

Deploying AI to Embedded and Enterprise Systems

Kyu-Hwan Jang ¹

1) *Application Engineering Group, MathWorks*

Corresponding Author: Kyu-Hwan Jang, pauljang@mathworks.com

ABSTRACT

Deploying AI raises challenges beyond those associated with developing a performant AI model, including:

- Meeting hardware constraints of the deployment environment, such as limited memory and power consumption
- Monitoring and maintaining model performance over their lifetime

Learn about expanded capabilities to address the above challenges for both compiler-based and embedded deployment using code generation:

1. Quantization: Fixed-point conversion for machine learning models and quantization for deep neural networks allow them to fit on hardware with limited memory and power.
2. Incremental learning and model updates: Code generation that separates parameters from prediction code and incremental learning make it possible to improve models continuously. DevOps provides a framework for managing and governing AI models across their life cycle.