

Forecast: AI Semiconductors, Worldwide, 2020-2026

Published 5 May 2022 - ID G00763796 - 3 min read

By Analyst(s): Alan Priestley

Initiatives: [Technology Market Essentials](#)

The use of artificial intelligence techniques in data centers, edge computing and endpoint devices has created a market for optimized semiconductor devices. Revenue from these AI semiconductors is forecast to be \$86 billion by 2026, growing by a five-year CAGR of 20%.

Table of Contents

- 1 Summary Tables Suitable for Printing
- 2 Pivot Table for Analysis — AI Semiconductors
- 3 Data Segmentation and Structure
- 4 Tips

List of Tables

- 1-1 AI Semiconductor Revenue by Electronic Equipment Category, 2020-2026 (\$M)
- 1-2 AI Semiconductor Revenue Growth by Electronic Equipment Category, 2020-2026 (%)
- 1-3 AI Semiconductor Revenue by Semiconductor Device Category, 2020-2026 (\$M)
- 1-4 AI Semiconductor Revenue Growth by Semiconductor Device Category, 2020-2026 (%)
- 2-1 AI Semiconductor Forecast Worldwide Revenue — By Semiconductor Device, 2020-2026 (\$M)

Overview

This forecast dataset provides Gartner's forecast for artificial intelligence (AI) semiconductors and is provided to enable a deeper insight into the emerging business opportunity for AI semiconductor devices.

This forecast is a subelement of the published [Semiconductors and Electronics Forecast Database, Worldwide, 1Q22 Update](#).

AI Semiconductors

The need to analyze complex “real-world” datasets is driving the demand for chips, optimized for execution of workloads, that leverage algorithms and models based on deep neural networks (DNNs). These workloads are often referred to as AI. Traditional microprocessors can be used to execute AI applications. However, they are not an optimal solution. Most AI-based workloads are highly parallel in nature and are best executed on chips specifically designed for these workloads. Currently, many application developers are using graphics processing unit (GPU)-based solutions and there is a wide range of semiconductor vendors developing chips specifically targeted for these workloads, either for use in data centers, edge computing systems or IoT endpoints.

This forecast dataset covers semiconductor devices designed to execute DNN-based workloads. This includes devices that can only execute DNN algorithms and also are designed for a range of applications, such as discrete application/multimedia processors, that have additional logic blocks included to support DNN execution. See Note 1 for the definition of the AI semiconductors included in this forecast.

Acronym Key and Glossary Terms

AI	artificial intelligence
DNN	deep neural network
GPU	graphics processing unit
MCU	microcontroller unit
MPU	microprocessor unit

Note 1. Definition — AI Semiconductors

This Gartner forecast covers the semiconductor revenue associated with devices designed specifically to execute the algorithms and programmatic models associated with DNNs used in AI. This is broken down into two major device segments: discrete AI semiconductor devices and integrated AI semiconductor devices.

Discrete AI Semiconductor Devices

These semiconductor devices are designed for the purpose of executing the highly parallel program algorithms associated with AI. These devices operate as dedicated AI processors or as application accelerators, in conjunction with a general-purpose or application-specific host processor.

The primary role of these devices is AI-related, and they may be used for either training or inference.

These semiconductor devices may be packaged on modules that include on-package memory and communications interfaces.

Examples include:

- Discrete GPUs from AMD, Intel and NVIDIA.
- AI chips from startups and established semiconductor vendors such as Graphcore, Gryfalcon Technology, Intel, Hailo, Qualcomm and SambaNova Systems.
- AI chips developed by the hyperscale cloud service providers, such as Amazon Web Services (AWS) Trainium, AWS Inferentia and Google Cloud TPU.

FPGAs can also be configured to execute AI algorithms and are included in this category.

Integrated AI Semiconductor Devices

Many modern MPUs and discrete application processors include dedicated IP blocks to execute AI neural network algorithms that complement the main processor functionality of the device. Executing AI algorithms is not the primary purpose of these devices. However, integration of AI functionality within the device is a requirement that enables the semiconductor vendor to successfully win business with the device. This forecast includes all revenue from the devices with integrated AI functional blocks.

The following are examples of devices that include AI functionality, but it is not their primary purpose:

- Smartphone application processors
- Application processors used in autonomous driving equipment
- Internet of Things (IoT) ASSPs and ASICs

What's Not Included in the AI Semiconductor Forecast

This forecast covers only the revenue generated from the sale of semiconductor devices that include (and are marketed with) specific functional logic blocks for running AI applications and whose AI functionality is a prime contributor to the vendor winning the socket. Therefore:

- General-purpose processors used to run AI workloads are not included in this forecast.
 - For example, x86 MPUs running AI applications with no additional AI processors or accelerators.
- General-purpose devices used to complement an AI-specific device in manufacturing an AI product are not included.
 - For example, DRAM, discrete components and MPUs/MCUs.

Revenue from AI application software and services is not included in this forecast.

Document Revision History

[Forecast: AI Semiconductors, Worldwide, 2019-2025, 4Q21 Update - 24 December 2021](#)

[Forecast: AI Semiconductors, Worldwide, 2019-2025, 3Q21 Update - 24 September 2021](#)

[Forecast: AI Semiconductors, Worldwide, 2019-2025, 2Q21 Update - 1 July 2021](#)

[Forecast: AI Semiconductors, Worldwide, 2019-2025, 1Q21 Update - 13 April 2021](#)

[Forecast Database, AI Neural Network Processing Semiconductors, 1Q20 - 16 April 2020](#)

Recommended by the Author

Some documents may not be available as part of your current Gartner subscription.

[Forecast Analysis: AI Semiconductors, Worldwide](#)

[Semiconductors and Electronics Forecast Database, Worldwide, 1Q22 Update](#)

[Market Definitions and Methodology: Semiconductor Devices and Applications](#)

[Emerging Technologies and Trends Impact Radar: Semiconductor and Electronics Technologies](#)

[Emerging Technologies and Trends Impact Radar: Artificial Intelligence, 2021](#)

[Emerging Technologies: Critical Insights on AI Semiconductors for Endpoint and Edge Computing](#)

© 2022 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)."