Artificial intelligence (AI) ecosystem for STM32N6 high-performance microcontroller with Neural-ART Accelerator
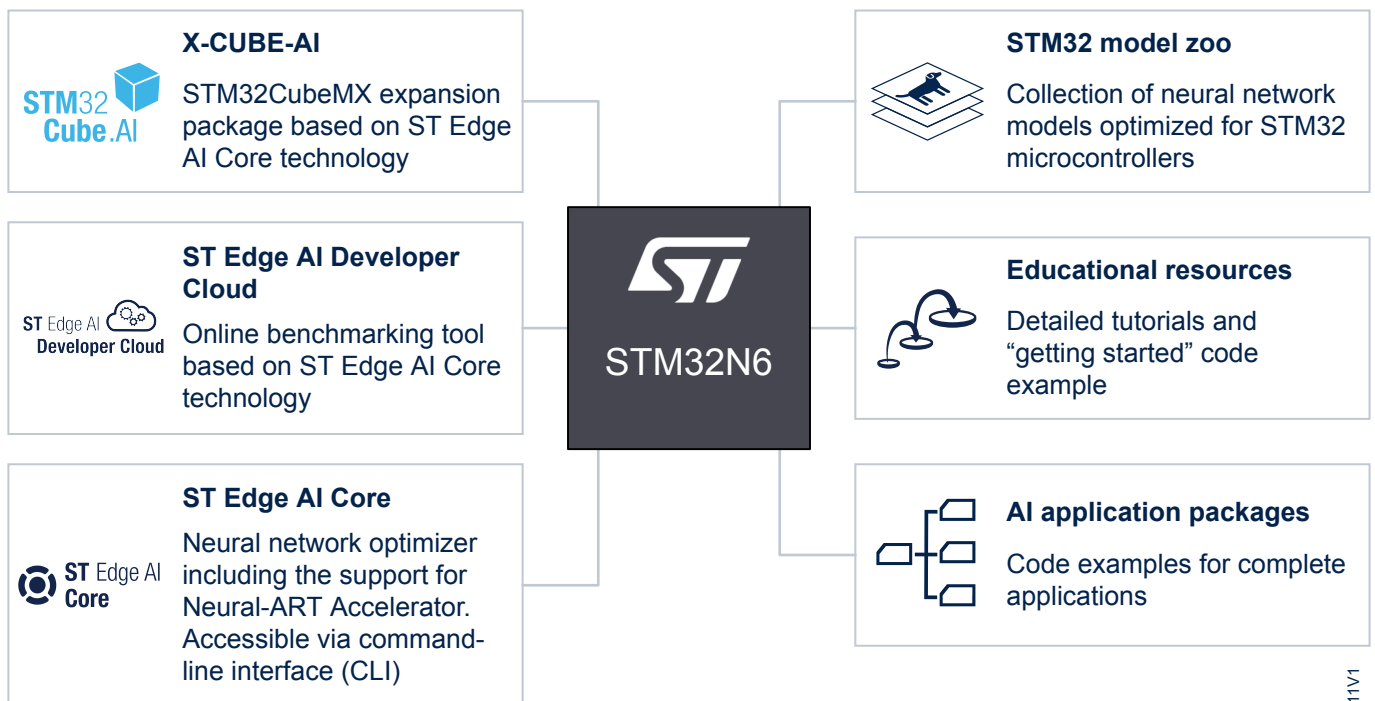
X-CUBE-AI

STM32CubeMX expansion package based on ST Edge AI Core technology

ST Edge AI Developer Cloud

Online benchmarking tool based on ST Edge AI Core technology

ST Edge AI Core

Neural network optimizer including the support for Neural-ART Accelerator. Accessible via command-line interface (CLI)

STM32N6

STM32 model zoo

Collection of neural network models optimized for STM32 microcontrollers

Educational resources

Detailed tutorials and "getting started" code example

AI application packages

Code examples for complete applications

DT76511V1

Product status link
STM32N6-AI

DB5437 - Rev 1 - December 2024
For further information contact your local STMicroelectronics sales office.

www.st.com

Features

- STM32Cube.AI (X-CUBE-AI):
    – Desktop tool for the optimization of NN models (plug-in for STM32CubeMX)
    – Automatic C code generation for STM32 microcontrollers
- ST Edge AI Developer Cloud (STEDGEAI-DC):
    – Online platform for benchmarking AI performance (inference time, memory footprints) on STM32 boards hosted online in a board farm
    – Automatic C code generation for STM32 microcontrollers
    – Process automation through a REST API
- ST Edge AI Core (STEdgeAI-Core):
    – Command-line interface (CLI) for the optimization of NN models
    – Automatic C code generation for STM32 microcontrollers
- STM32 model zoo:
    – Access to a curated collection of optimized AI models with associated performances, ready-to-use application examples, and scripts for model training, evaluation, quantization, benchmarking, and deployment
    – Utilization of ST Edge AI Developer Cloud services through a REST API
- Application packages:
    – Access to optimized source code for several AI applications
    – Seamless implementation on user's boards
- Getting started example codes:
    – Simple AI code examples and tutorials describing how to implement an AI application using STMicroelectronics' Neural-ART Accelerator
    – Optional link with the STM32 model zoo resources
- Large offering of tools to match the user's profile and goals
- Straightforward integration of the Neural-ART Accelerator into the AI software ecosystem for optimal AI application performance
- Easy access to model selection, training scripts, and key model metrics, directly available for benchmarking
- ML benchmarking automation service with Python™ scripts (REST API)
- Broad selection of use cases and application code examples to quickly get started with AI on MCU
- Native support for various deep learning frameworks such as Keras and TensorFlow™ Lite, and support for all frameworks that can export to the ONNX standard format such as PyTorch™, MATLAB®, and more
- Free and user-friendly license terms

Description

The STM32N6 AI ecosystem (STM32N6-AI) is STMicroelectronics' collection of tools and resources to support the development and deployment of AI models on the high-performance STM32N6 series microcontrollers.

The STM32N6 microcontroller (MCU) is the first to feature STMicroelectronics' proprietary Neural-ART Accelerator neural processing unit (NPU) for a major leap in AI/ML performance.

STM32N6-AI tools are designed to streamline the development process and ensure that developers can achieve optimal performance and efficiency. They offer support for both bring your own data (BYOD) and bring your own model (BYOM) approaches to match users' best development practices and preferences.

Tools such as STM32 model zoo (github.com/STMicroelectronics/stm32ai-modelzoo), ST Edge AI Developer Cloud (STEDGEAI-DC), STM32Cube.AI (X-CUBE-AI), and ST Edge AI Core (STEdgeAI-Core) mobilize the STM32N6 potential for AI and computer vision (CV) applications.

Additionally, STM32N6-AI offers various software packages that serve as examples and starting points for a user's AI projects:

- For AI: Include simple CV and audio applications such as people detection or image classification, as well as more complex and optimized applications such as pose estimation, instance segmentation, hand landmark detection, and audio scene recognition.
- For video: Demonstrate a complete application involving H264 encoding and USB video device class stream output data to a PC.

Overall, STM32N6-AI provides access to a full ecosystem of software and tools dedicated to help build next-generation machine learning applications at the edge with the STM32N6. The Neural-ART Accelerator NPU embedded in the STM32N6 efficiently handles AI inference tasks and provides an exceptional acceleration for NN models execution. This integration makes edge AI on MCU both practical and widespread, and offers a powerful, efficient, and scalable solution for a diverse range of applications.

The Neural-ART Accelerator is fully supported by ST Edge AI Core, including STM32Cube.AI and ST Edge AI Developer Cloud. These tools optimize the NN models and generate the code to be executed in hardware by the NPU. They analyze the neural network, prepare the data, and map its operators to the appropriate hardware resources to fully leverage the NPU capabilities without hassle and to benefit from optimal AI acceleration. Streamlined support for models from the most popular AI frameworks such as Keras, TensorFlow™, and ONNX ensures a smooth AI development pipeline.

ST Edge AI Suite

All the tools and software packages contributing to the STM32N6 AI ecosystem are part of STMicroelectronics ST Edge AI Suite, which is an integrated collection of software tools designed to facilitate the development and deployment of embedded AI applications. This comprehensive suite supports both optimization and deployment of machine learning algorithms and neural network models, from data collection to the final deployment on hardware, streamlining the workflow for professionals across various disciplines.

The ST Edge AI Suite supports various STMicroelectronics products: STM32 microcontrollers and microprocessors, Neural-ART Accelerator, Stellar microcontrollers, and smart sensors.

The ST Edge AI Suite represents a strategic move to democratize edge AI technology, making it a pivotal resource for developers looking to harness the power of AI in embedded systems efficiently and effectively.

Note: Arm is a registered trademark of Arm Limited (or its subsidiaries) in the US and/or elsewhere.

arm

1 Access information

Get access to the STM32N6 AI ecosystem home page (STM32N6-AI) on STMicroelectronics website at *www.st.com*.

Direct accesses to the main STM32N6 AI ecosystem resources are indicated below:

- Free access to ST Edge AI Developer Cloud is available at stedgeai-dc.st.com/home. Log in with STMicroelectronics mySt credentials.
- ST Edge AI Core (STEdgeAI-Core) is available for free download from the *www.st.com* website.
- STM32 model zoo models, application examples, and scripts are available for free download from github.com/STMicroelectronics/stm32ai-modelzoo.
- STM32Cube.AI (X-CUBE-AI) is available for free download from the *www.st.com* website.

Revision history

Table 1. Document revision history

Date	Revision	Changes
04-Dec-2024	1	Initial release.

IMPORTANT NOTICE – READ CAREFULLY

STMicroelectronics NV and its subsidiaries ("ST") reserve the right to make changes, corrections, enhancements, modifications, and improvements to ST products and/or to this document at any time without notice. Purchasers should obtain the latest relevant information on ST products before placing orders. ST products are sold pursuant to ST's terms and conditions of sale in place at the time of order acknowledgment.

Purchasers are solely responsible for the choice, selection, and use of ST products and ST assumes no liability for application assistance or the design of purchasers' products.

No license, express or implied, to any intellectual property right is granted by ST herein.

Resale of ST products with provisions different from the information set forth herein shall void any warranty granted by ST for such product.

ST and the ST logo are trademarks of ST. For additional information about ST trademarks, refer to www.st.com/trademarks. All other product or service names are the property of their respective owners.

Information in this document supersedes and replaces information previously supplied in any prior versions of this document.