

RN00459

Ara2 Software Packages Release Notes

Rev. 1.0 — 13 April 2026

Release notes

Document information

Information	Content
Keywords	RN00459, Ara240, Ara2 Runtime SDK, Ara2 Vision Example, LLM, VLM, eIQ AAF Connector, CNN
Abstract	This document contains important information about the contents of different packages and demos available for Ara240 16GB M.2 Module with FRDM i.MX 8M Plus and FRDM i.MX 95 boards.



1 Overview

This document contains important information about the contents of different packages and demos available for Ara240 16GB M.2 Module with FRDM i.MX 8M Plus and FRDM i.MX 95 boards. It covers supported features, known issues, and limitations in this release.

2 Features

Table 1. Ara2 software packages

Package	Features
Ara2 Runtime Package	Provides a runtime environment for Ara240. Provides scripts to download and run pre-compiled CNN models on Ara240 modules.
Ara2 Vision Example	Demonstrate a Vision pipeline capable of handling up to 8 simultaneous video streams for object detection on the Ara240 discrete NPU.
LLM Edge Studio	Showcase unimodal capabilities on i.MX for supported Large Language Models (LLMs) accelerated by the Ara240 discrete NPU at the edge.
VLM Edge Studio	Showcase Ara240 multimodal capabilities on i.MX using VLMs for video and image analysis.
eIQ AAF Connector	Service interface that exposes an OpenAI-compatible REST API to client applications while brokering requests to locally hosted LLM/VLM backends on Ara240 discrete NPU.

2.1 Models

CNNs

- <https://huggingface.co/nxp/YOLOv8>
 - Detection (n, s, m, l, x)
 - Pose Estimation (n, s, m, l, and x)
 - Segmentation (n, s, m, l, and x)

LLMs

- <https://huggingface.co/nxp/Qwen2.5-Coder-1.5B-Ara240>
- <https://huggingface.co/nxp/Qwen2.5-7B-Instruct-Ara240>

VLMs

- <https://huggingface.co/nxp/Qwen2.5-VL-7B-Instruct-Ara240>

3 Licenses

Table 2. Package license information

Package name	Downloadable file	License
Ara2 Runtime SDK	rt-sdk-ara2-2.0.4.deb	LA_OPT_NXP_Software_License
Ara2 Vision Examples	ara-vision-examples_1.0.deb	LA_OPT_Online Code Hosting NXP_Software_License
LLM Edge Studio	llm-edge-studio_2.0.0.deb	LA_OPT_Online Code Hosting NXP_Software_License

Table 2. Package license information...continued

Package name	Downloadable file	License
VLM Edge Studio	vlm-edge-studio_1.0.0.deb	LA_OPT_Online Code Hosting NXP_Software_License
eIQ Connector	eiq-aaf-connector_2.0_all.deb	LA_OPT_Online Code Hosting NXP_Software_License

4 Supported devices

Target Hosts:

- [FRDM i.MX 8M Plus](#)
- [FRDM i.MX 95](#)

5 System configuration

Supported OS:

- i.MX Embedded Linux Versions:
 - LF6.18.2_1.0.0

Supported/Required hardware:

- Supported FRDM i.MX platform
- [Ara240 DNPU](#)
- [Ara240 16GB M.2 Module](#)
- microSD >= 64 GB storage
- USB-C debug cable
- Internet (Ethernet)
- HDMI monitor
- 100 W power supply ([Anker Charger](#) is recommended)
- USB mouse
- USB keyboard
- 1080p USB camera

6 Release contents

6.1 Ara2 runtime SDK package contents

Package: rt-sdk-ara2

Version: 2.0.4

The `rt-sdk-ara2` Debian package provides a complete runtime environment for AI/ML acceleration using the Ara240 NPU on i.MX 8M Plus and i.MX 95.

This package includes:

- Runtime libraries for Ara240 discrete NPU integration
- Python bindings (DVAPI) for custom inference applications
- Optimum-Ara framework for LLMs and VLMs
- GStreamer plugins for real-time detection object applications
- Helper scripts for monitoring, benchmarking, and model management
- Systemd service for automatic hardware initialization

- UIO DMA driver for PCIe communication

6.2 Ara2 vision example release contents

Package: ara2-vision-examples

Version: 1.0

Comprehensive vision example demonstrating real-time multi-stream object detection using YOLOv8 models accelerated by Ara240 DNPU on FRDM i.MX 8M Plus development board and FRDM i.MX 95 development board.

Key features:

- Multi-stream processing: Supports 1-8 concurrent video streams with real-time object detection
- Hardware acceleration: Leverages Ara240 DNPU for efficient YOLOv8 detection inference
- GStreamer integration: Robust pipeline management with hardware-accelerated video decoding
- Visual output: Mosaic display with bounding boxes for detected objects

Supported platforms:

- FRDM i.MX 8M Plus (up to 35 FPS single stream, 15 FPS with 8 streams)
- FRDM i.MX 95 (up to 55 FPS single stream, 24 FPS with 8 streams)

Models supported:

- [nxp/YOLOv8](#)
 - YOLOv8n (nano)
 - YOLOv8s (small)
 - YOLOv8m (medium)
 - YOLOv8l (large)
 - YOLOv8x (extralarge)

Target applications:

- Multi-camera surveillance systems
- Industrial monitoring and quality control
- Smart city infrastructure
- Retail analytics
- Autonomous vehicle testing

License: NXP Proprietary (LA_OPT_Online Code Hosting)

6.3 LLM Edge Studio release contents

Package: llm-edge-studio

Version: 2.0.0

Large Language Model (LLM) Edge Studio is a Qt/QML-based GUI application designed for NXP's FRDM i.MX platforms (8M Plus and i.MX 95). It interacts with large language models accelerated by the Ara240 DNPU at the edge.

Key features:

- User-friendly GUI for model selection and prompt submission
- Real-time streaming token responses via Server-Sent Events (SSE)
- Integration with eIQ AAF connector for HTTP-based communication

- Support for Qwen2.5 models (7B-Instruct and Coder-1.5B)
- Model loading/unloading with progress tracking
- Performance metrics display (TTFT, TPS)

Architecture:

- Frontend: Qt 6.5+ with QML for responsive UI
- Backend: C++ with AAF Connector for REST API communication
- Communication: HTTP POST/GET with SSE streaming
- Target: Embedded Linux (>= LF6.18.2_1.0.0)

Supported platforms:

- FRDM i.MX 8M Plus
- FRDM i.MX 95

Models supported:

- [nxp/Qwen2.5-Coder-1.5B-Ara240](#)
- [nxp/Qwen2.5-7B-Instruct-Ara240](#)

License: NXP Proprietary (LA_OPT_Online Code Hosting)

6.4 VLM Edge Studio release contents

Package: vlm-edge-studio

Version: 1.0.0

Vision Language Model (VLM) Edge Studio is a Qt/QML-based GUI application designed for NXP's FRDM i.MX (8M Plus and i.MX 95). It interacts with vision language models accelerated by the Ara240 discrete NPU at the edge.

Key features:

- User-friendly GUI for model selection and prompt submission
- Real-time streaming token responses via Server-Sent Events (SSE)
- Integration with eIQ AAF connector for HTTP-based communication
- Model loading/unloading with progress tracking
- Performance metrics display (TTFT, TPS)
- Support for Qwen2.5 VL 7B model
- Video or image mode analysis support
- USB-C type camera support for video recording/image capture for vision analysis

Architecture:

- Frontend: Qt 6.5+ with QML for responsive UI
- Backend: C++ with AAF connector for REST API communication
- Communication: HTTP POST/GET with SSE streaming
- Target: Embedded Linux (>= LF6.18.2_1.0.0)

Supported platforms:

- FRDM i.MX 8M Plus
- FRDM i.MX 95

Models supported:

- [nxp/Qwen2.5-VL-7B-Instruct-Ara240](#)

License: NXP Proprietary (LA_OPT_Online Code Hosting)

6.5 eIQ AAF connection release contents

Package: eiq-aaf-connector

Version: 2.0

eIQ AAF connector — OpenAI compatible server for LLMs. The eIQ AAF connector provides a REST-based server for LLM inference on Ara240 discrete NPU. It offers OpenAI-compatible APIs for language model interactions on i.MX devices.

Supported platforms:

- FRDM i.MX 8M Plus
- FRDM i.MX 95

License: NXP Proprietary (LA_OPT_Online Code Hosting)

7 Known issues

7.1 Ara240 firmware upgrade on FRDM i.MX 8M Plus

Description:

On FRDM i.MX 8M PLUS during firmware updates of the ARA240 DNPU using the script `program_flash.sh` on BSP 6.18.2_1.0.0 (Q1'2026), running the firmware update may cause the board to freeze and fail the flashing process.

Workaround:

1. Install Q1'2026 BSP 6.18.2_1.0.0 (as normal).
2. If `program_flash.sh` execution freezes, power off the FRDM i.MX board.
3. To stop at U-Boot, reboot and press any key.
4. Append `pcie_aspm=off` to the `mmcargs`, save setting using `saveenv`, and then continue boot.
5. Power cycle the FRDM i.MX 8M Plus board and continue firmware upgrade using steps in [README](#).

```
u-boot=> editenv mmcargs
edit: setenv bootargs ${jh_clk} ${mcore_clk} console=${console} root=
${mmccroot} pcie_aspm=off
u-boot => saveenv
u-boot => boot
```

7.2 FRDM i.MX 8MP and FRDM i.MX 95 freeze after firmware upgrade

Description

After updating ARA240 firmware and rebooting, the board freezes during boot.

Mitigation

Doing a power cycle after this condition happens, the board should boot correctly this time.

7.3 Model load stalls at ~95 % in VLM-Edge-Studio and LLM-Edge-Studio

Description

When loading a model in VLM Edge Studio or LLM Edge Studio, the progress bar may halt around 95 % and never complete.

This can occur due to two independent root causes:

1. **SSL certificate verification failure due to no Internet connectivity**

As part of the model load sequence, `eiq-connector` attempts to download/update Qwen2.5 VL resources by making HTTPS requests to the model's source repository. If the host system has no Internet connectivity, the `eiq-connector` cannot validate or retrieve metadata, causing the load to stall silently.

Workaround:

Ensure that the system is connected to the Internet and has the correct system time before loading a model.

2. **Model not fully cleaned up when switching between applications**

If you run `vlm-edge-studio` and then open `llm-edge-studio` (or vice versa), closing the first application does not always cause `eiq-connector` to fully unload the previously loaded model. When the second application attempts to load a new model, the connector may partially reject the load, again causing a 95 % stall.

Workaround:

Before switching between VLM and LLM Edge Studio applications, reboot the system to ensure that the connector starts clean.

Recommended combined action

To avoid the 95 % stall entirely, apply both workarounds together:

1. Reboot the system before switching between Edge Studio applications.
2. Ensure that the machine has Internet access and the correct system time before loading a model.

This gives the highest reliability until the underlying issues are fixed. As an extra precaution, monitor `/tmp/connector.log` for any error message.

7.4 `ara2_metrics` script failure on FRDM i.MX 8MP and FRDM i.MX 95

Description

Running `ara2_metrics.sh` and selecting option 6 (Utilization, Temperature, DRAM continuously) causes an immediate segmentation fault in the underlying binary `hw_metrics.out`. The issue appears after the first temperature printout.

Expected behavior

The script should start printing periodically the information.

Workaround

Wait one second after the menu is fully displayed.

7.5 Silent hugging face model download failures under poor network conditions

When installing either `llm-edge-studio` or `vlm-edge-studio`, the installation triggers the `fetch_models` script from `rt-sdk-ara2`. The download process appears to complete normally, but the model files, `model.dvm`, are downloaded incompletely. As a result, the model fails to load or initialize during runtime.

Observed behavior

- Download process shows no errors.
- Runtime components fail when trying to load the corrupted/incomplete model.

Expected behavior

- All model files should download completely and load successfully after installation.

Workaround

Do a fresh installation of the package by removing `/usr/share/llm/<AFECTED_MODEL>`, purging the package with `dpkg -P`, and trying again. Lastly, as a double check you can manually compute the `md5sum` and compare it with the one under `/usr/share/llm/<MODEL>/model.json`.

8 Revision history

[Table 3](#) summarizes revisions to this document.

Table 3. Revision history

Document ID	Release date	Description
RN00459 v.1.0	13 April 2026	Initial public release

Legal information

Definitions

Draft — A draft status on a document indicates that the content is still under internal review and subject to formal approval, which may result in modifications or additions. NXP Semiconductors does not give any representations or warranties as to the accuracy or completeness of information included in a draft version of a document and shall have no liability for the consequences of use of such information.

Disclaimers

Limited warranty and liability — Information in this document is believed to be accurate and reliable. However, NXP Semiconductors does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information and shall have no liability for the consequences of use of such information. NXP Semiconductors takes no responsibility for the content in this document if provided by an information source outside of NXP Semiconductors.

In no event shall NXP Semiconductors be liable for any indirect, incidental, punitive, special or consequential damages (including - without limitation - lost profits, lost savings, business interruption, costs related to the removal or replacement of any products or rework charges) whether or not such damages are based on tort (including negligence), warranty, breach of contract or any other legal theory.

Notwithstanding any damages that customer might incur for any reason whatsoever, NXP Semiconductors' aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms and conditions of commercial sale of NXP Semiconductors.

Right to make changes — NXP Semiconductors reserves the right to make changes to information published in this document, including without limitation specifications and product descriptions, at any time and without notice. This document supersedes and replaces all information supplied prior to the publication hereof.

Suitability for use — NXP Semiconductors products are not designed, authorized or warranted to be suitable for use in life support, life-critical or safety-critical systems or equipment, nor in applications where failure or malfunction of an NXP Semiconductors product can reasonably be expected to result in personal injury, death or severe property or environmental damage. NXP Semiconductors and its suppliers accept no liability for inclusion and/or use of NXP Semiconductors products in such equipment or applications and therefore such inclusion and/or use is at the customer's own risk.

Applications — Applications that are described herein for any of these products are for illustrative purposes only. NXP Semiconductors makes no representation or warranty that such applications will be suitable for the specified use without further testing or modification.

Customers are responsible for the design and operation of their applications and products using NXP Semiconductors products, and NXP Semiconductors accepts no liability for any assistance with applications or customer product design. It is customer's sole responsibility to determine whether the NXP Semiconductors product is suitable and fit for the customer's applications and products planned, as well as for the planned application and use of customer's third party customer(s). Customers should provide appropriate design and operating safeguards to minimize the risks associated with their applications and products.

NXP Semiconductors does not accept any liability related to any default, damage, costs or problem which is based on any weakness or default in the customer's applications or products, or the application or use by customer's third party customer(s). Customer is responsible for doing all necessary testing for the customer's applications and products using NXP Semiconductors products in order to avoid a default of the applications and the products or of the application or use by customer's third party customer(s). NXP does not accept any liability in this respect.

Limiting values — Stress above one or more limiting values (as defined in the Absolute Maximum Ratings System of IEC 60134) will cause permanent damage to the device. Limiting values are stress ratings only and (proper) operation of the device at these or any other conditions above those given in the Recommended operating conditions section (if present) or the Characteristics sections of this document is not warranted. Constant or repeated exposure to limiting values will permanently and irreversibly affect the quality and reliability of the device.

Terms and conditions of commercial sale — NXP Semiconductors products are sold subject to the general terms and conditions of commercial sale, as published at <https://www.nxp.com/profile/terms>, unless otherwise agreed in a valid written individual agreement. In case an individual agreement is concluded only the terms and conditions of the respective agreement shall apply. NXP Semiconductors hereby expressly objects to applying the customer's general terms and conditions with regard to the purchase of NXP Semiconductors products by customer.

No offer to sell or license — Nothing in this document may be interpreted or construed as an offer to sell products that is open for acceptance or the grant, conveyance or implication of any license under any copyrights, patents or other industrial or intellectual property rights.

Quick reference data — The Quick reference data is an extract of the product data given in the Limiting values and Characteristics sections of this document, and as such is not complete, exhaustive or legally binding.

Export control — This document as well as the item(s) described herein may be subject to export control regulations. Export might require a prior authorization from competent authorities.

Suitability for use in non-automotive qualified products — Unless this document expressly states that this specific NXP Semiconductors product is automotive qualified, the product is not suitable for automotive use. It is neither qualified nor tested in accordance with automotive testing or application requirements. NXP Semiconductors accepts no liability for inclusion and/or use of non-automotive qualified products in automotive equipment or applications.

In the event that customer uses the product for design-in and use in automotive applications to automotive specifications and standards, customer (a) shall use the product without NXP Semiconductors' warranty of the product for such automotive applications, use and specifications, and (b) whenever customer uses the product for automotive applications beyond NXP Semiconductors' specifications such use shall be solely at customer's own risk, and (c) customer fully indemnifies NXP Semiconductors for any liability, damages or failed product claims resulting from customer design and use of the product for automotive applications beyond NXP Semiconductors' standard warranty and NXP Semiconductors' product specifications.

HTML publications — An HTML version, if available, of this document is provided as a courtesy. Definitive information is contained in the applicable document in PDF format. If there is a discrepancy between the HTML document and the PDF document, the PDF document has priority.

Translations — A non-English (translated) version of a document, including the legal information in that document, is for reference only. The English version shall prevail in case of any discrepancy between the translated and English versions.

Security — Customer understands that all NXP products may be subject to unidentified vulnerabilities or may support established security standards or specifications with known limitations. Customer is responsible for the design and operation of its applications and products throughout their lifecycles to reduce the effect of these vulnerabilities on customer's applications and products. Customer's responsibility also extends to other open and/or proprietary technologies supported by NXP products for use in customer's applications. NXP accepts no liability for any vulnerability. Customer should regularly check security updates from NXP and follow up appropriately.

Customer shall select products with security features that best meet rules, regulations, and standards of the intended application and make the ultimate design decisions regarding its products and is solely responsible for compliance with all legal, regulatory, and security related requirements concerning its products, regardless of any information or support that may be provided by NXP.

NXP has a Product Security Incident Response Team (PSIRT) (reachable at PSIRT@nxp.com) that manages the investigation, reporting, and solution release to security vulnerabilities of NXP products.

NXP B.V. — NXP B.V. is not an operating company and it does not distribute or sell products.

Trademarks

Notice: All referenced brands, product names, service names, and trademarks are the property of their respective owners.

NXP — wordmark and logo are trademarks of NXP B.V.

eIQ — is a trademark of NXP B.V.

Contents

1	Overview	2
2	Features	2
2.1	Models	2
3	Licenses	2
4	Supported devices	3
5	System configuration	3
6	Release contents	3
6.1	Ara2 runtime SDK package contents	3
6.2	Ara2 vision example release contents	4
6.3	LLM Edge Studio release contents	4
6.4	VLM Edge Studio release contents	5
6.5	eIQ AAF connection release contents	6
7	Known issues	6
7.1	Ara240 firmware upgrade on FRDM i.MX 8M Plus	6
7.2	FRDM i.MX 8MP and FRDM i.MX 95 freeze after firmware upgrade	6
7.3	Model load stalls at ~95 % in VLM-Edge- Studio and LLM-Edge-Studio	6
7.4	ara2_metrics script failure on FRDM i.MX 8MP and FRDM i.MX 95	7
7.5	Silent hugging face model download failures under poor network conditions	7
8	Revision history	8
	Legal information	9

Please be aware that important notices concerning this document and the product(s) described herein, have been included in section 'Legal information'.
