Origami

Bringing Machine Learning to the very end of IoT Edge

09/01/2023

PROBLEM: No Cloud, No AI

How AI works now:

1. IoT device sends data to Cloud DC
2. Data is processed by Cloud AI
3. Prediction is sent back to IoT device
4. IoT device actuates control

Cloud AI is

dependent on network connection, too **far** away to respond quickly, too **big** to run locally, and too **expensive**!

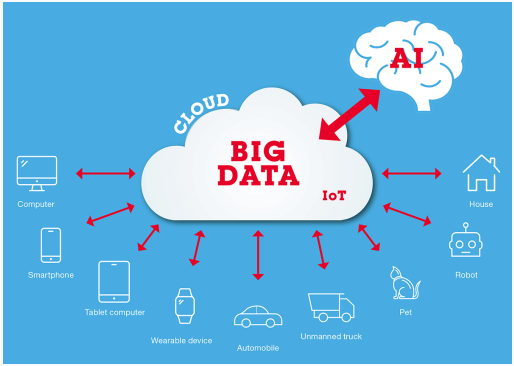


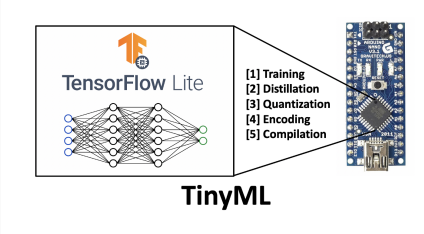
Figure 1: Big AI on Cloud

SOLUTION: Cloudless AI with tiny Machine Learning

We squash ML smaller, run on microcontroller in IoT device, process IoT sensor data locally. This is called,

Tiny Machine Learning

1. Cloudless & Networkless
2. small, fast & energy efficient
3. cost effective



Market opportunity: *Now AI Is Eating IoT*

- 2011, Why Software Is Eating the World, a16z
- 2017, The End of Cloud Computing, a16z
- 2019, Software Ate The World, Now AI Is Eating Software, Forbes
- 2022, AI Accelerators Enter IoT SoCs, EETIMES
- 2023, AI Ate Software, Now AI Is Eating IoT, Origami

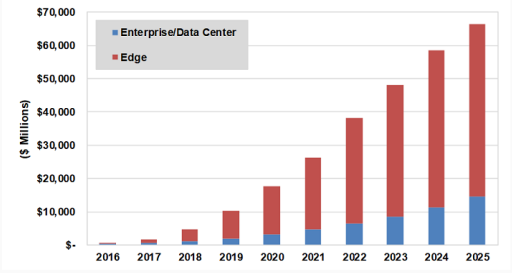


Figure 2: Deep learning chipset revenue

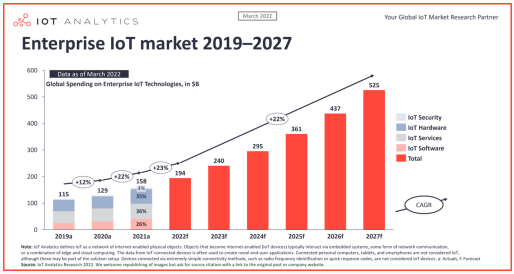


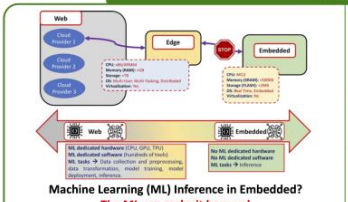
Figure 3: Enterprise IoT market

Product: TinyML as-a-Service (TinyMLaaS)

Why join the Cloud ML navy when you can be a TinyML pirate?

Hiroshi Doyu (Origami) and Roberto Morabito (University of Helsinki, Origami)

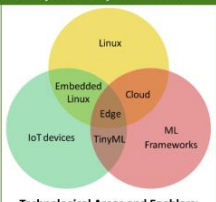
What is TinyML? Why we need it?



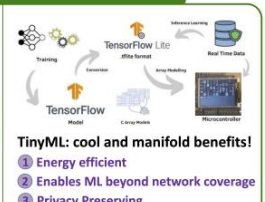
Machine Learning (ML) Inference in Embedded?

TinyML can make it happen!

And goodbye to slow and expensive cloud-based ML inference!



Technological Areas and Enablers: Where does TinyML stand?

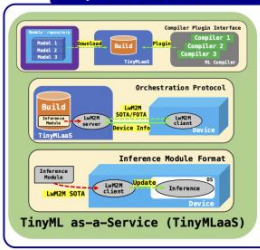


TinyML: cool and manifold benefits!

1. Energy efficient
2. Enables ML beyond network coverage
3. Privacy Preserving

But...is it enough?

TinyML as-a-Service: three "interfaces" for *democratizing* TinyML



Compiler Plugin Interface
Middleware abstraction bridging the requests submitted to the TinyMLaaS front-end with the compiling parameters requested by the back-end

Orchestration Protocol
Through OMA Lightweight M2M (LwM2M) and standardized data representation models, it manages the interactions between devices and the TinyMLaaS infrastructure (ex. Software/Firmware Over-The-Air – SOTA/FOTA)

Inference Module Format
Provides the output processed by ML compilers via LwM2M SOTA to devices

	Device and ML service management capabilities	Orchestration and ML lifecycle capabilities	Connectivity management Capabilities	Network latency dependency
Cloud ML	Yes	Yes	No	High
TinyML	No	No	No	Low
TinyMLaaS	Yes	Yes	Yes	Low

TinyMLaaS in Action

Business model

Table 1: Price list

Type	Description	Price (EUR)
License	Per TinyML model annually	30K/model
Subscription	Per TinyMLaaS traffic	0.1/Request
Professional Service	Time and Materials	150/H

Marketing and sales strategy

Basically TinyML could be applied to any industries. **Now We focus on:**

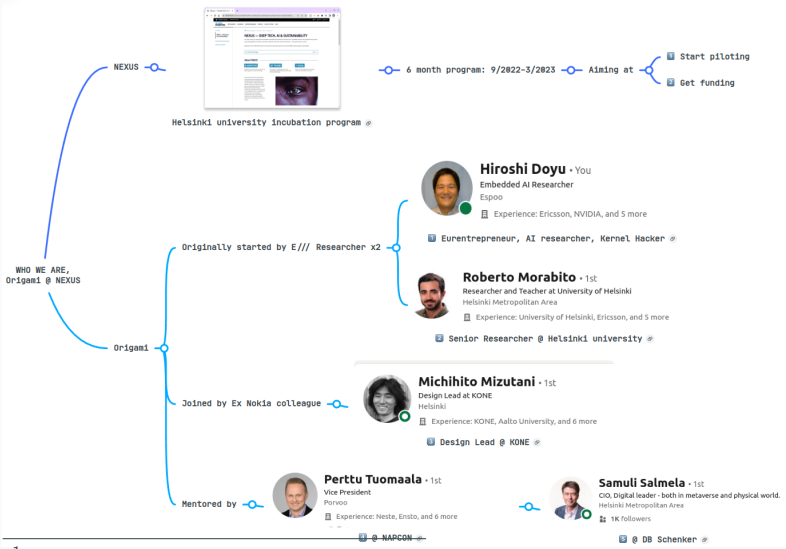
Factory Automation

- Networkless AI Automation
- Cloudless Predictive maintenance
  - without big data storage
- Autonomous Emergency shutdown

HealthCare

- Cloudless Medical Gadget
  - with longer battery life
  - with privacy preservation
  - with small ML models

Team[1]



[1] https://www.mindmeister.com/map/2547854225

History & Progress

- 09/2022 Origami started under NEXUS incubation program[2]
- 12/2022 On-site workshop held on customer premises
- 01/2023 A workshop report & piloting proposals sent
- 02/2023 Agree on piloting project with *Letter of Intent* (LoI)
- 04/2023 Acquire funding with LoI
- 05/2023 Close a piloting contract
- 06/2023 Hire more developers
- 08/2023 Start development
- 10/2023 End of development
- Q1/2024 Start TinyMLaaS subscription service

[2] 20 promising teams selected for mentor-driven NEXUS incubator programme

Ask

Looking for piloting projects with TinyMLaaS!

We could offer to:

1. hold an on-site workshop to consult ML challenges
2. propose multiple (Tiny)ML solutions
3. run & verify TinyML project together

Contact information

Origami

https://Origami-TinyML.github.io/blog/about.html