



Less Power, Less Cost, More Impact – tinyML[®] and Edge AI To The Rescue

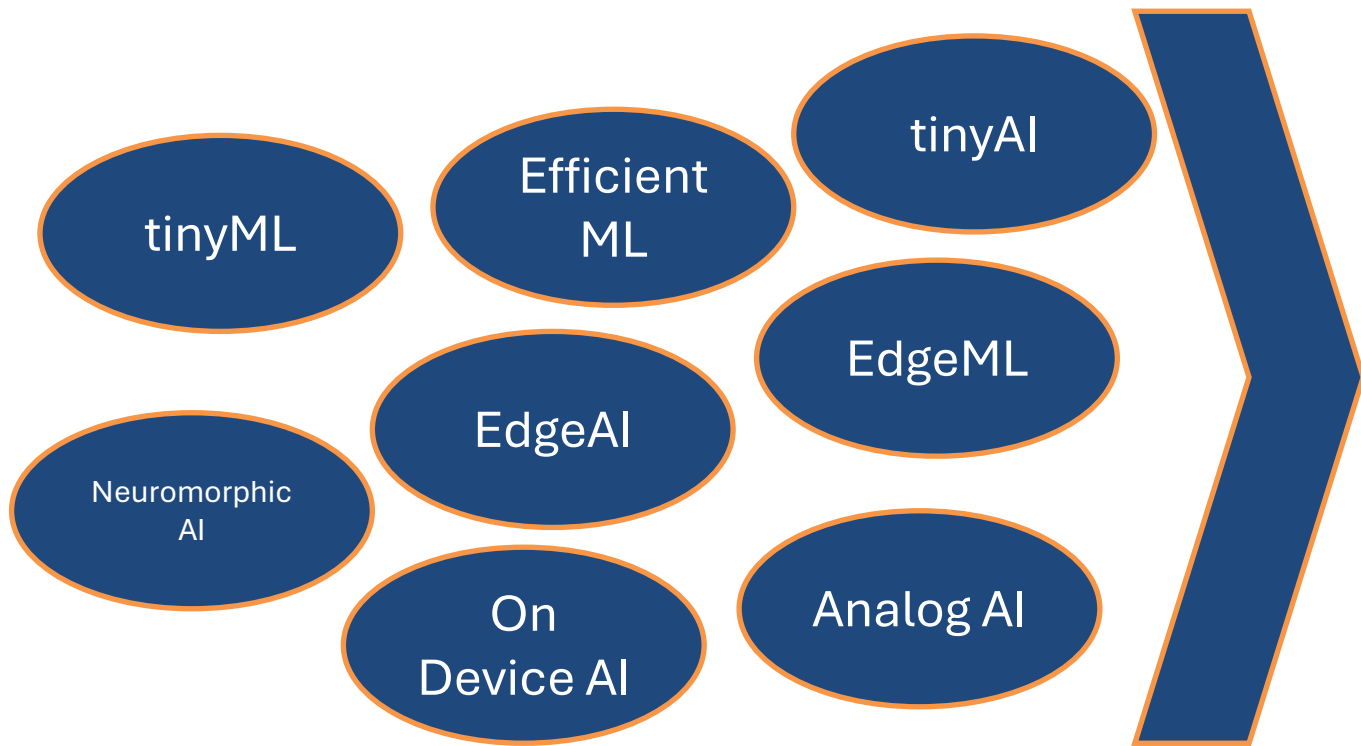
Pete Bernard, Executive Director, tinyML Foundation

Pete Bernard, Executive Director



- EDGECELSIOR - Founder & Principal
- Microsoft - 18 years
 - Senior Director - Azure Edge HW/5G
 - Principal Group PM – Windows & Devices Products
- Silicon Valley - 14 years
 - Insignia Solutions - Chief Product Officer
 - E Color - VP Products
 - Phoenix Technologies – Director of Technology
- Seven patents in security, imaging, smart wearables, and music streaming

Defining Terms

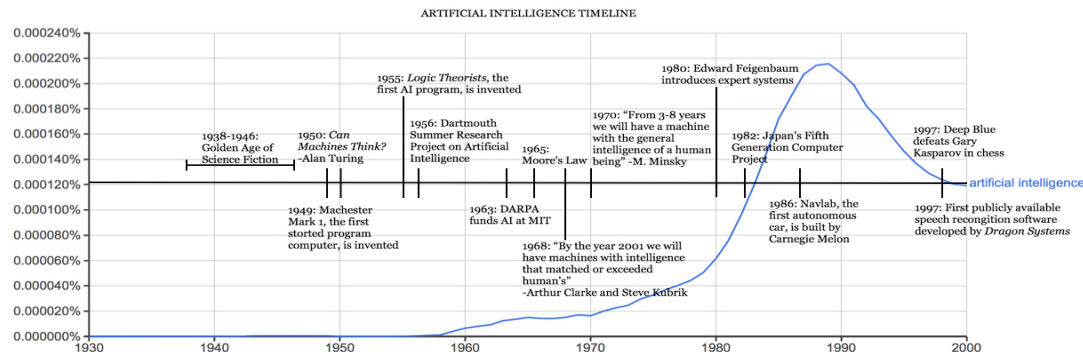


Where AI
Meets the
Real World.

How Did AI Get So Weird?



“AI” has been around for almost 70 years . . .



In 1970 Marvin Minsky told Life Magazine, “from three to eight years we will have a machine with the general intelligence of an average human being.”

AI Technology Reflects Society

- Loneliness poses risks as deadly as smoking: surgeon general (May 2023)
- Poll of 2,000 executives, ... showed that 41% of them expect to employ fewer people because of AI – April 2024
- Impose 'robot tax' for AI-induced job loss; India in July 2024

"To be perfectly blunt, AI girlfriends are not your friends. Although they are marketed as something that will enhance your mental health and well-being, they specialize in delivering dependency, loneliness, and toxicity, all while prying as much data as possible from you."

MISHA RYKOV, RESEARCHER @ *PRIVACY
NOT INCLUDED



AI “Democratization” Is Critical – but Complex

- the democratization of AI use
 - **reducing the costs of acquiring and running AI tools** and providing intuitive interfaces to facilitate human-AI interaction without extensive training or technical know-how.
- the democratization of AI development
 - **involving more people** (e.g., academics, individual developers, smaller labs) in AI development processes provides a critical external evaluation and auditing mechanism.
- the democratization of AI profits
 - **avoid widening a socioeconomic divide between AI-leading and -lagging nations**, to ease the financial burden of job loss to automation, to smooth economic transition in case of the rapid growth of the AI industry, and to provide mechanisms for labs to demonstrate their commitment to pursuing advanced AI for the common good.
- the democratization of AI governance
 - ensure that decisions around questions such as AI usage, development, and profits **reflect the interests and preferences of the people being impacted**.

For example, the circle of individuals who would greatly benefit from access to an AI drug discovery tool is relatively small (mainly pharmaceutical researchers), however, these tools can be repurposed to discover new toxins that might be used as chemical weapons.

Is AI Sustainable?

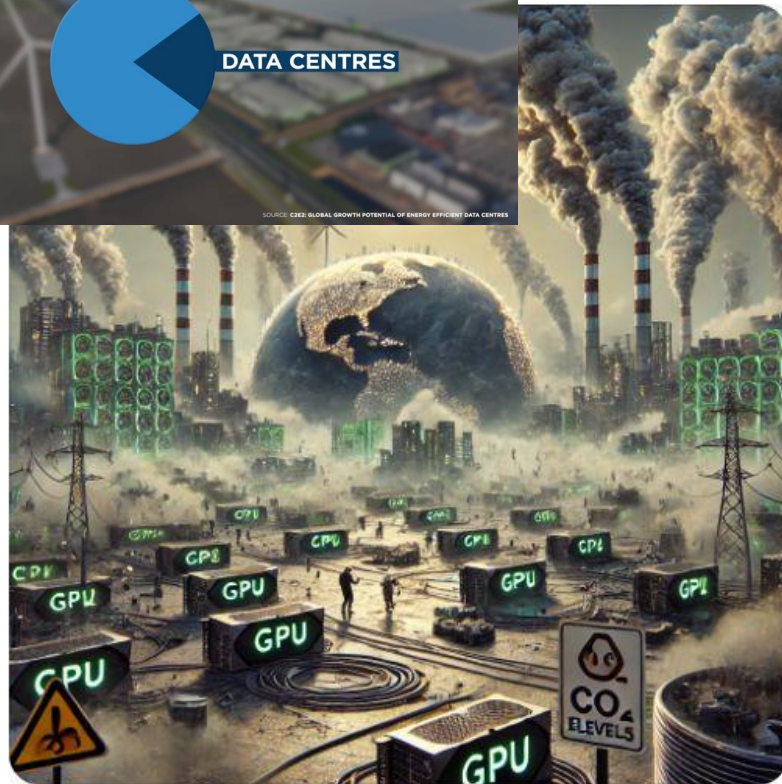
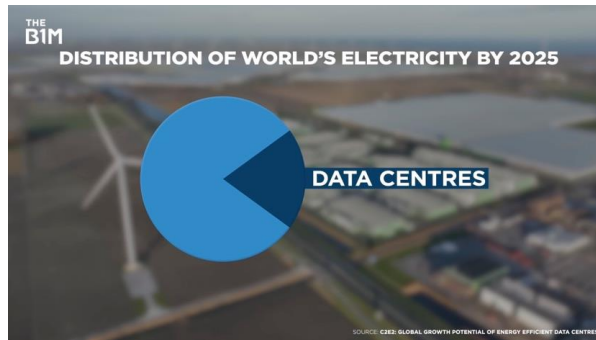
- A single NVIDIA H100 GPU results in the same emissions as a Toyota Corolla traveling 10,000 miles a year.
- NVIDIA's high-end GPUs will generate the same emissions as over TWO MILLION gas powered cars.

Common carbon footprint benchmarks

in lbs of CO2 equivalent

Roundtrip flight b/w NY and SF (1 passenger)	1,984
Human life (avg. 1 year)	11,023
American life (avg. 1 year)	36,156
US car including fuel (avg. 1 lifetime)	126,000
Transformer (213M parameters) w/ neural architecture search	626,155

Chart: MIT Technology Review • Source: Strubell et al. • Created with Datawrapper

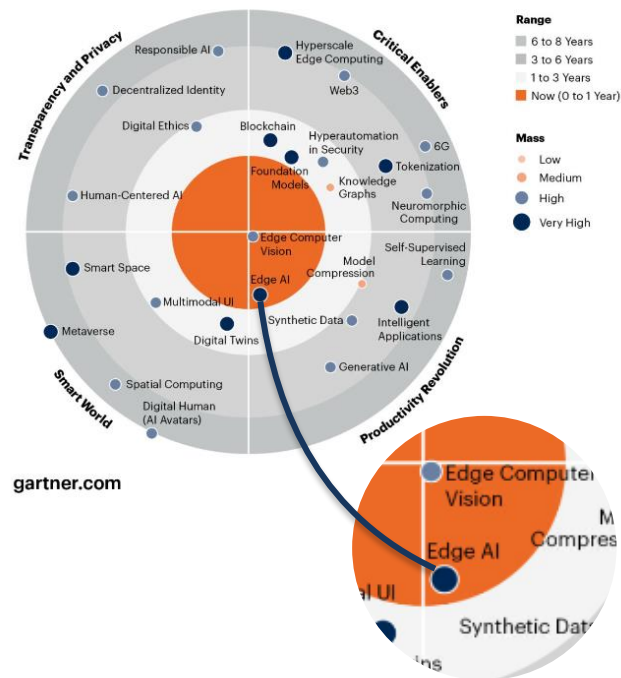


Edge AI To The Rescue!

Technology is Accelerating Digital Transformations

- Advances in semiconductors, connectivity and software have enabled a wave of AI on devices
- Digitally transforming healthcare, agriculture, manufacturing, retail and many more...
- From low power sensors, imaging, gateways and robotics platforms – running AI Models with cloud-to-edge management and security

2023 Gartner Emerging Technologies and Trends Impact Radar

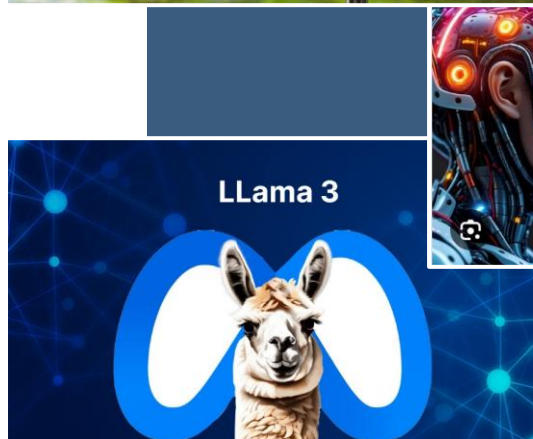


Edge AI Technology State-of-the Art is Moving Quickly

- AI models can run on solar-powered sensors
- LoRA and 5G Redcap are enabling far-flung deployments
- Gen AI models are coming to the edge – Meta's LLAMA model shrank 10x in 9 months



"up to 99% of use cases could be addressed using SLMs . . 2024 will be the year of the SLM."
– CEO of Hugging Face



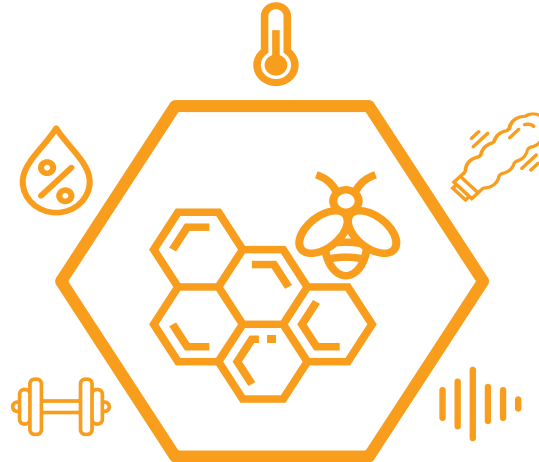
Example: Agriculture



Hive too **dumpy** may lead to pest & disease infestation, fungi & Mould growth

Measuring hive **weight** to keep track of honey production & bee colony population

Hive too **hot** that may reduce honey quality & also indication of brooding cycle



ibees

An AI powered beehive monitoring solution to help farmers monitor the health of their beehives in real time on their **phone**.

Hive **knocked over** or in the process of being stolen

- Queen not laying
- Colony about to swarm
- Bees disturbed

Example: Critical Infrastructure



Progress to Date



Sydney
WATER

WaterNSW

SA Water

yarra **trams**

**Keolis
Downer**

ARTC

AURIZON

Woolworths



coles



~ 100,000
Devices deployed

~ 2.8 billion
Measurements captured
annually

3 x
Devices awarded highest
engineering design award
in Australia

100+ Global Customers

Across water, rail and commercial industry

Example: Smart Cities

Use cases : Smart Cities

Traffic management



- Traffic counting
- Vehicle type
- Speed

Car parking management



- Outdoor / Indoor / Street parking
- Highway rest area

Dynamic lighting



- Lighting control based on presence of vehicle / pedestrian

Road surface management



National / State park



- Entrance queue
- Parking availability

Road safety



- Wrong way detection

TINY
ML BLUEPRINTS



- Alert for crowd gathering

HOW DOES EDGE AI KEEP PEDESTRIANS SAFE?

with

SONY

LEONARDO
IMAGING

SAN JOSE
SCHOOL OF ENGINEERING



AITR OS

tinyML[®] Foundation is “the place” for the Edge AI Technology Community



100+ partners across Technology and Fortune 500



500K+ YouTube Views; **13k+** YouTube subscribers;
21k+ followers on LinkedIn



Deep academic partnerships - **~100k** students taking tinyML classes worldwide



In-person symposiums / summits in the **US, EMEA, ASIA**



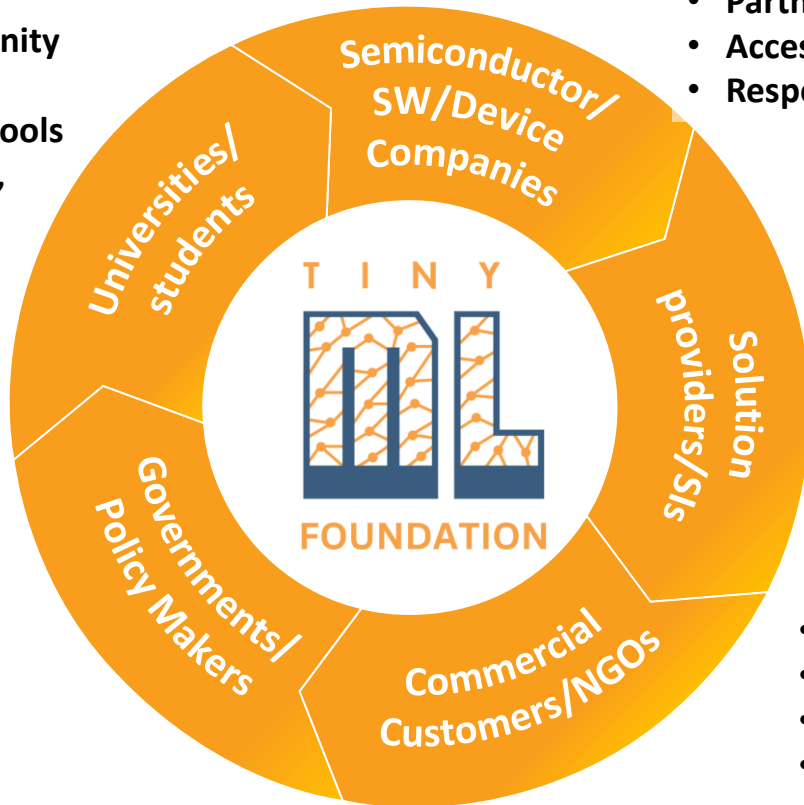
50 meetup groups in **41** countries

A 501c3 non-profit/charitable organization

A Broad Community of Expertise

- World-wide academic community
- Coursework/training
- Access to tech for platforms, tools
- Opportunities for mentorship, internship, career paths

- Education on Edge AI
- Access to leading edge tech
- Feedback and input on policy



- Knowledge of state-of-the-art
- Partnership and Business Dev
- Access to talent
- Responsible AI opps

- Access to leading edge tech
- Exposure to tech suppliers
- Commercial customers
- Access to talent

- Access to leading edge tech
- Responsible AI solutions
- Find solutions with partners
- Access to talent

Working Groups

Working Groups are populated by Partners and Academia to drive cross industry initiatives and best practices in a neutral and collaborative environments

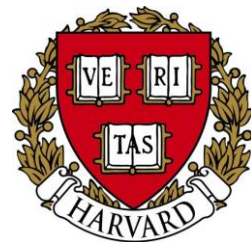
- **Datasets and Benchmarks**; how do we develop and contribute high quality datasets and models to the community and establish best practices on comparing “performance”?
- **Blueprints**; what are solutions deployed in key industry verticals that we can turn into blueprints including guidance documents code snippets, datasets, models and reference hardware to accelerate deployments?
- **Sustainability**; what are the best way to focus the community on the United Nations Sustainability Development Goals by working with UNICC and ITU? Input and guidance into the World Economic Forum AI Governance efforts
- **Marketing**; effective storytelling to highlight the impact of “AI in the real world” across our partner organizations

Academia + Industry Partnership

- tinyML EDU efforts have educated over 100k students worldwide
- tinyML Foundation Symposiums highlight academic research alongside industry events
- We support scholarships for next gen AI leadership
- Providing opportunities for mentorships, internships and talent pipelines



The Abdus Salam
International Centre
for Theoretical Physics



MAYO
CLINIC



JOHNS HOPKINS
UNIVERSITY

Navigating Responsible AI and AI for Good Engagements for our Community

- Member of the ITU AI for Good effort
 - tinyML Workshop at the 2024 Summit in Geneva
- World Economic Forum member
 - Developing best practices across industries, educating policy makers
 - Member, AI Governance Council
- Sponsoring Hackathons tackling societal challenges
 - Pedestrian Safety Hackathon with Infineon, Sony and Brainchip



tinyML Foundation Community Members

– a mix of startup, enterprise and Fortune 100



Strengthening Community with In-Person Events



"It is my first time attending the tinyML Summit, and the experience has been fantastic. I had the opportunity to network with top edge ML researchers from both industry and academia, discussing the current and future challenges of implementing ML on low-power devices..."

– Wilfredo Lugo-Beauchamp,
Assistant Professor, University of Puerto Rico



tinyML Foundation Milan – June 2024



*"The forum was not just about sharing knowledge but also about absorbing the creative energy from many friends and colleagues I met for the first time. The future of innovation is bright, and it's built on the **dense layers of human connections** we make..."*

– Nabarun Dasgupta, ST Microelectronics



Less Power, Less Cost, More Impact – tinyML and Edge AI To The Rescue ...

Accelerate Business

- Regional Community Events
- Working groups
- Partnerships
- Hackathons for Industry Verticals
- Knowledge Certification programs

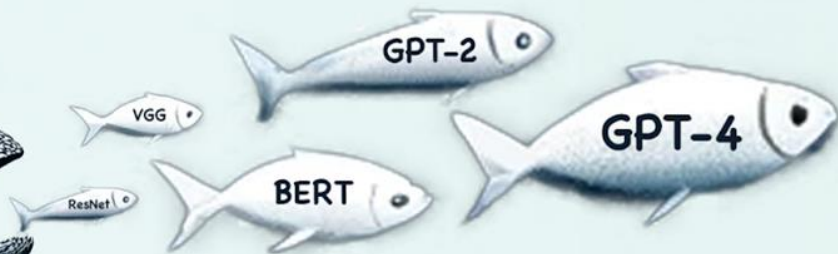
Accelerate Academia

- Symposia to highlight cutting edge research
- Scholarships and funding
- Talent pipelines, and mentorship/internship programs

Accelerate Society

- AI for Good via United Nations and World Economic Forum ++
- Responsible AI thru policymaker education
- Security/transparency working groups
- Hackathons for societal solutions

Make Edge AI a Digital AND a Societal Transformation



THANK YOU.

tinymml.org

Image courtesy: Prof. Vijay Reddi and his Harvard team

