SNAPDRAGON RIDE™ SoCs
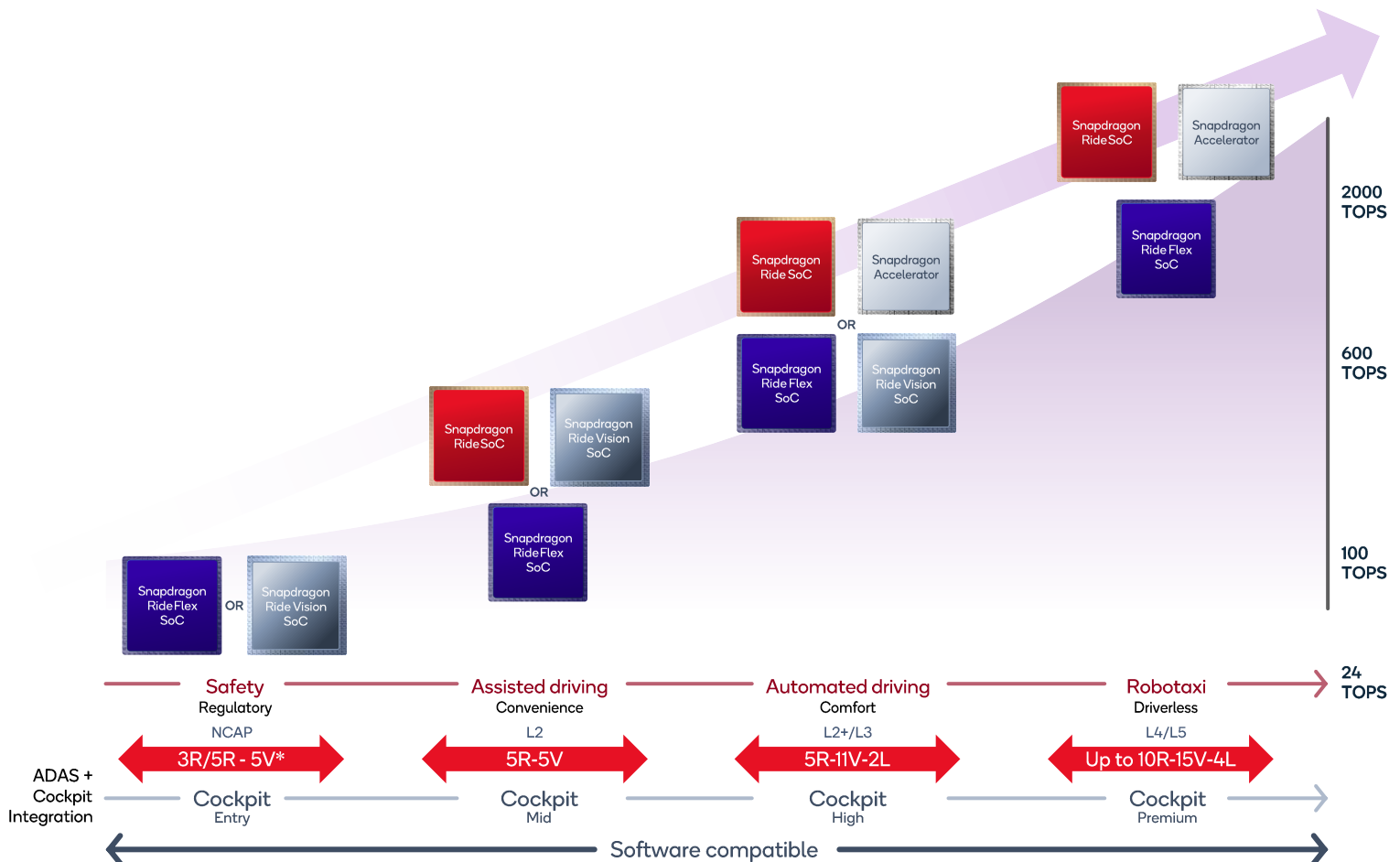
A family of multipurpose, power-efficient SOCs and software designed for ADAS system scalability, safety and reliability.
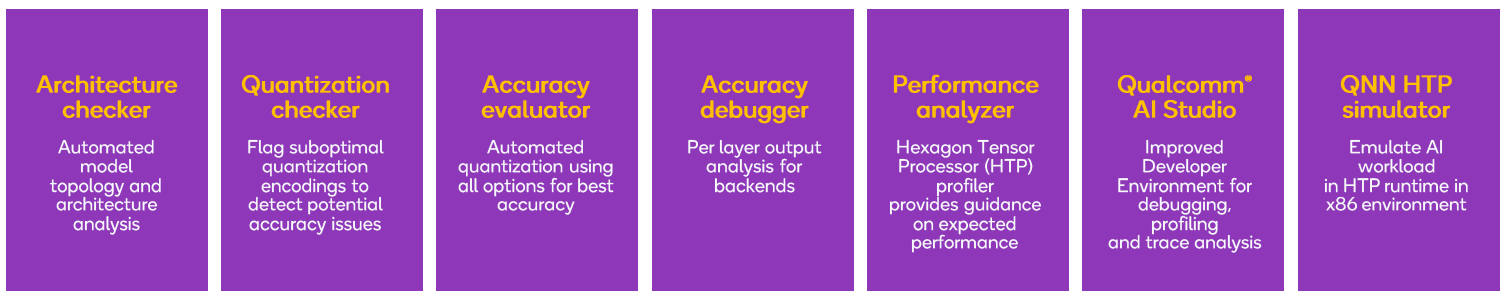
Snapdragon

MOST COMPLETE AND SCALABLE SYSTEM PORTFOLIO

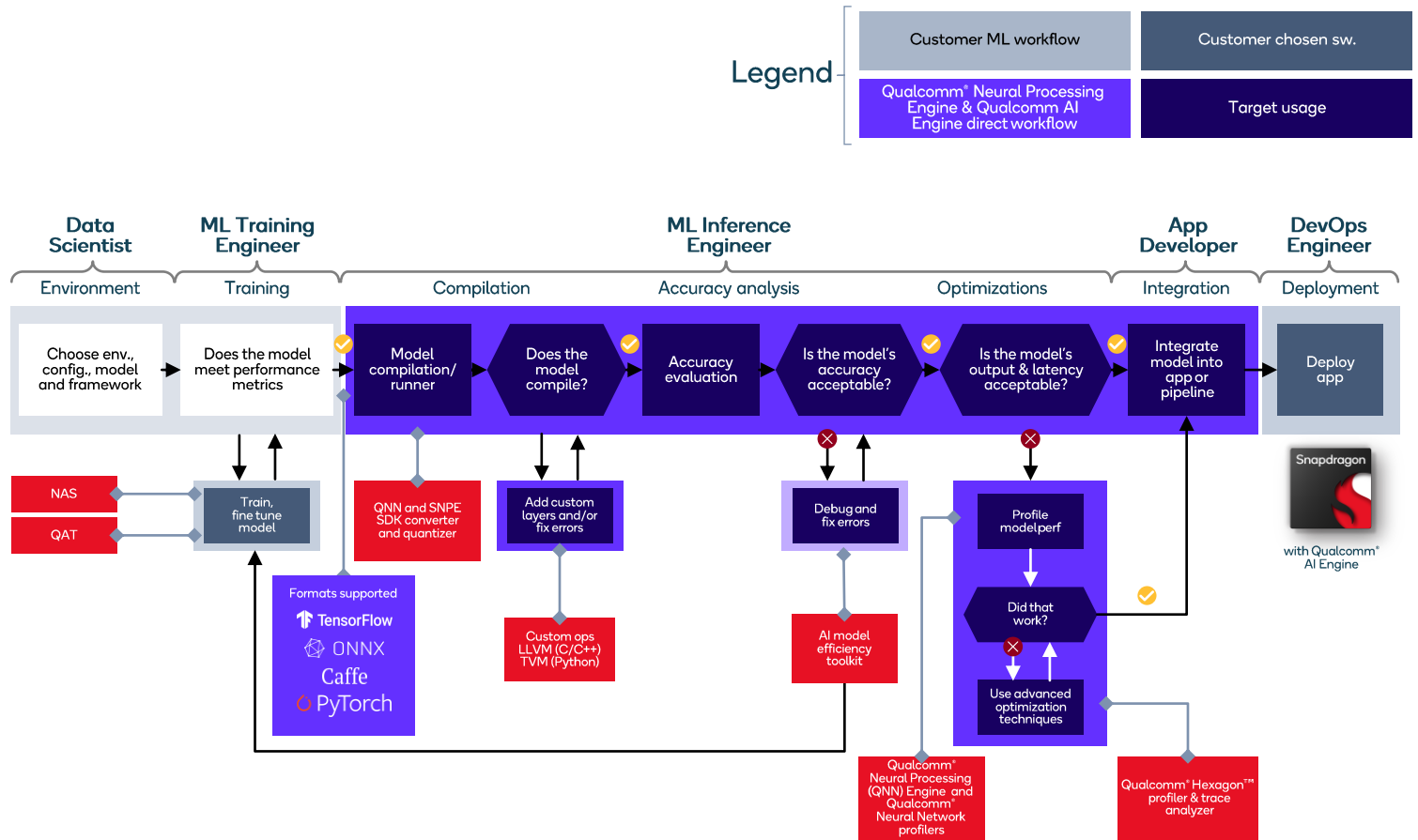|	Safety	Assisted driving	Automated driving	Robotaxi
	Regulatory	Convenience	Comfort	Driverless
	NCAP	L2	L2+/L3	L4/L5
	3R/5R - 5V*	5R-5V	5R-11V-2L	Up to 10R-15V-4L
ADAS + Cockpit Integration	Cockpit Entry	Cockpit Mid	Cockpit High	Cockpit Premium

Software compatible
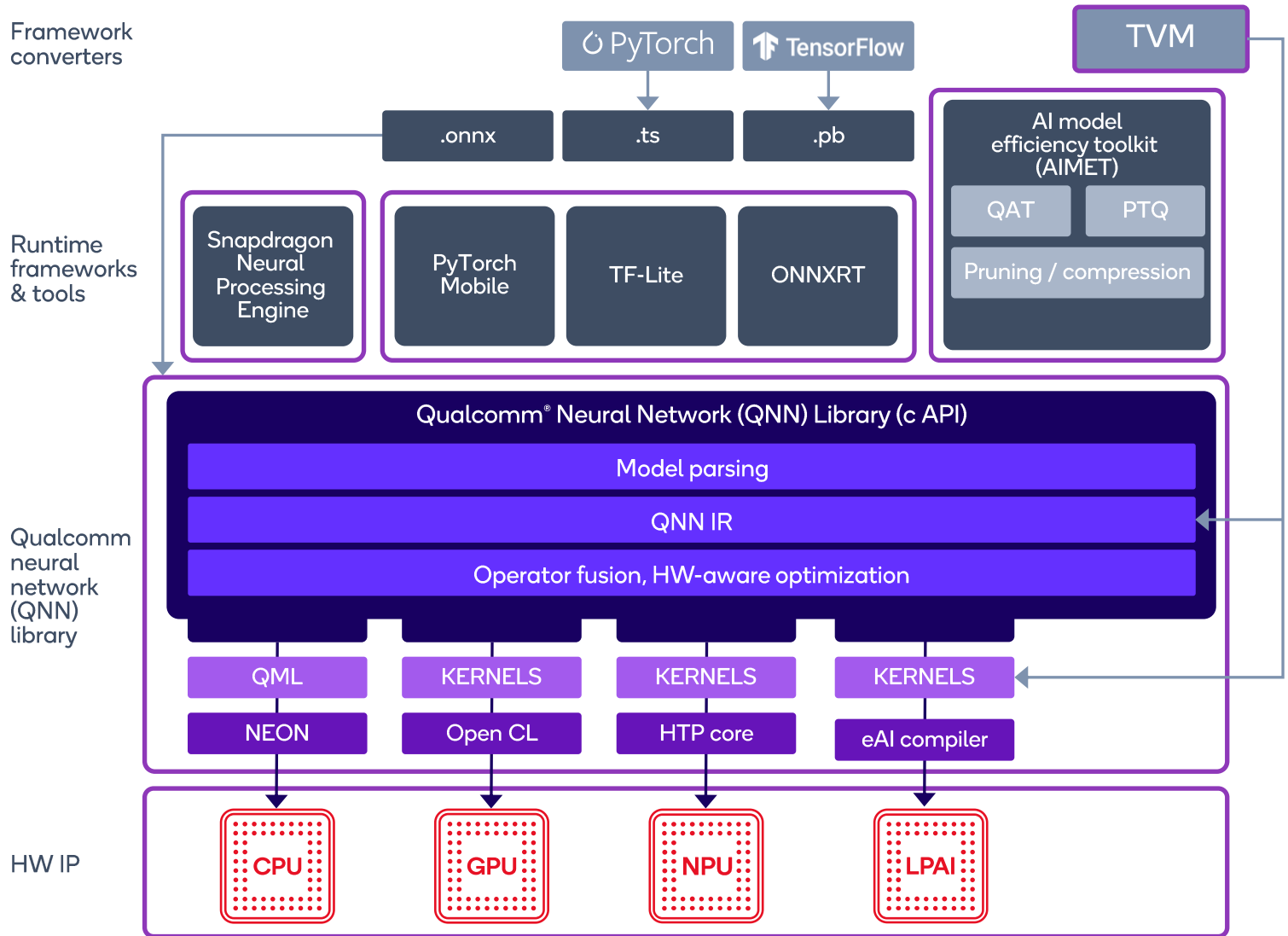
TOPS scale: 24 TOPS, 100 TOPS, 600 TOPS, 2000 TOPS

SoCs that scale with your system design, interface requirements and performance needs.

These high performance SOCs have the best power/inference on the market with multiple benchmarks indicating 2x power efficiency and performance for AI networks over other solutions.

DEVELOPER TOOLS TO ACCELERATE AI WORKFLOWS

Legend

Customer ML workflow	Customer chosen sw.
Qualcomm® Neural Processing Engine & Qualcomm AI Engine direct workflow	Target usage

Roles

Data Scientist	ML Training Engineer	ML Inference Engineer				App Developer	DevOps Engineer
Environment	Training	Compilation	Accuracy analysis		Optimizations	Integration	Deployment

Choose env., config., model and framework → Does the model meet performance metrics → Model compilation/runner → Does the model compile? → Accuracy evaluation → Is the model's accuracy acceptable? → Is the model's output & latency acceptable? → Integrate model into app or pipeline → Deploy app

NAS
QAT
Train, fine tune model

QNN and SNPE SDK converter and quantizer

Add custom layers and/or fix errors

Custom ops LLVM (C/C++) TVM (Python)

Debug and fix errors

AI model efficiency toolkit

Profile model perf

Did that work?

Use advanced optimization techniques

Qualcomm® Neural Processing (QNN) Engine and Qualcomm® Neural Network profilers

Qualcomm® Hexagon™ profiler & trace analyzer

Formats supported: TensorFlow, ONNX, Caffe, PyTorch

Snapdragon — with Qualcomm® AI Engine

Tools

Architecture checker
Automated model topology and architecture analysis

Quantization checker
Flag suboptimal quantization encodings to detect potential accuracy issues

Accuracy evaluator
Automated quantization using all options for best accuracy

Accuracy debugger
Per layer output analysis for backends

Performance analyzer
Hexagon Tensor Processor (HTP) profiler provides guidance on expected performance

Qualcomm® AI Studio
Improved Developer Environment for debugging, profiling and trace analysis

QNN HTP simulator
Emulate AI workload in HTP runtime in x86 environment

Snapdragon Ride

MODULAR AND EXTENSIBLE AI TOOLS COMPATIBLE WITH COMMON PUBLIC FRAMEWORKS

Framework converters

PyTorch · TensorFlow · TVM

.onnx · .ts · .pb

Runtime frameworks & tools

Snapdragon Neural Processing Engine · PyTorch Mobile · TF-Lite · ONNXRT

AI model efficiency toolkit (AIMET)
- QAT
- PTQ
- Pruning / compression

Qualcomm neural network (QNN) library

Qualcomm® Neural Network (QNN) Library (c API)
- Model parsing
- QNN IR
- Operator fusion, HW-aware optimization

QML · KERNELS · KERNELS · KERNELS

NEON · Open CL · HTP core · eAI compiler

HW IP

CPU · GPU · NPU · LPAI

Key takeaways:

✓ Unified QNN API for neural networking

✓ Common tools and converters across Snapdragon® SoCs

✓ Customers and applications portable across SoCs

✓ Modular and extensible per-accelerator backend libraries for reusable full stack AI solutions

Snapdragon
Ride

3