research reports annually, reaching more than 14,000 subscribers, and cover thousands of technology, media, and telecommunications companies. Our exhaustive intelligence and deep technology expertise enable us to uncover actionable insights that help our customers connect the dots in today's constantly evolving technology environment and empower them to improve their businesses–today and tomorrow.

*The majority of Omdia technology research products and solutions were acquired by Informa in August 2019 and are now part of Omdia.

Tables, cont'd & Figures

Deep learning SoC accelerator edge chipset revenue, training vs. inference, world markets: 2019–25

Deep learning SoC accelerator enterprise chipset revenue by power consumption, world markets: 2019–25

Deep learning SoC accelerator enterprise chipset revenue by compute capacity, world markets: 2019–25

Deep learning SoC accelerator edge chipset revenue by power consumption, world markets: 2019–25

Deep learning SoC accelerator edge chipset revenue by compute capacity, world markets: 2019–25

Types of devices with enterprises using AI accelerators

Edge devices shipping in high volume and chipset requirements

Key players in different deep learning chipsets

Key CPU products and vendors

Key players in GPU

Deep learning CPU enterprise chipset revenue by compute capacity, world markets: 2019–25

Deep learning CPU edge chipset revenue by power consumption, world markets: 2019–25

Deep learning CPU edge chipset revenue by compute capacity, world markets: 2019–25

Figures

Deep learning chipset revenue, world markets: 2019–25

Estimated AI workloads on enterprise GPUs and CPUs

Deep learning chipset revenue, world markets: 2019–25

Deep learning chipset year-on-year revenue growth rates, world markets: 2020–25

Deep learning chipset revenue by chipset type, world markets: 2019–25

Deep learning chipset revenue, inference vs. training, world markets: 2019–25

Deep learning chipset revenue by compute capacity, world markets: 2019–25

Deep learning chipset revenue by power consumption, world markets: 2019–25

Deep learning chipset revenue by market sector, world markets: 2019–25

Deep learning CPU chipset revenue, world markets: 2019–25

Deep learning GPU chipset revenue, world markets: 2019–25

Deep learning ASIC chipset revenue, world markets: 2019–25

Deep learning FPGA chipset revenue, world markets: 2019–25

Deep learning SoC accelerator chipset revenue, world markets: 2019–25

Key players in FPGA

Comparison of deep learning chipset parameters

Selected benchmarks for AI chipsets

Data formats used in AI chipsets

Popular deep learning frameworks

Deep learning chipset companies

IP companies

CONNECT WITH US

CUSTOMER SUCCESS
customersuccess@omdia.com

SALES
US: +1 (212) 652 5335
APAC: +61 (0)396 016 700
EMEA: +44 (0)7771 980 316

@OmdiaHQ | omdia.com

ABOUT OMDIA

Omdia is a global technology research powerhouse, established following the merger of the research division of Informa Tech (Ovum, Heavy Reading, and Tractica) and the acquired Omdia technology research portfolio.* We combine the expertise of more than 400 analysts across the entire technology spectrum, covering 150 markets. We publish over 3,000 research reports annually, reaching more than 14,000 subscribers, and cover thousands of technology, media, and telecommunications companies. Our exhaustive intelligence and deep technology expertise enable us to uncover actionable insights that help our customers connect the dots in today's constantly evolving technology environment and empower them to improve their businesses–today and tomorrow.

*The majority of Omdia technology research products and solutions were acquired by Informa in August 2019 and are now part of Omdia.