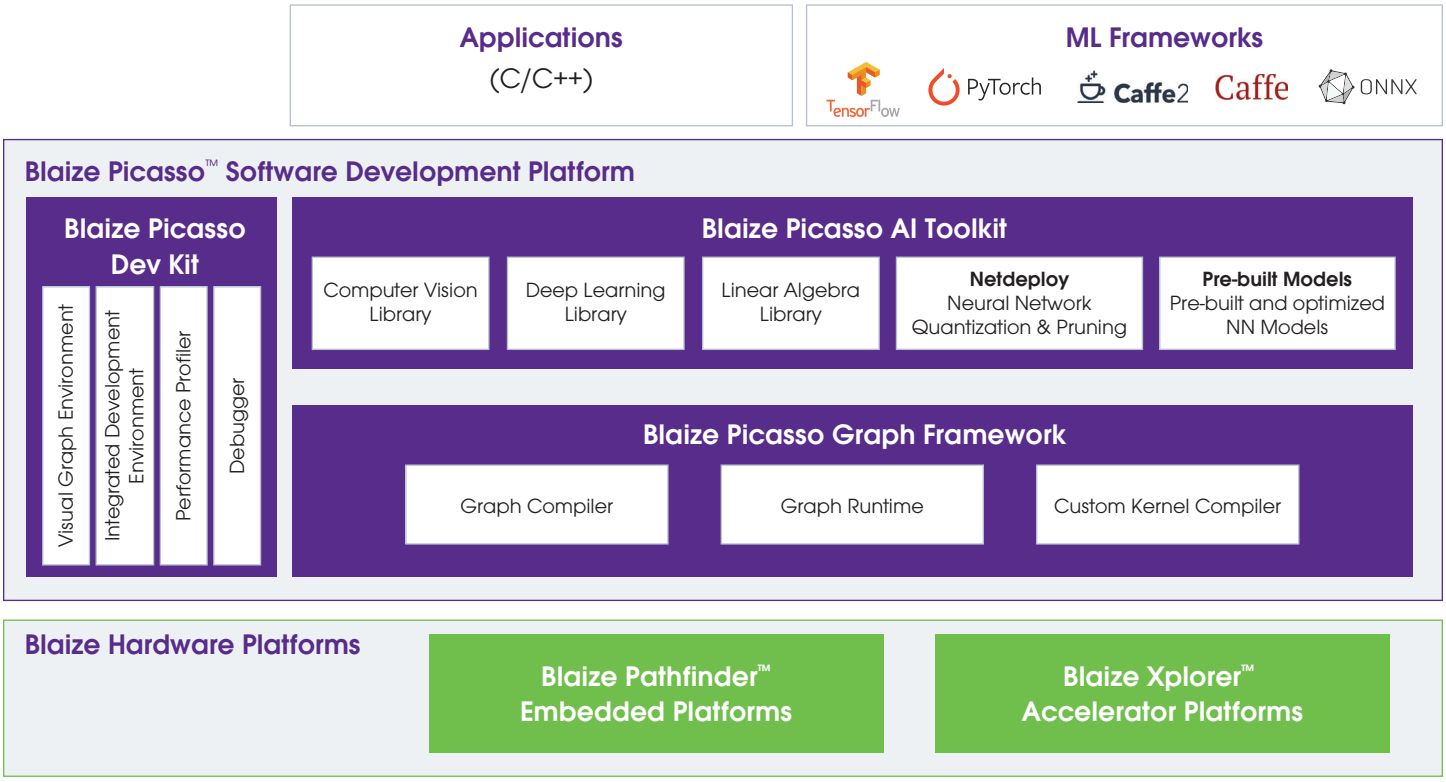
Blaize Picasso™ Software Development Platform for Graph Streaming Processors (GSP)

Graph-native Software Platform

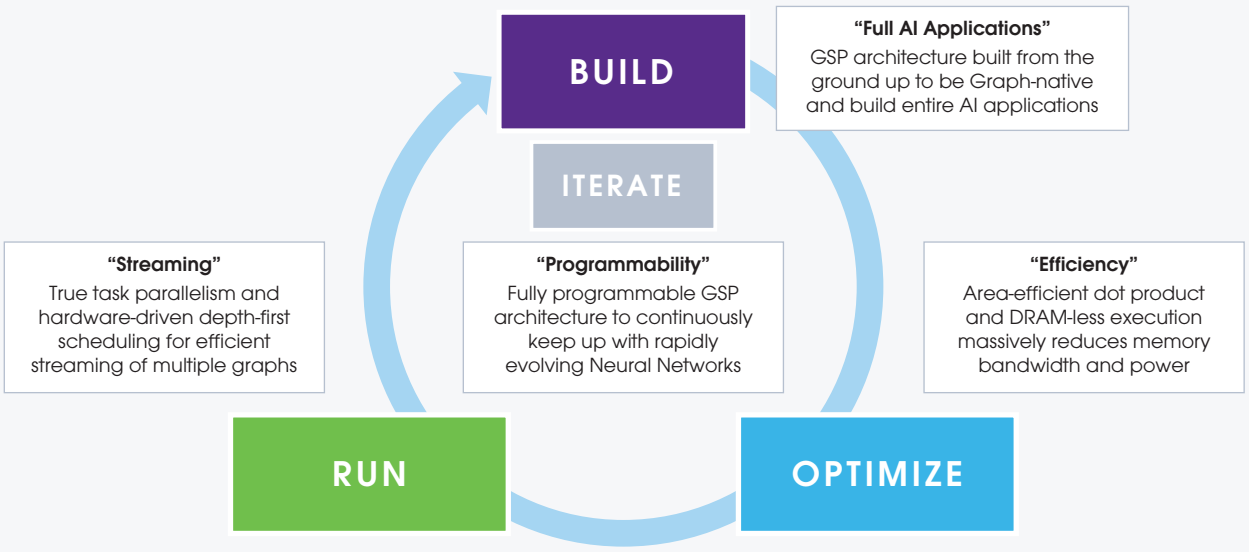
Efficiently Build and Optimize Complete Artificial Intelligence Applications

Blaize Graph Streaming Processor™, or GSP, is a new architectural paradigm for building and accelerating Artificial Intelligence (AI) applications. In order to take full advantage of this new architecture, Blaize has developed the industry's first truly graph-native software platform that enables customers to build and optimize many AI application from end-to-end as well as deep in hardware.



Blaize's Picasso™ Software Development Platform enables customers to accelerate their entire AI Application development cycle—build, integrate, optimize, run and continuous improvement.

Graph Streaming Processor Architecture Enables Graph-Native AI Application Development



Blaize's Picasso Platform consists of three major components:

- Blaize Picasso AI Toolkit
- Blaize Picasso Graph Framework
- Blaize Picasso Dev Kit

Blaize Picasso AI Toolkit: Libraries and Tools to Build Entire AI Apps

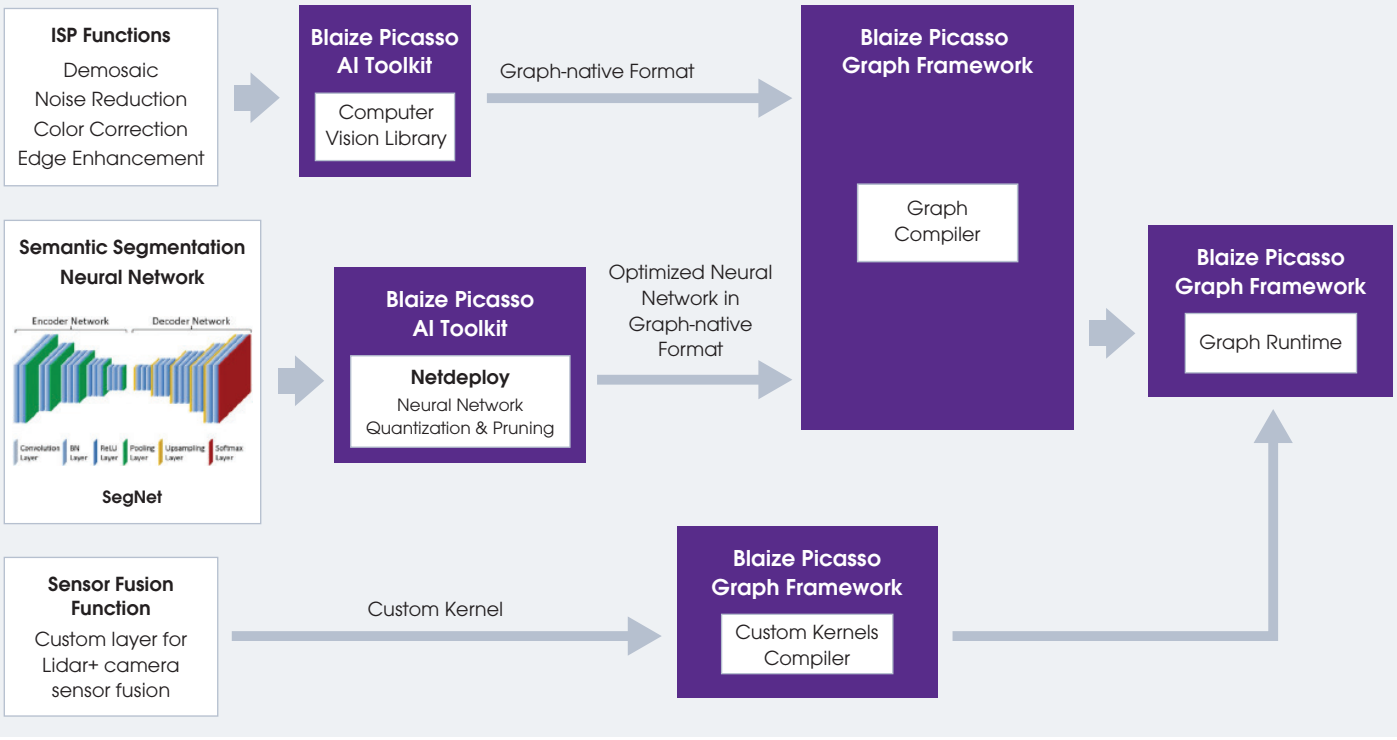
Blaize Picasso AI Toolkit consists of libraries and tools to build entire AI Applications, including neural network and non-neural network functions. AI application pipelines typically include both neural network operations for object detection, classification, segmentation, etc. as well as functions for Sensor Fusion and Image Signal Processing (ISP) such as noise filtering, color conversion, etc. The Blaize Picasso AI toolkit consists of multiple libraries for Computer Vision, Deep Learning and Linear Algebra to build these non-neural network functions in a graph-native manner. Other existing functions, written in C/C++, can be readily converted into a graph-native format with full GSP hardware acceleration.

NetDeploy: Automated Optimization and Compression

Neural network functions, which are typically trained in Machine Learning frameworks such as Tensorflow or Pytorch, need to be optimized to run efficiently on hardware. Blaize NetDeploy tool automates the optimization and compression process through techniques such as quantization, block sparsity induction and pruning. The core feature of NetDeploy is its ability to perform hardware-aware optimization of neural networks by balancing hardware performance with neural network model accuracy. Using sensitivity analysis, NetDeploy determines which operations are good candidates for compression, selectively removes minimally-contributing parameters, evaluates the resulting

The different tools in the Blaize Picasso Software Development Platform enable developers to build, optimize and run this complete automotive sensor fusion application on the GSP.

Automotive Sensor Fusion Software Workflow on Blaize Picasso Software Development Platform



The Graph-native nature of the GSP architecture enables customers to run this entire sensor fusion application, both neural network and non-neural network functions, efficiently on the GSP cores in a streaming fashion, providing great advantages in terms of performance and power efficiency. Further, as the overall sensor fusion workflow and neural network architectures evolve, customers can easily update the entire application, taking advantage of the extensive software tools in the Blaize Picasso Platform and the programmable nature of the GSP architecture.

Corporate Headquarters
4370 Town Center Blvd
Suite 240
El Dorado Hills, CA 95762 USA
T: +1 916 347-0050
info@blaize.com

India Office
Block 2, 4th Floor,
DLF Cybercity, APHB Colony,
Gachibowli
Hyderabad-500 032

UK Office
Suite 1, Concept House
Home Park Road
Kings Langley, Herts
WD4 8UD UK