

Unified Compute in RISC-V

Evolving RISC-V to address AI/ML
with Open Standards and Global Collaboration



Dr. Philipp Tomsich
Chief Technologist & Founder, VRULL GmbH

RISC-V is a unique opportunity to build a **custom AI/ML accelerators** based on an open standards ecosystem

INDUSTRY-LEADING DIFFERENTIATION

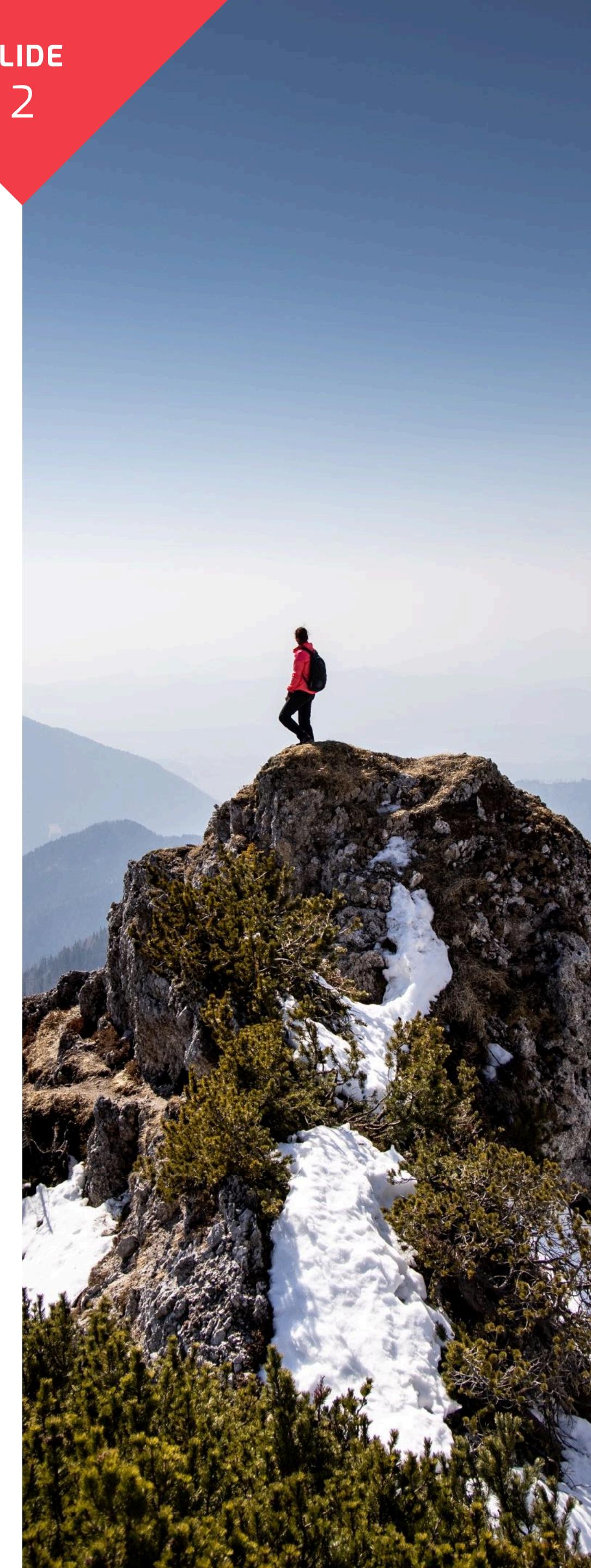
Match your RISC-V products to your application needs by adding standard extensions around a minimal base set

SUPPORTED BY ONE UNIFIED ECOSYSTEM

Build on a software ecosystem that is built around feature detection and the adaptation to differentiated implementations

DOMAIN-SPECIFIC CUSTOMISATION

Bring together the best local knowledge in the development of a global standard to define domain-specific extensions

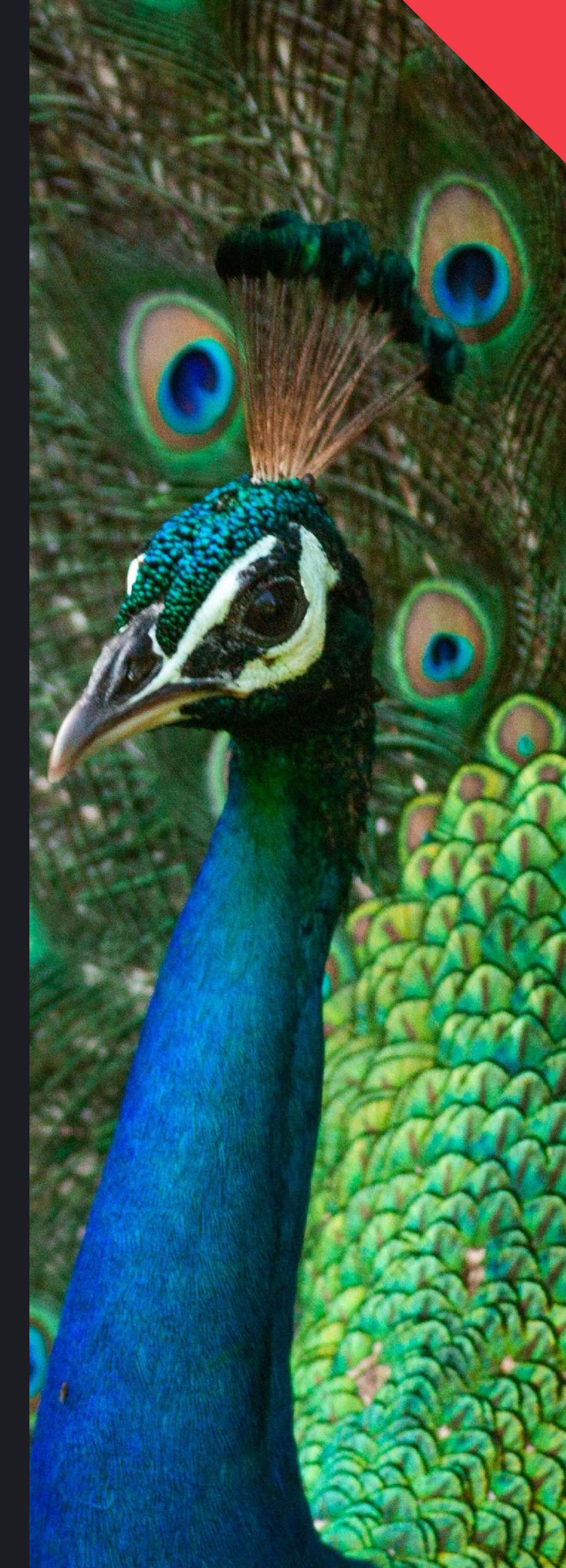


UPENDING THE ECONOMICS OF CUSTOM SILICON

ONE GLOBAL ECOSYSTEM

INDUSTRY-LEADING DIFFERENTIATION

STRENGTHS & OPPORTUNITIES



FLEXIBILITY TO CUSTOMISE

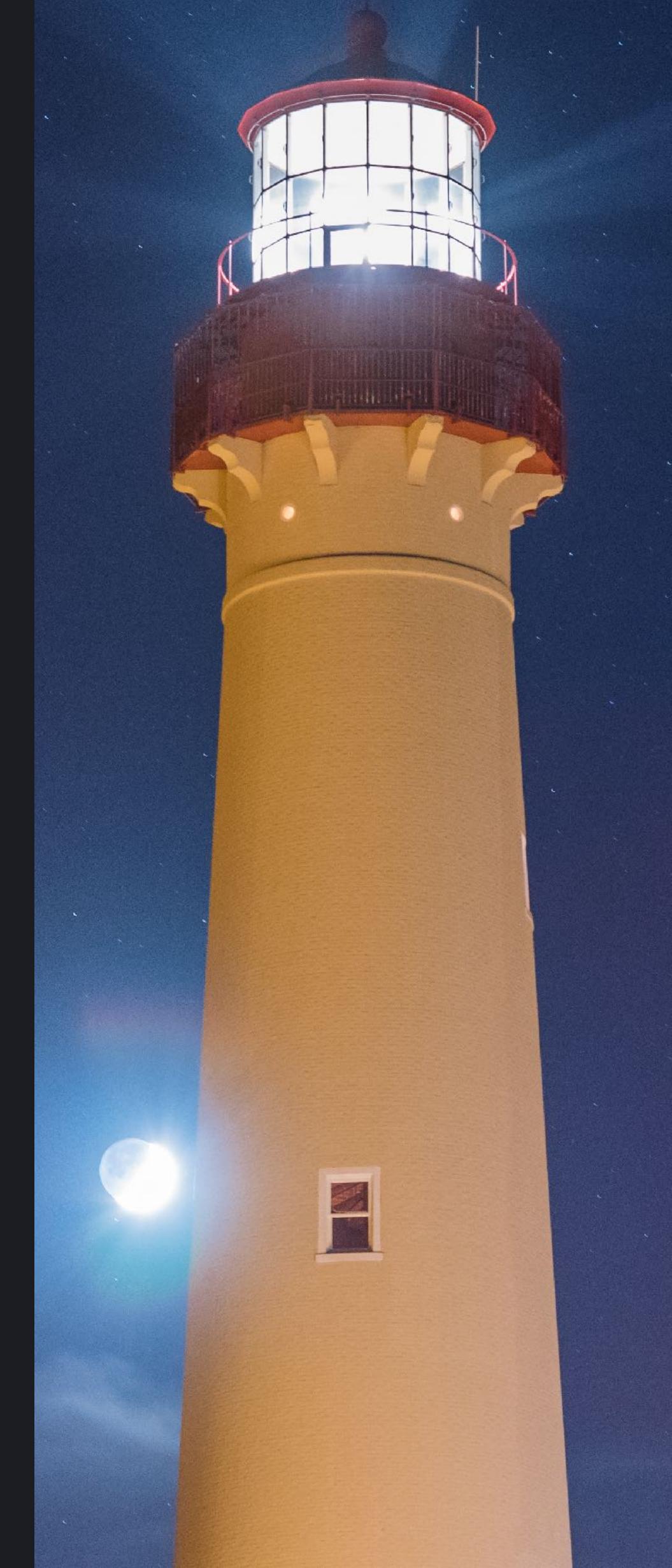
RISC-V standards are built for your freedom to customise:

- RISC-V defines small base feature sets for implementors to add only those extensions that their markets require
- RISC-V permits the addition of vendor-specified extensions to differentiate products with application-specific accelerators
- RISC-V provides guidance and branding to implementors for interoperability requirements in market-segments through the definition of "RISC-V Profiles"

RISC-V leads the industry in providing the opportunity to build customised and differentiated solutions from standard and non-standard extensions.

ONE UNIFIED ECOSYSTEM

STRENGTHS & OPPORTUNITIES



INTEROPERABLE CHOICES

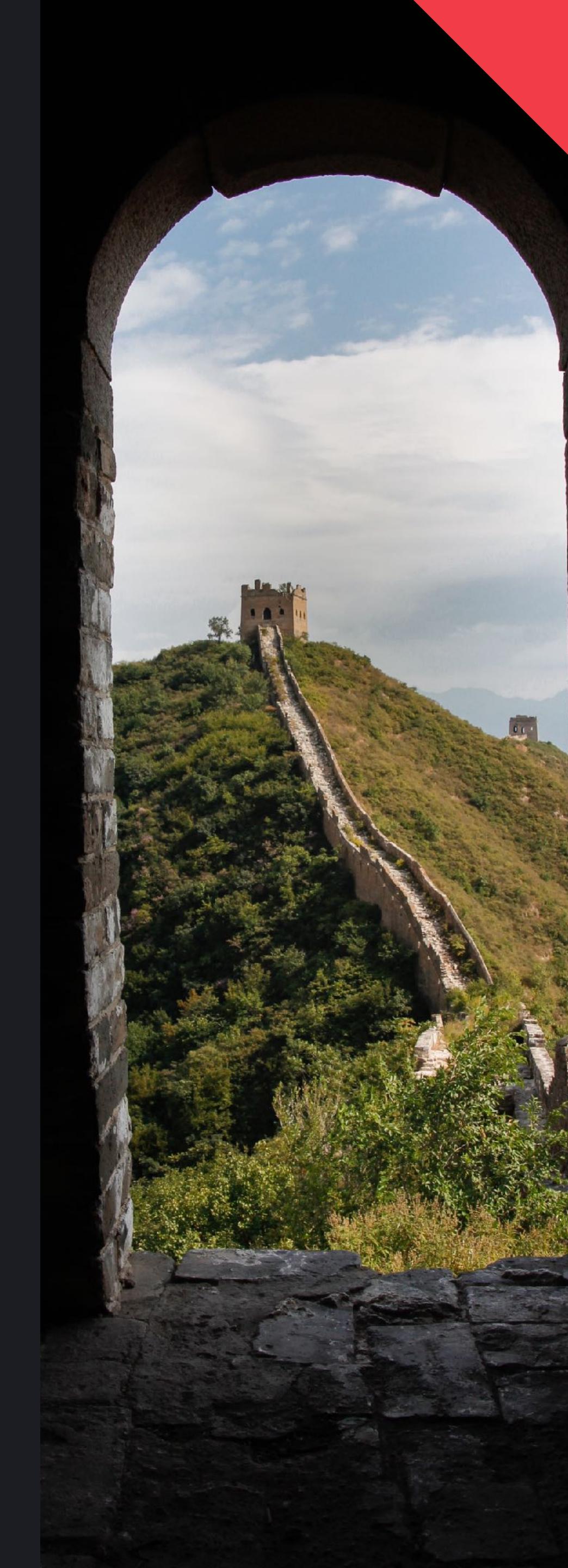
Enabling rapid innovation and differentiation means that innovators should leverage one unified ecosystem:

- Common development tools are adapted to support flexible combinations of standard and vendor-defined extensions
- Reusable software libraries and extensible software stacks support incremental innovation by providing commodity base functionality and extend it for application-specific acceleration

We carefully evolve the RISC-V software ecosystem to specify state-of-the-art feature detection mechanisms and to ensure the flexible adaptation to different application profiles.

DOMAIN-SPECIFIC CUSTOMISATION

STRENGTHS & OPPORTUNITIES



ADAPTED TO APPLICATIONS

RISC-V International offers a neutral forum that brings together the best local and domain-specific knowledge to develop the standards that will define tomorrow's computing solutions:

- High-performance computing
- Artificial intelligence and machine learning extensions
- Secure and confidential computing
- Control-flow integrity and runtime integrity extensions
- Functional safety extensions

Bringing together experts from different geographies and industries, RISC-V can quickly evolve its standards to address emerging applications and opportunities.

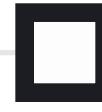
RISC-V changes the way that AI/ML accelerators are built

PUBLICLY REPORTED ADOPTIONS IN AI/ML



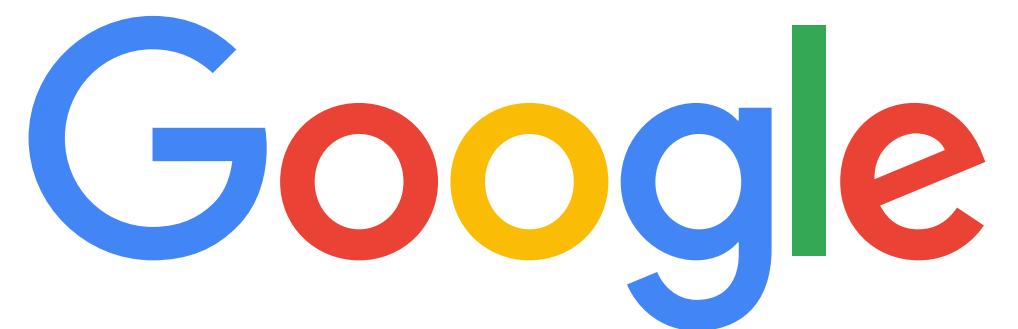
IN-HOUSE ACCELERATOR BASED ON RISC-V

ANNOUNCED ITS META TRAINING INFERENCE ACCELERATOR (MTIA) TO BE BUILT AROUND RISC-V CORES AND FABRICATED IN 7NM AT TSMC



MASS-PRODUCTION AI/ML ACCELERATORS

STREAM COMPUTING HAS BEEN SHIPPING ITS RISC-V BASED AI/ML ACCELERATORS RIVALLING THE NVIDIA A10 PRODUCT SINCE EARLY 2023

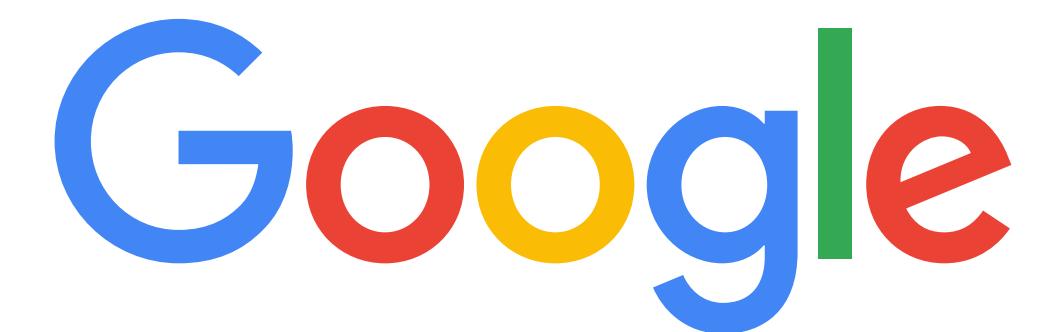


NEXT-GENERATION TPU BASED ON RISC-V

ANNOUNCED ITS NEXT-GENERATION TPU TO COMBINE A RISC-V CLUSTER FROM SIFIVE WITH GOOGLE'S OWN MATRIX UNIT

RISC-V changes the way that AI/ML accelerators are built

PUBLICLY REPORTED ADOPTIONS IN AI/ML



RISC-V changes the way that AI/ML accelerators are built

UPENDING THE ECONOMICS OF AI/ML ACCELERATORS

Sharing of a common
Base instruction set



REDUCING DUPLICATION

COLLABORATING ON THE COMMON PARTS OF
EVERY IMPLEMENTATION, RISC-V REDUCES
DUPLICATION AND FACILITATES REUSE

Custom instructions
allow differentiation and acceleration



ENABLING COMPETITION AND INNOVATION

TAILOR AND OPTIMISE PERFORMANCE WITH
DOMAIN-SPECIFIC ACCELERATION AND
CUSTOM INSTRUCTIONS FOR A COMPETITIVE EDGE

Shared ecosystem
supporting rapid innovation



LOWERING THE BARRIER OF ENTRY

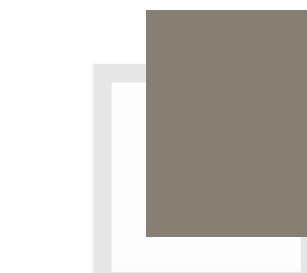
A UNIFIED FOUNDATION FOR DEVELOPERS
TO BUILD ON A CONSISTENT, OPEN PLATFORM
ACCELERATING INNOVATION AND COMPATIBILITY

AI/ML IS SOFTWARE DEFINED



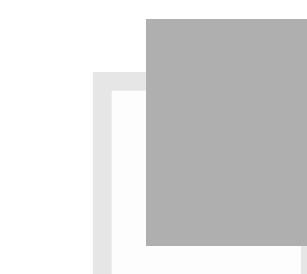
BUILT ON AI/ML FRAMEWORKS

AI/ML is built on frameworks (e.g., TensorFlow, ONNX, OneAPI, IREE, OpenXLA) that distribute “intermediate code” and bind to hardware late



RAPIDLY EVOLVING ALGORITHMS

Transformers have first been introduced in a research paper in 2017. While we can't predict the next algorithms, but we can prepare for innovation.



DIFFERENTIATION OPPORTUNITY

AI/ML is not held back by a focus on binary compatibility, but rather driven by cost and performance metrics.

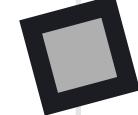
Building RISC-V into a “common language” for AI/ML, HPC and IoT

A TALE OF TWO STANDARDS

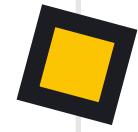
RISC-V Integrated Matrix Extension

Scalable matrix extension built on the Vector Extension
intended to reuse its micro-architectural resources

Designed to provide similar levels of geometry agnosticism and
integrate with the memory model of the RISC-V Vector Extension



RISC-V Attached Matrix Extension



Matrix extension in a self-contained, orthogonal execution unit
that scales down to IoT and up to HPC.
Designed for maximum freedom for integrators and accepting
higher software complexity in return.

Intended to provide a path to encompass tensor operations in
the future.

Building RISC-V into a “common language” for AI/ML, HPC and IoT ADDRESSING ALL MARKET SEGMENTS

RVV+IME vs. AME
provide choices for implementors



SOLUTIONS FOR ALL MARKET SEGMENTS
DEPENDING ON THE MARKET SEGMENTS
(SUCH AS HPC VS. AI/ML), RISC-V WILL
OFFER DISTINCT SOLUTIONS

AME provides a coherent strategy for
Matrix and Tensors



AME WILL OFFER A LONG-TERM EVOLUTION
IME BUILDS ON THE 1D RVV IMPLEMENTATION,
WHILE AME OFFERS A CLEAN APPROACH TO
1D (VECTOR), 2D (MATRIX) AND HIGH-D (TENSOR)

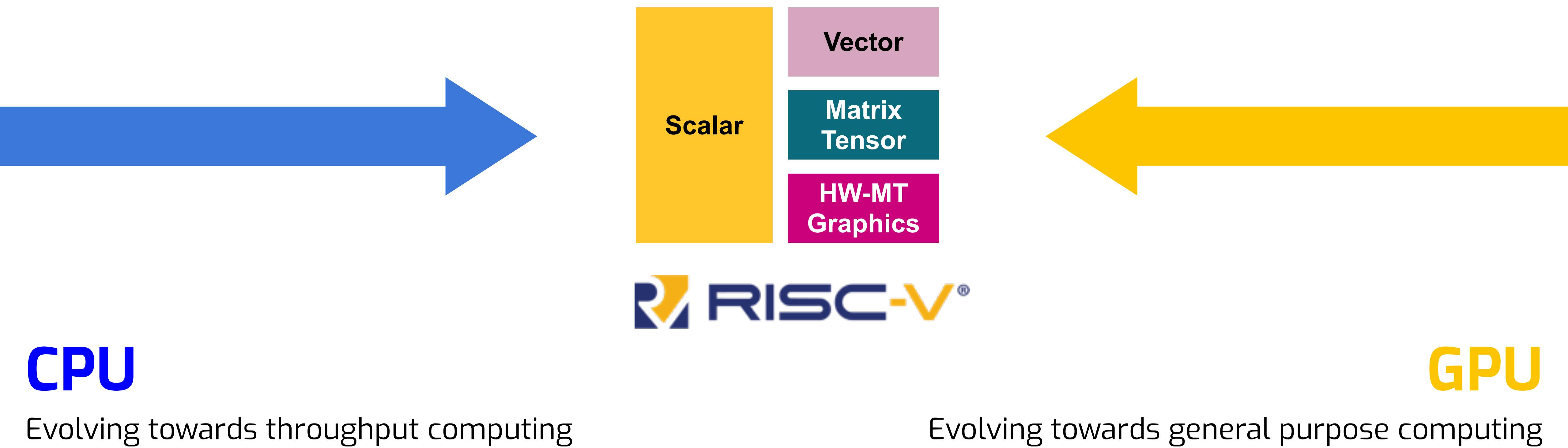
Unified enablement
through software abstractions



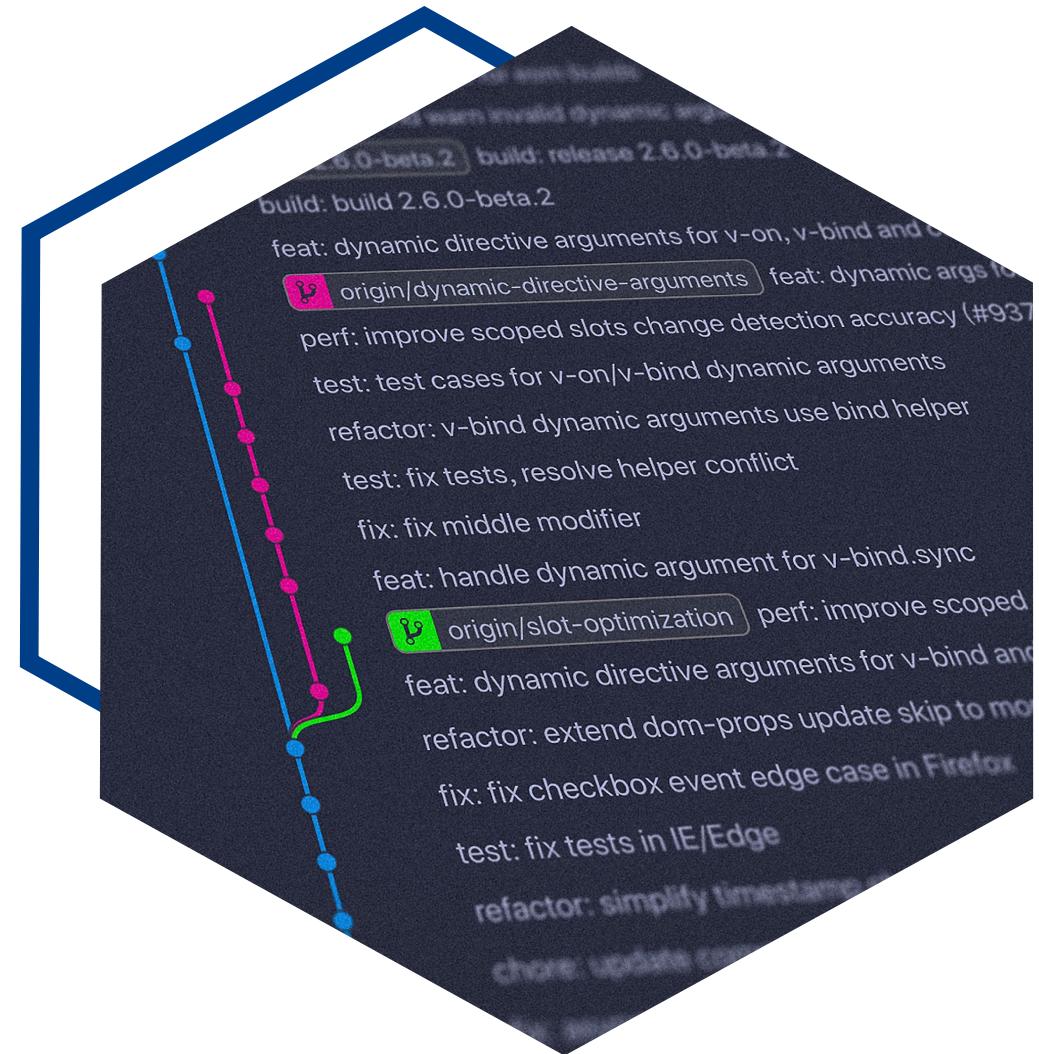
RAISING THE ABSTRACTION LEVELS
IME AND AME ARE AIMING TO PROVIDE
HOLISTIC APPLICATION COMPATIBILITY
RATHER THAN BINARY COMPATIBILITY

Towards the convergence of CPU, and (GP)GPU for improved programmer productivity

UNIFIED COMPUTE FOR RISC-V



VRULL is leading the charge on programmer productivity with Unified Compute DRIVING INNOVATION FOR AI/ML ON RISC-V



**SOFTWARE AND TOOLCHAINS
FOR INDUSTRY-STANDARD
COMPUTE & ACCELERATOR APIs**



**ALGORITHMS & WORKLOAD
INTELLIGENCE FOR OPTIMISED
MICRO-ARCHITECTURES**



**STANDARDS-DEVELOPMENT
FOR INTEROPERABLE
ECOSYSTEMS**



DEVELOPING STANDARDS CREATING MOMENTUM FOR AN INTELLIGENT TOMORROW

VRULL GmbH
Beatrixgasse 32, 1030
Wien/Vienna, Austria
contact@vrull.eu