A Big "Hello!"

LatentAI

About LatentAI

- We develop tools to enable efficient, adaptive AI that is optimized for compute, memory, and power consumption that integrates with existing AI/ML infrastructure and frameworks to overcome the challenges of AI for the edge.

- Vitals:
  - SRI startup spin-off backed by DARPA technologies.
  - Founded in Dec 2018.
  - Seed round led by Future Ventures.
  - Offices in Menlo Park, CA, and Princeton, NJ.

LatentAI

Our Goals with TVM Project

- We are relatively new to the TVM project and community so we are in full learning mode.

- Integrate TVM stack services into our developer tools to explore new edge compute use cases.

- Share our learnings and continue to engage with the community for future collaborations.

LatentAI

Tool-flow integration

- We have learned the following from exploring ways to integrate our tools framework with TVM:

    - Standardizing performance sweeps through scripts including *xgb* and *random* (20, 100 and 500 iterations) for x86 (Skylake) and Arm (A52/A72).

    - A CLI that enables scripting is more appropriate for running performance and inference accuracy sweeps which can discuss through an RFC.

    - Developed working demos to showcase performance including SSD_MobileNet_v1 running on a Raspberry Pi and Tiny YOLO on iOS.

LatentAI

Future collaboration

- Continue learning the TVM code base and getting acclimated to the community.

- Explore ways to best contribute back to the project and community.

- Continue to meet with the community in person at events to share experience and findings.

LatentAI

Thank you!

Abelardo Lopez-Lagunas
abelardo@latentai.com

Mark Griffin
mark@latentai.com

LatentAI