



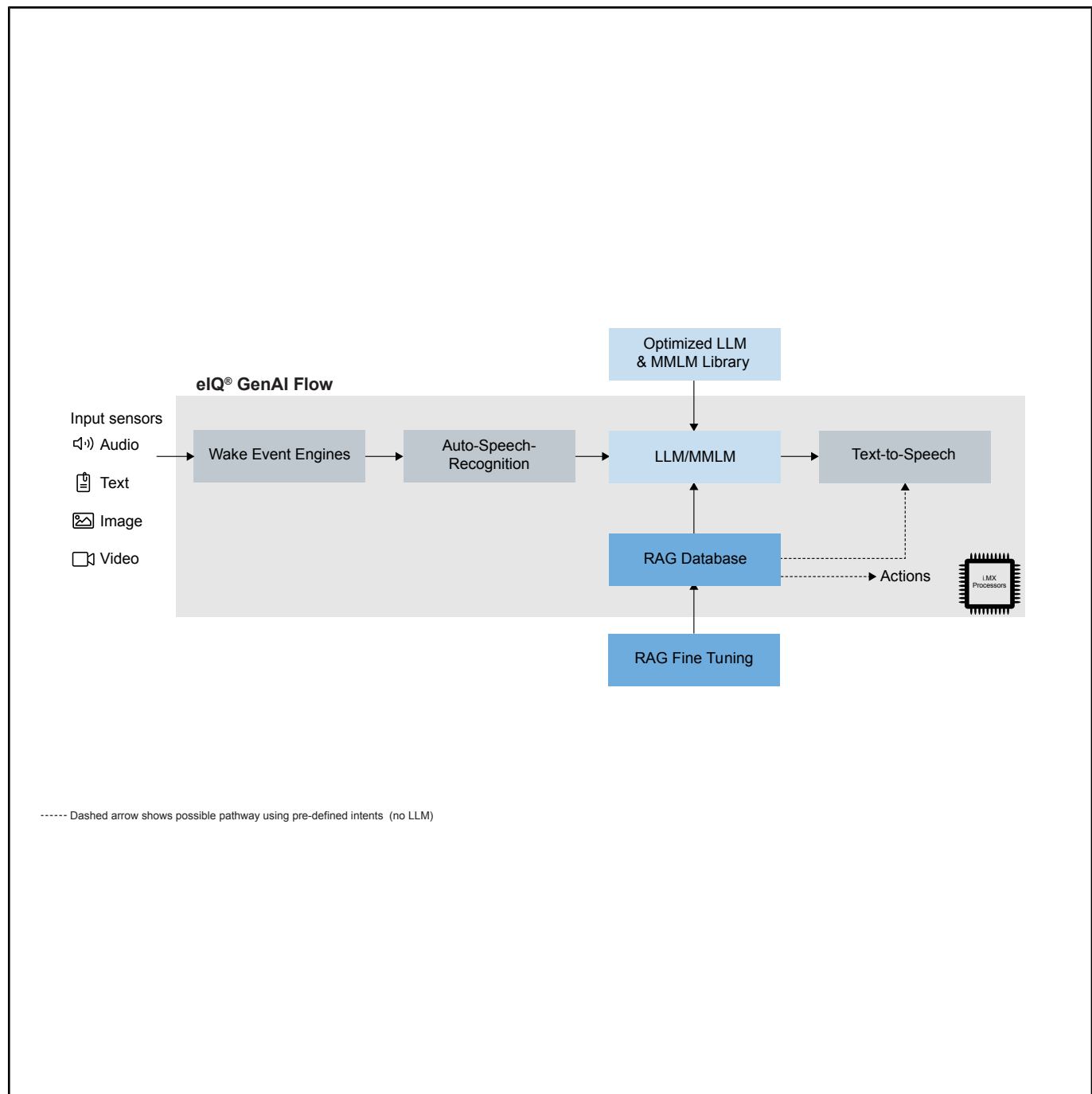
# **Simplified and Optimized Generative AI at the Edge with eIQ GenAI Flow**

## **GEN-AI-FLOW**

Last Updated: Nov 7, 2025

eIQ Gen-AI Flow is a modular, end-to-end software pipeline optimized to bring Large and Small Language Models (LLMs and SLMs) to NXP applications processors like the i.MX 95, 93 and 8M Plus. The pipeline includes wake word, speech-to-text, text-to-speech and Gen-AI models to run on device CPU and/or NPU. Gen-AI Flow also includes Retrieval Augmented Generation (RAG) to securely and easily fine-tune LLMs with custom, private knowledge sources like PDFs.

## Bringing Generative AI to the Edge Block Diagram



View additional information for [Simplified and Optimized Generative AI at the Edge with eIQ GenAI Flow](#).

**Note:** The information on this document is subject to change without notice.

**[www.nxp.com](http://www.nxp.com)**

NXP and the NXP logo are trademarks of NXP B.V. All other product or service names are the property of their respective owners. The related technology may be protected by any or all of patents, copyrights, designs and trade secrets. All rights reserved. © 2025 NXP B.V.