2nd April 2025

Edge AI 2.0

for Future Computing

Aaron Ding

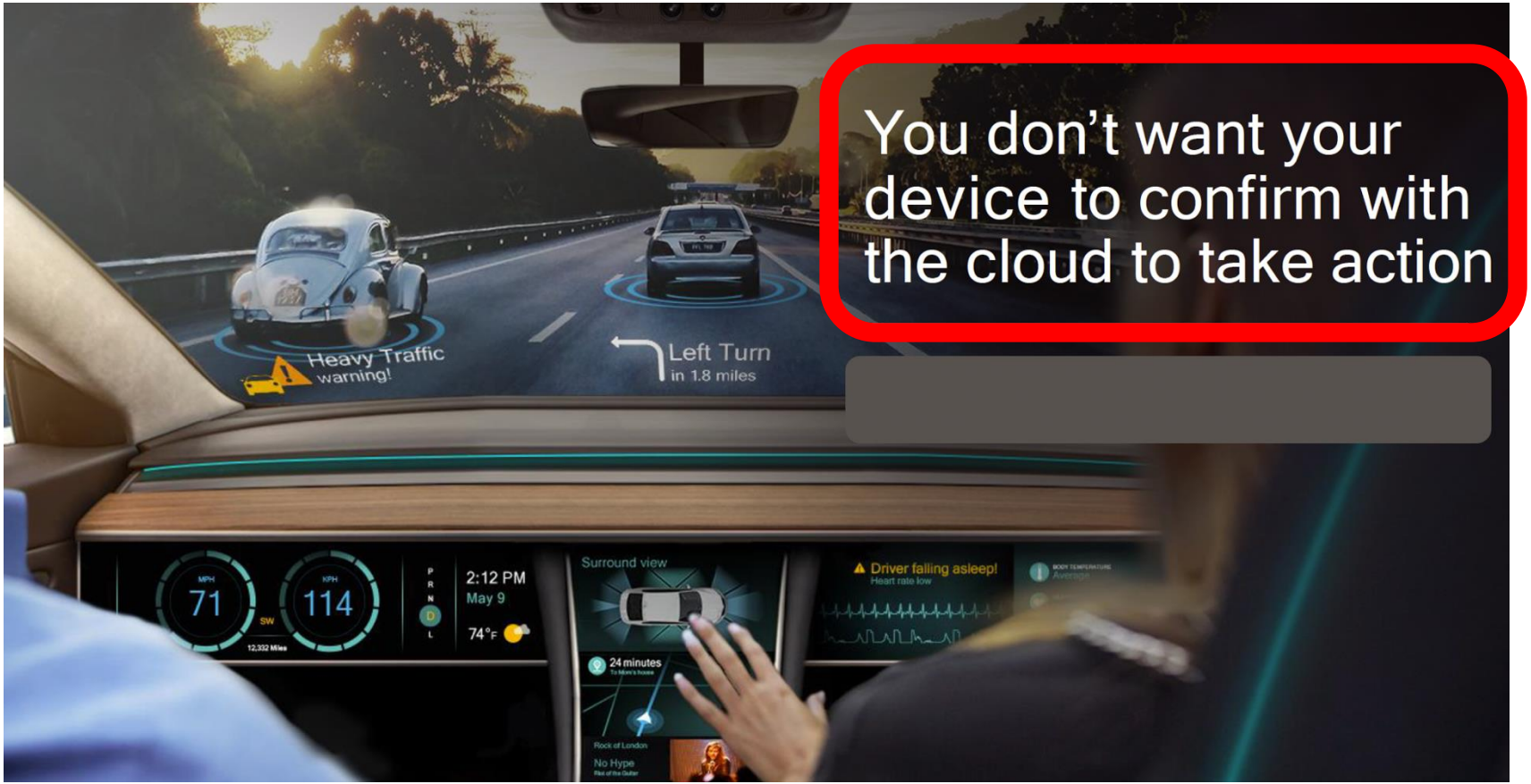
Director of CPI, Senior Associate Professor (UHD1 Ius Promovendi)

Why + Lessons + Directions

Why + *Lessons* + *Directions*

Challenges

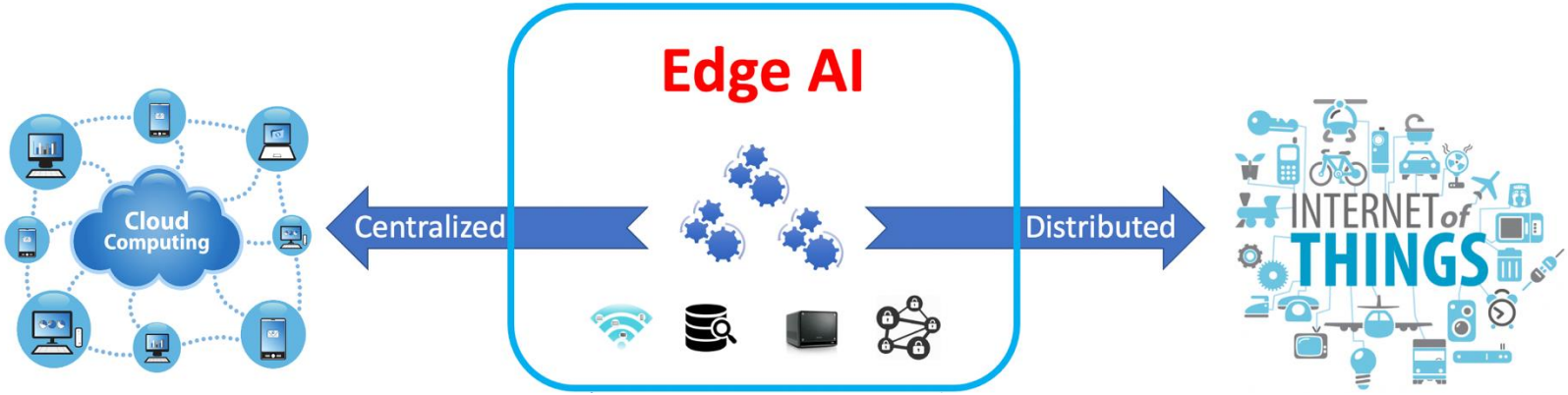
latency, reliability, security, data privacy and trust …



TU Delft

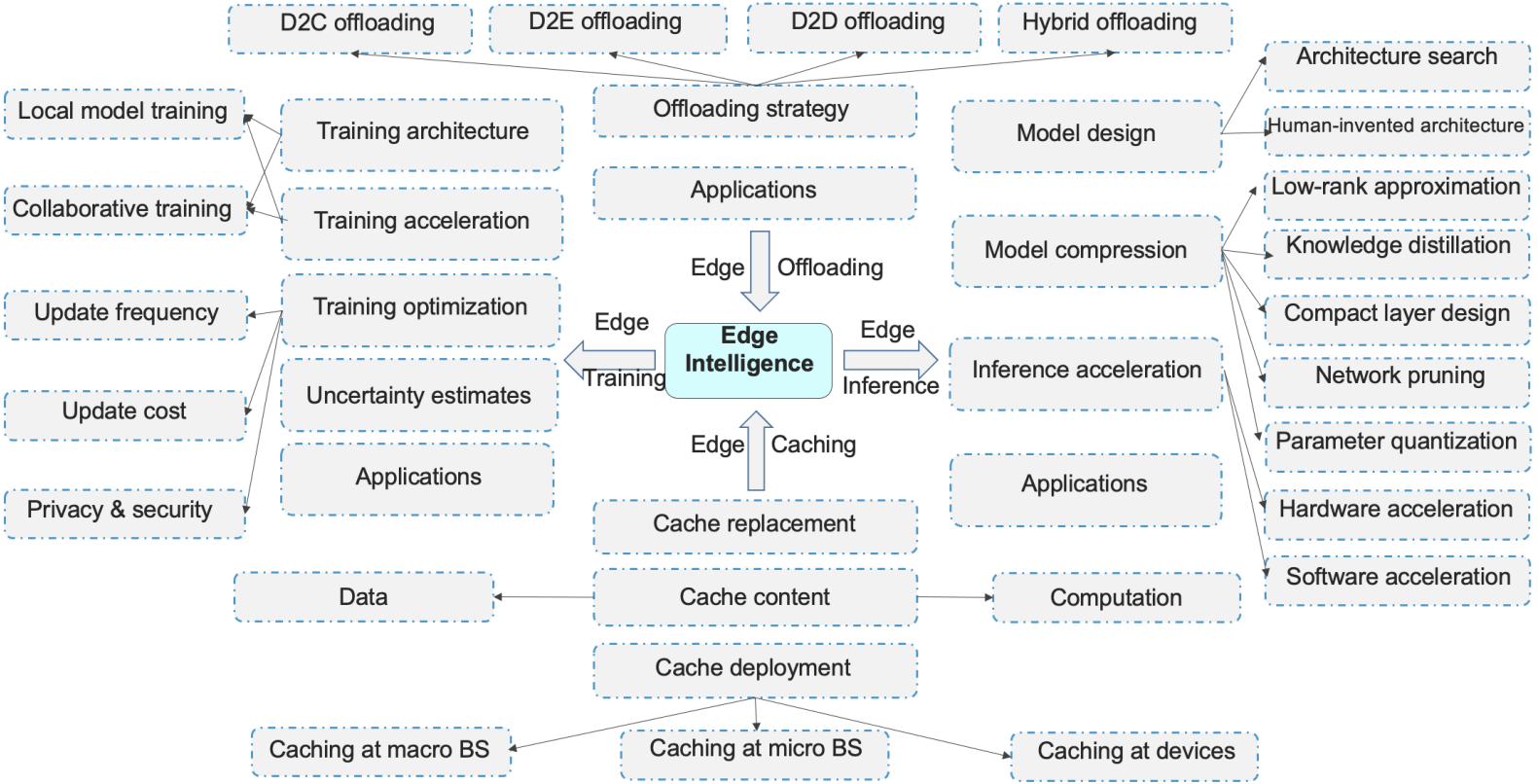
Motivations

Bridge the Gap



Consolidate Cloud & IoT

Enabling Techniques



"Edge Intelligence: Empowering Intelligence to the Edge of Network", Proc of IEEE, Vol 109, No 11, 2021

But Where Are They ?

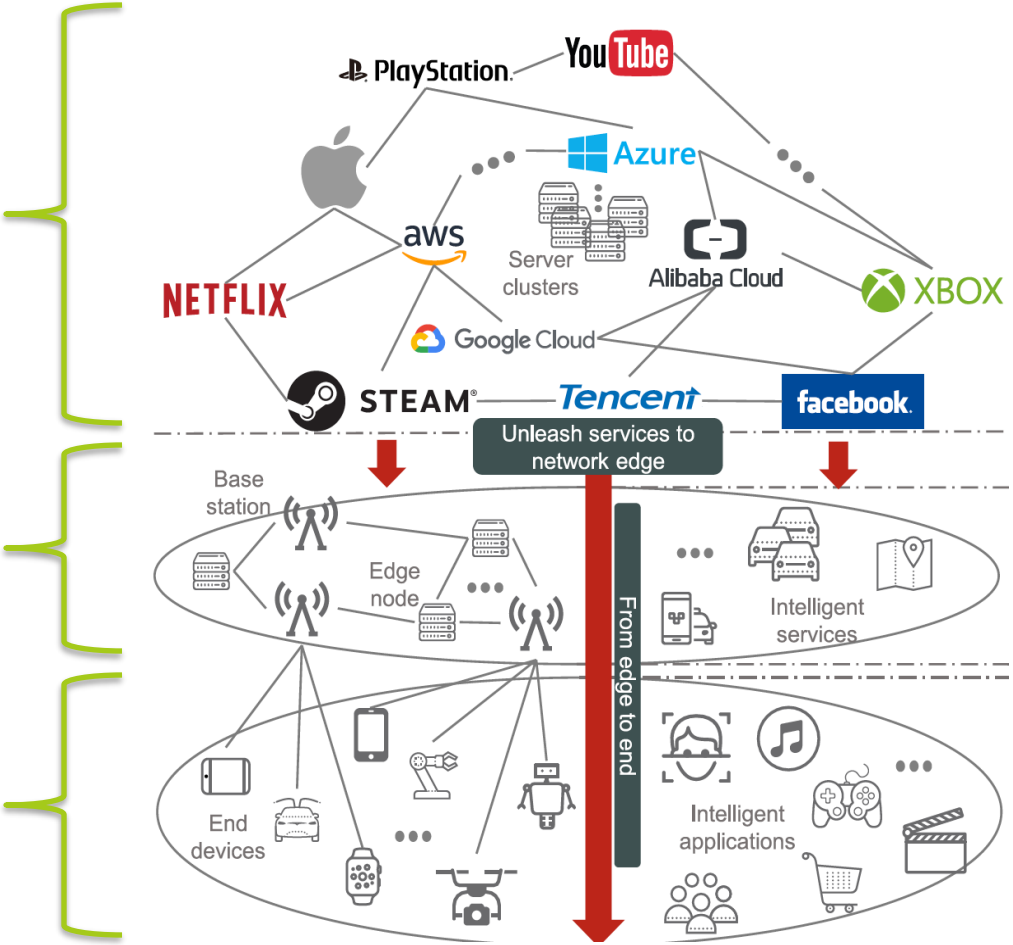
Holistic View

Cloud View

Cloud Layer

Edge Layer

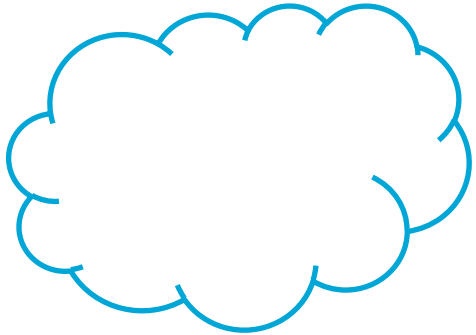
IoT & End Devices



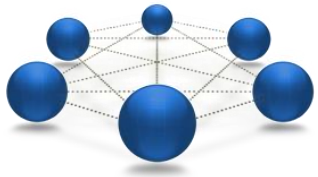
Convergence of Edge Computing and Deep Learning: A Comprehensive Survey, IEEE COMST 2020

What exactly is it ?

Edge Paradigm

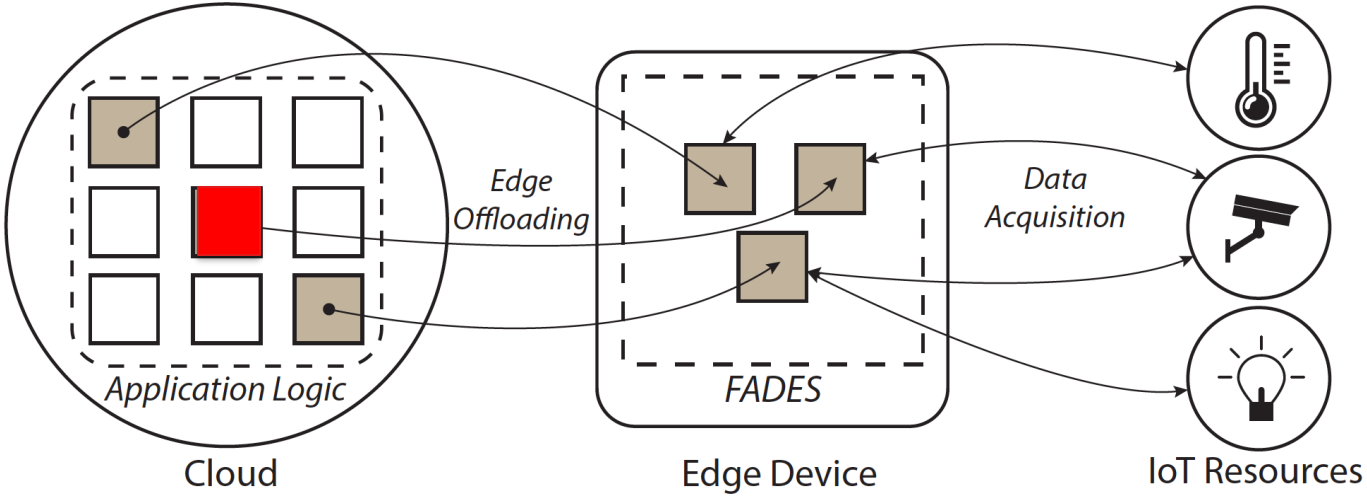


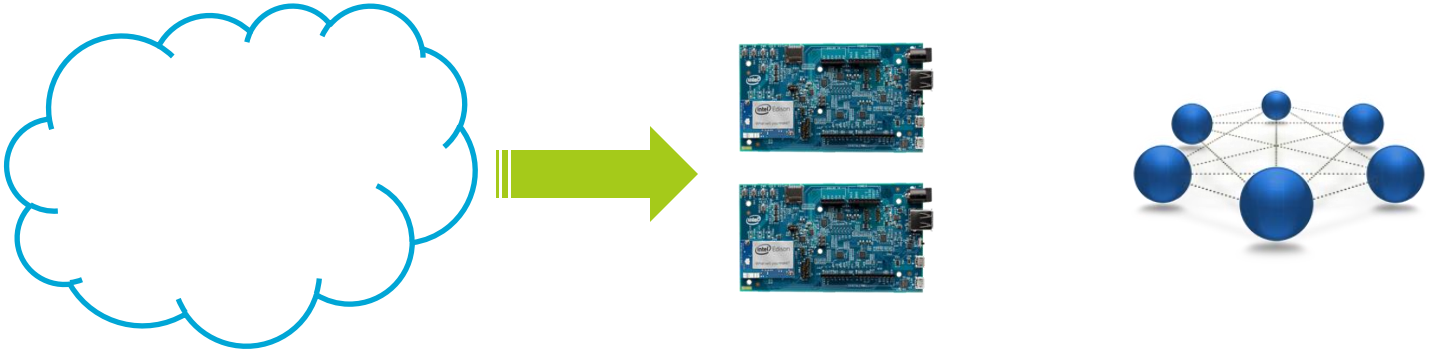
Edge



IoT

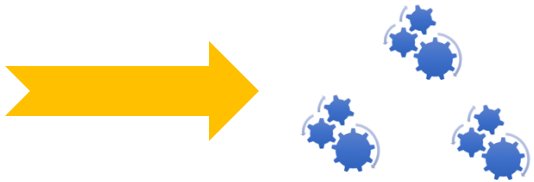
Example

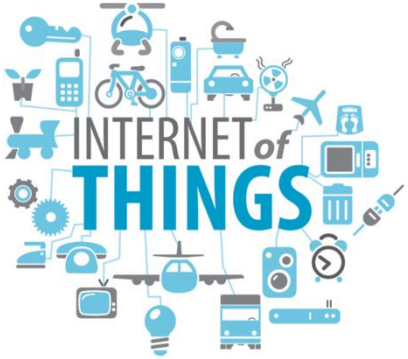

Edge
Offloading
Data
Acquisition
Application Logic
FADES
Cloud
Edge Device
IoT Resources



Why Bringing Intelligence to Edge

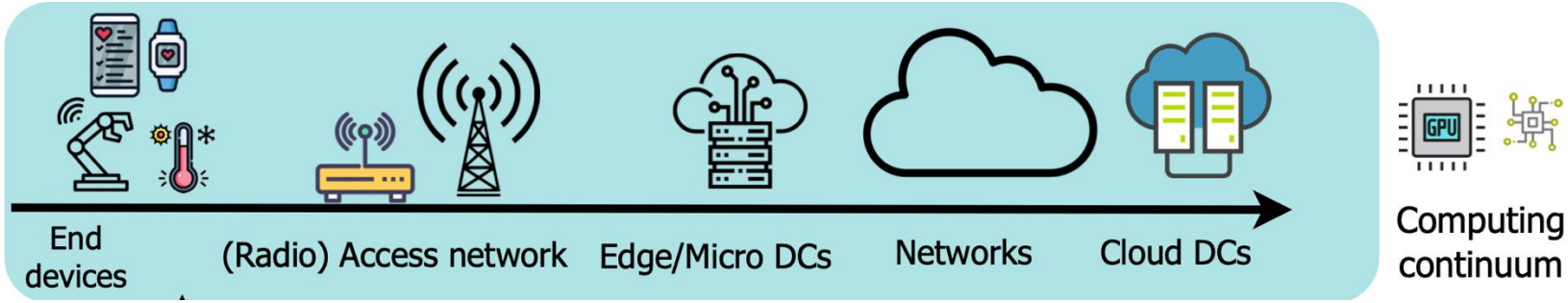
Problem Solved !?

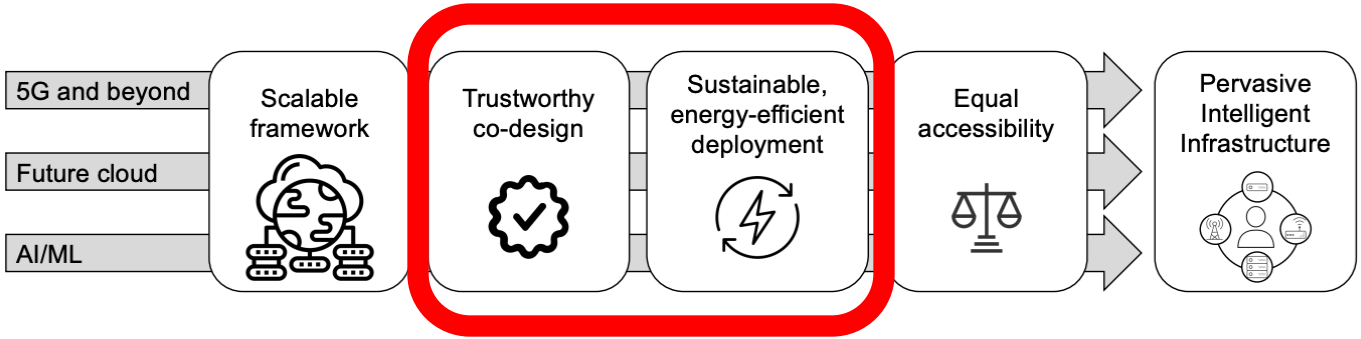




Are You Sure …?

Future Computing – Edge AI 2.0

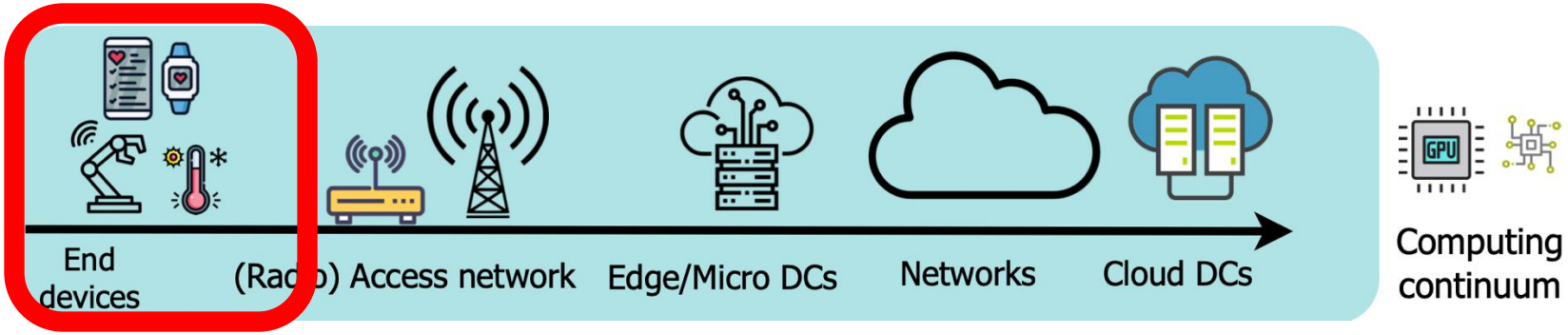


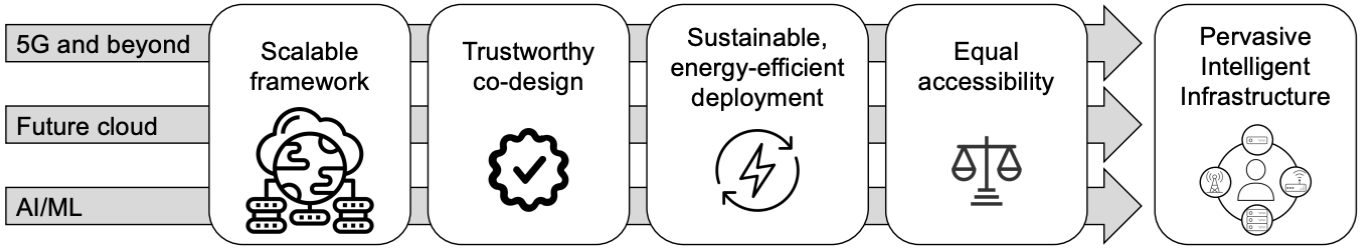


"Roadmap of Edge AI: A Dagstuhl Perspective", ACM SIGCOMM CCR 2022

*Why + **Lessons** + Directions*

Tiny but not Wasted: Circular Economy?





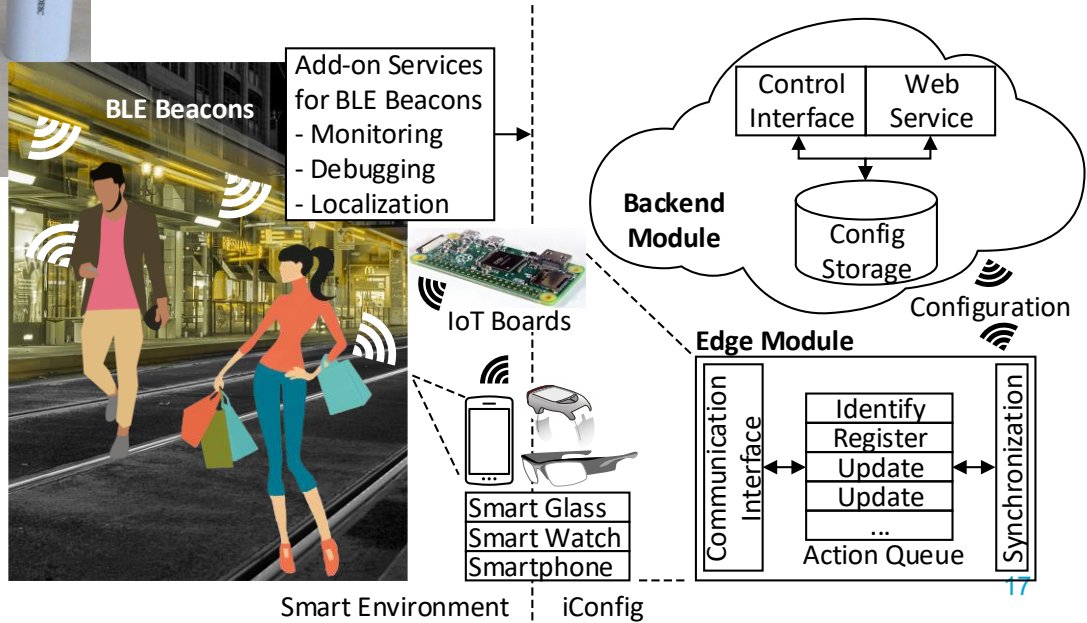
"Roadmap of Edge AI: A Dagstuhl Perspective", ACM SIGCOMM CCR 2022

Case: Tiny IoT

what I see is what i Configure

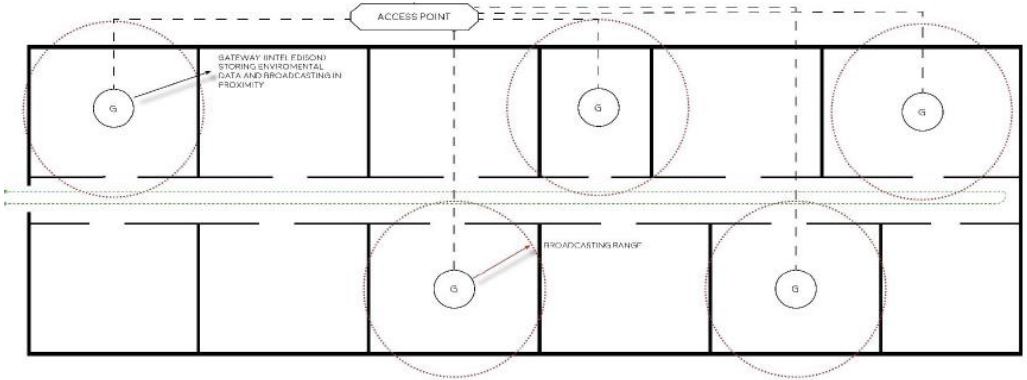
“ I don’t get what’s this really useful for …”

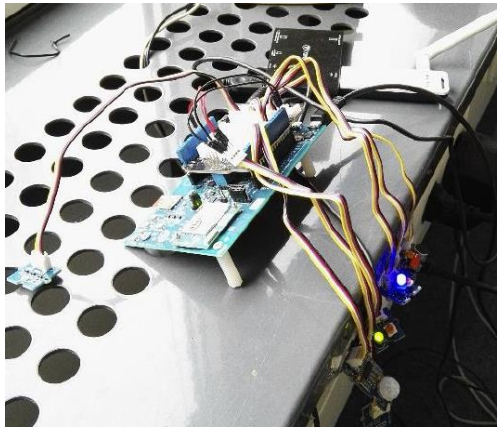
“ Why those BLEs are ever needed ? Why configure them ? ”



Starting on Campus

- Live testbed on campus
  - Office room context monitoring (IoT boards)
  - Proximity reasoning (BLE beacons)

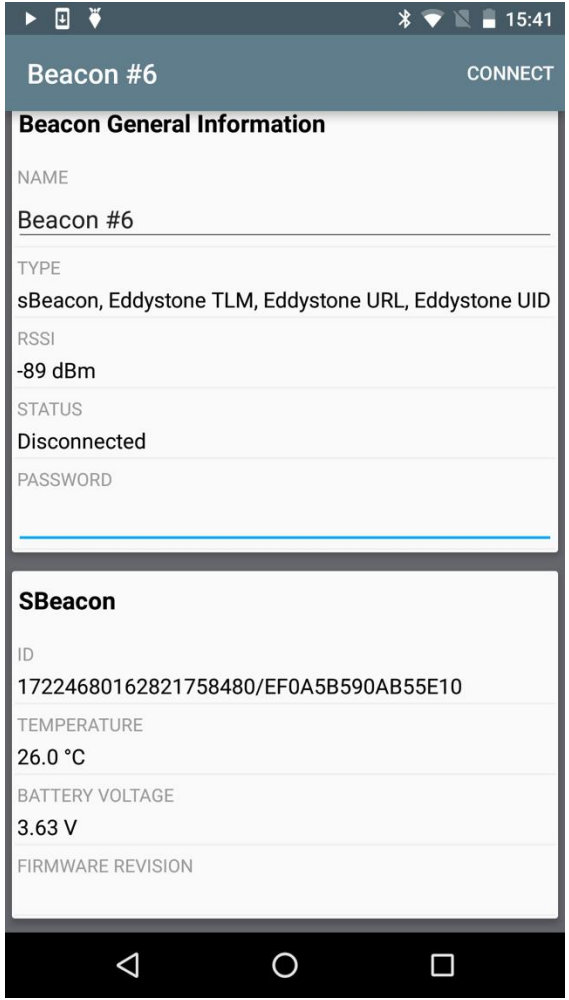




How to Configure ?

- Complications started
  - No backend, no fixed connectivity
  - Poor UI
  - Purely manual
  - Error prone
  - Time consuming

- Scalable ?
  - 10-20, perhaps doable
  - 100+, will drive people mad



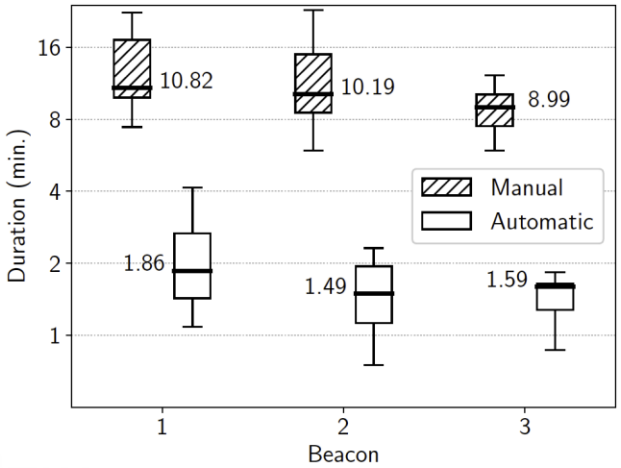
Where to track ?

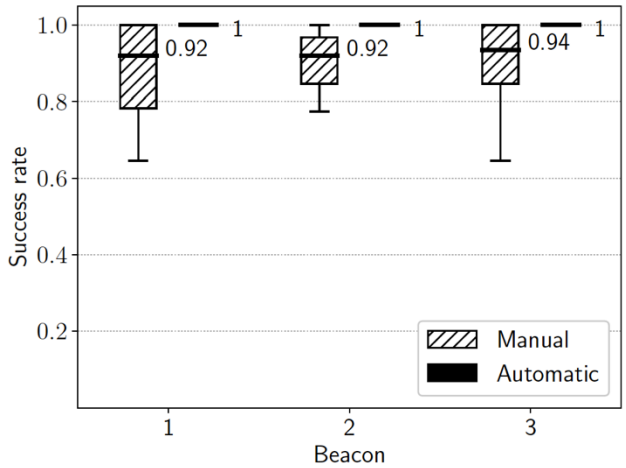
- Distributed across buildings
  - Paper, manual logs got lost…
  - Lack of tools, process
- Management nightmare

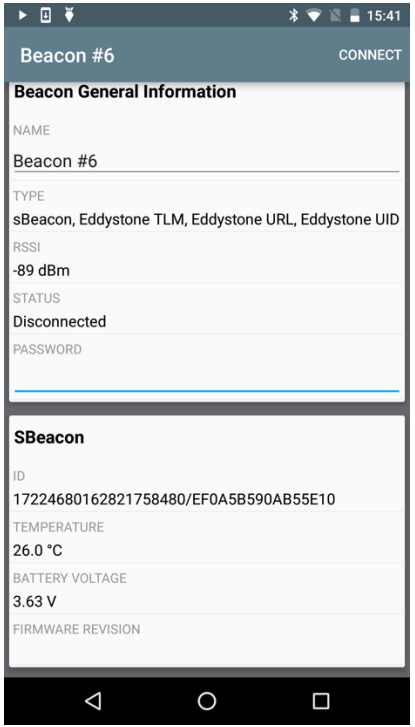
ACTION	UPDATE STATUS	MAC ADDRESS	BEACON ID	NEAREST ROOM	MAINTENANCE
Edit	up-to-date	C9:ED:20:D2:6A:5F	8EC968106F2	01.05.038	PKT COUNT: 35155830; VOLTAGE: 3.6000001; UPTIME: 459853408
Edit	up-to-date	EB:26:C9:E4:DC:5B	C28C3BC4AE8	01.05.038	PKT COUNT: 24238715; VOLTAGE: 3.6000001; UPTIME: 459856770
Edit	up-to-date	D6:82:27:05:E4:0B	2F0446B54A6	01.05.038	PKT COUNT: 31240056; VOLTAGE: 3.6000001; UPTIME: 459853016

Usability Study

- Basic smartphone tool
  - Six times longer, manual vs. automatic
  - Lowest rate 58% + only 1/3 got entirely right







Challenges behind

- The issues of being tiny
  - Battery only
  - No wire connectivity

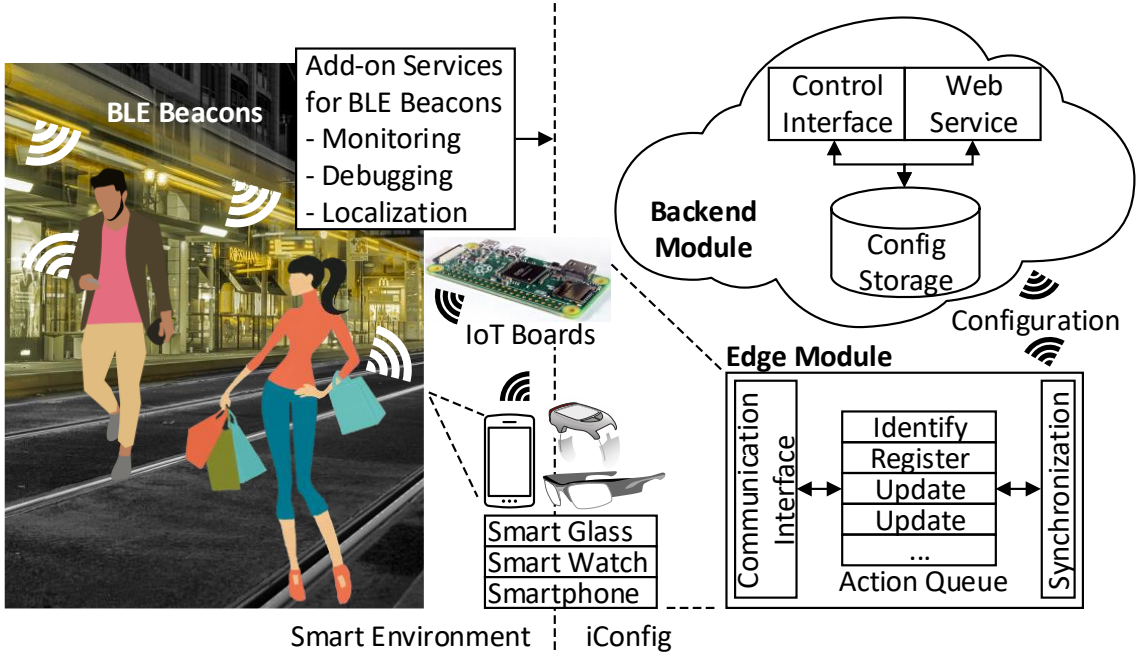
What's missing ?

- Device life cycles
  - Registration: location, tags
  - Configuration
  - Update
  - Monitoring
  - Debugging

- More automation

iConfig

- For configuration automation
  - Portable frontend + central backend

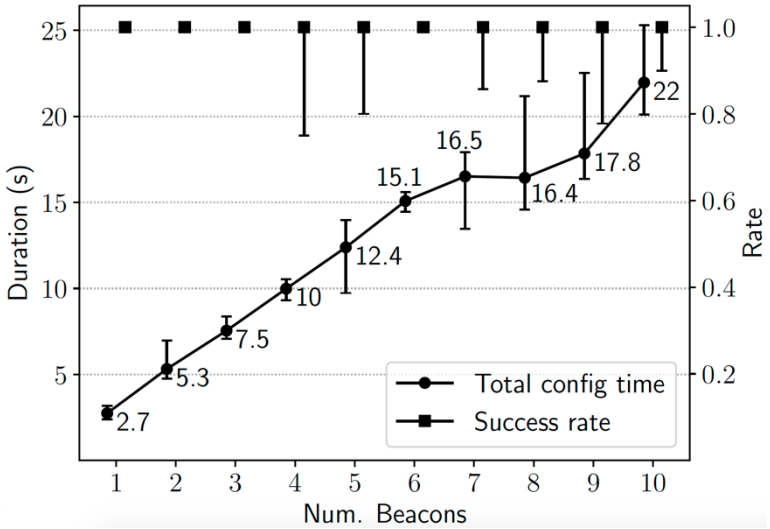


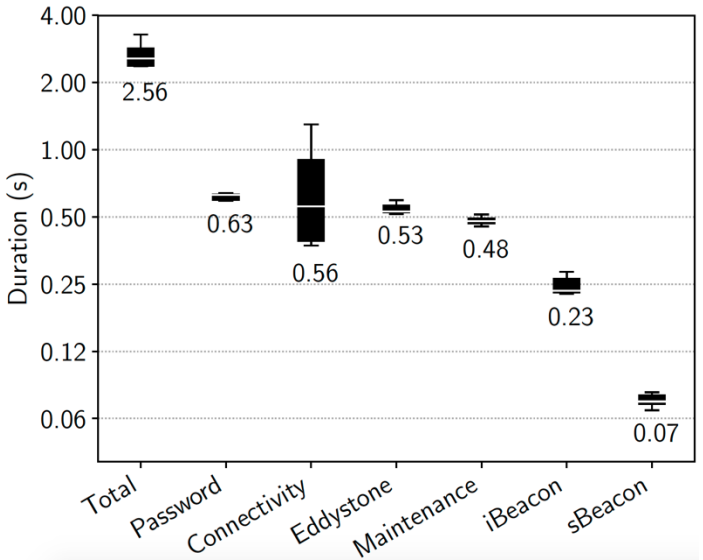
Does it work ?

- Testbed experiments
  - ~2.2s per beacon
  - Password 25%

But not enough:

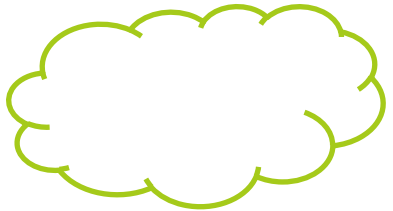
- Novelty ?
- Another App ?





Wait …

- Something's missing
  - Conventional design
  - Web UI, cloud backend
  - Old-fashion interaction
  - Poor usability



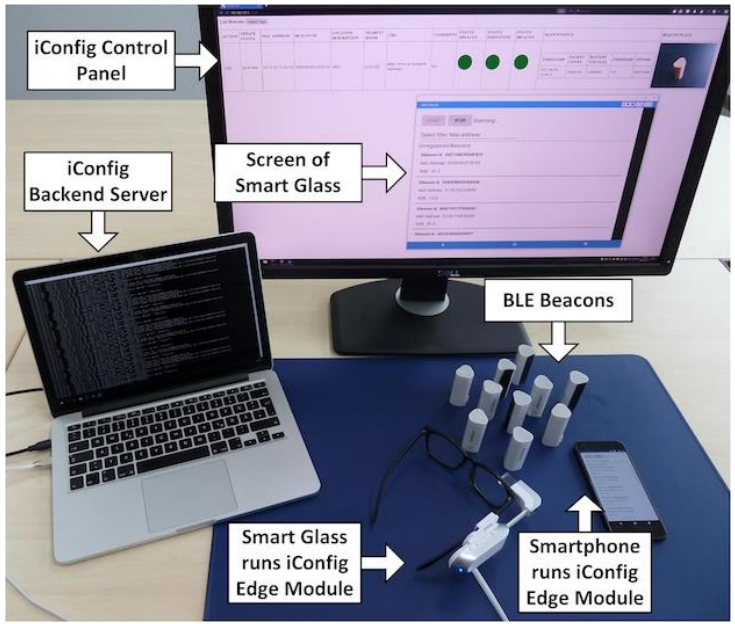
ACTION	UPDATE STATUS	MAC ADDRESS	BEACON ID	NEAREST ROOM	MAINTENANCE
Edit	up-to-date	C9:ED:20:D2:6A:5F	8EC968106F2	01.05.038	PKT COUNT: 35155830; VOLTAGE: 3.6000001; UPTIME: 459853408
Edit	up-to-date	EB:26:C9:E4:DC:5B	C28C3BC4AE8	01.05.038	PKT COUNT: 24238715; VOLTAGE: 3.6000001; UPTIME: 459856770
Edit	up-to-date	D6:82:27:05:E4:0B	2F0446B54A6	01.05.038	PKT COUNT: 31240056; VOLTAGE: 3.6000001; UPTIME: 459853016

Next shot

- Voice!
  - No hands

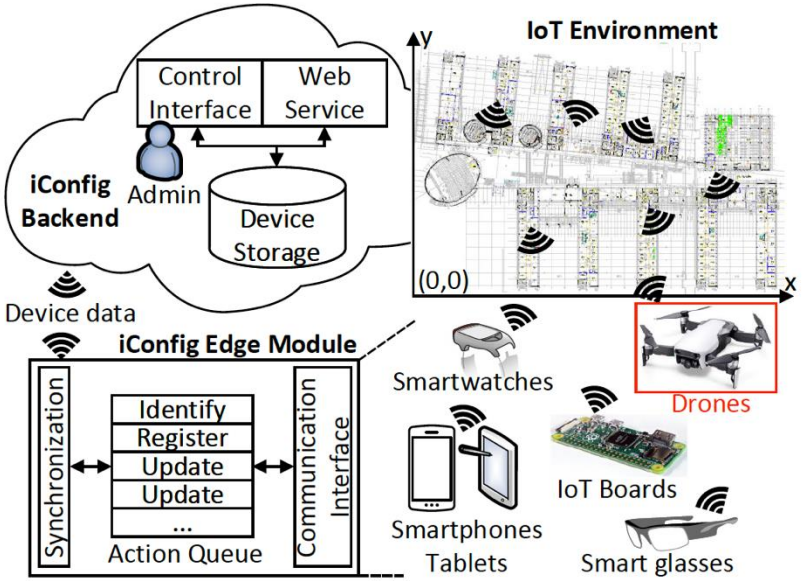
what I see is what i Configure

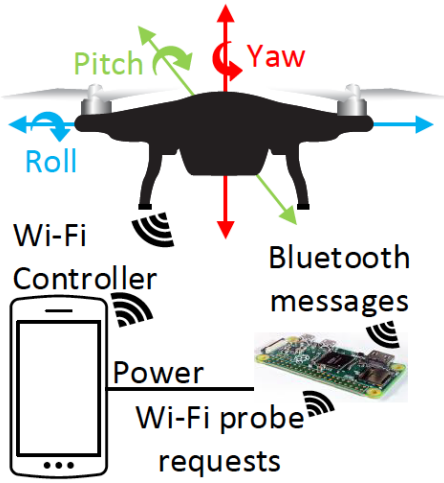
- Use cases
  - Configuration Automation
  - Debug + Monitor
  - Energy-aware management



Sky is the limit ☺

- Drone based IoT management
  - ACM SIGCOMM MAGESys





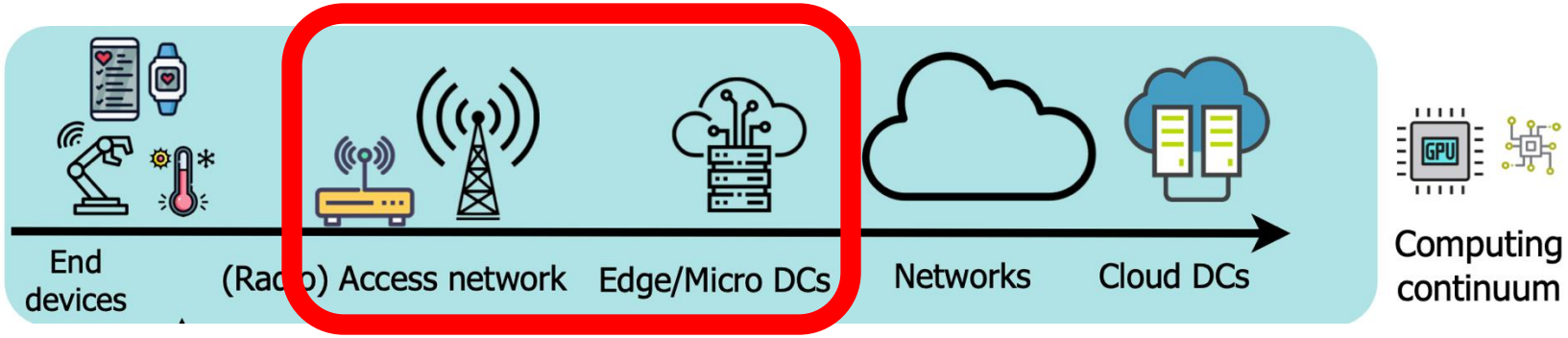
Lesson #1

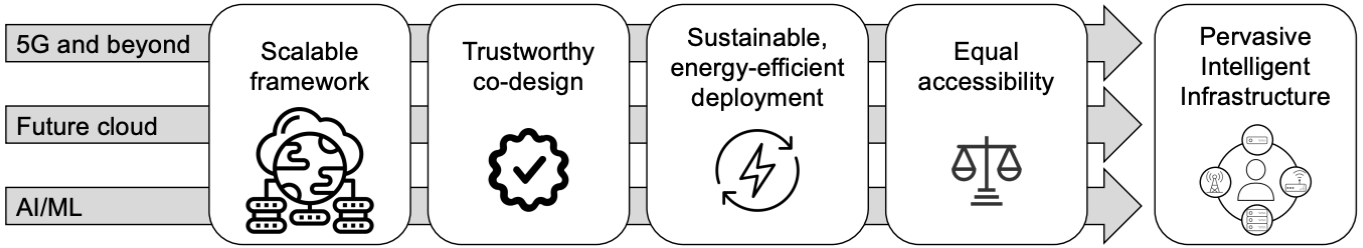
Tiny things hard to manage

Recycle is a MUST for Sustainability

Tools (Edge) are Missing

Bigger but not Static: Programmable ?





"Roadmap of Edge AI: A Dagstuhl Perspective", ACM SIGCOMM CCR 2022

Lesson #2

“ What’s **Killer App** on the Edge? ”

Case: Crowd Intelligence on Edge

- **Societal impact** of past years
  - Responding and coping with emergency/pandemics
  - Urban activity/mobility sensing on the edge

Motivation

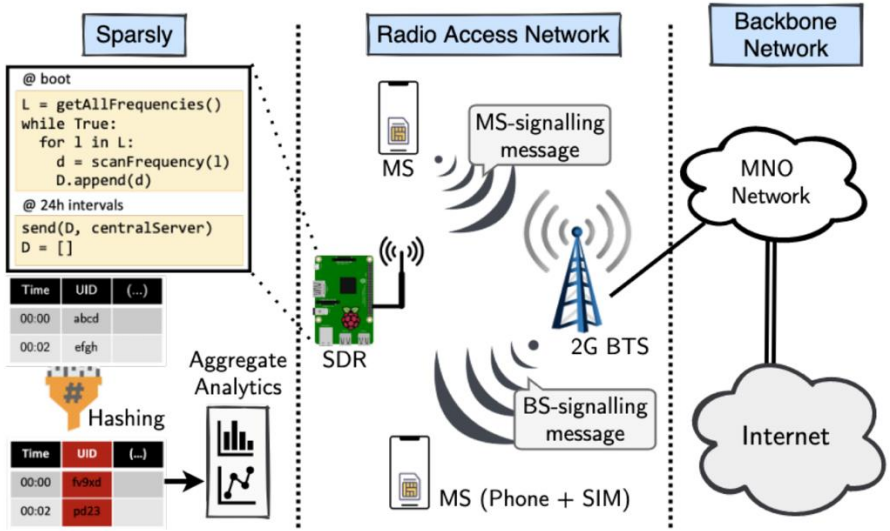
- Low cost and scalable
  - User equipment
  - Deployment and coverage

Gap: high fidelity entails high cost, infrastructure dependency, privacy intervention

- Passive (non-intrusive)
  - No need to force user interactions nor mandatory engagement
- Privacy-aware/friendly
  - Balance fidelity and data (local) regulations

Design

- SDR: Software Defined Radio
- Cellular signal probing
- Paging requests (PR, BS-signaling) and location update requests (LUR, MS-signaling)
- International Mobile Subscriber Identity (IMSI) for sensing purpose
- R-Pi with Noolec NESDR dongle, spectrum (900MHz)

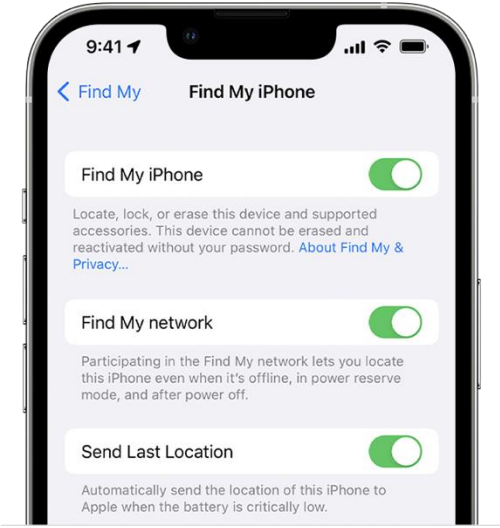


Unexpected

- Project ends…
  - Regulatory and legal considerations
  - Privacy in local context

Reusable is Key

- Reflections
  - Privacy on Edge? regulations and legal
  - Difference across countries
  - Programmability has saved us



Where is my Tag? Unveiling Alternative Uses of Apple FindMy Service

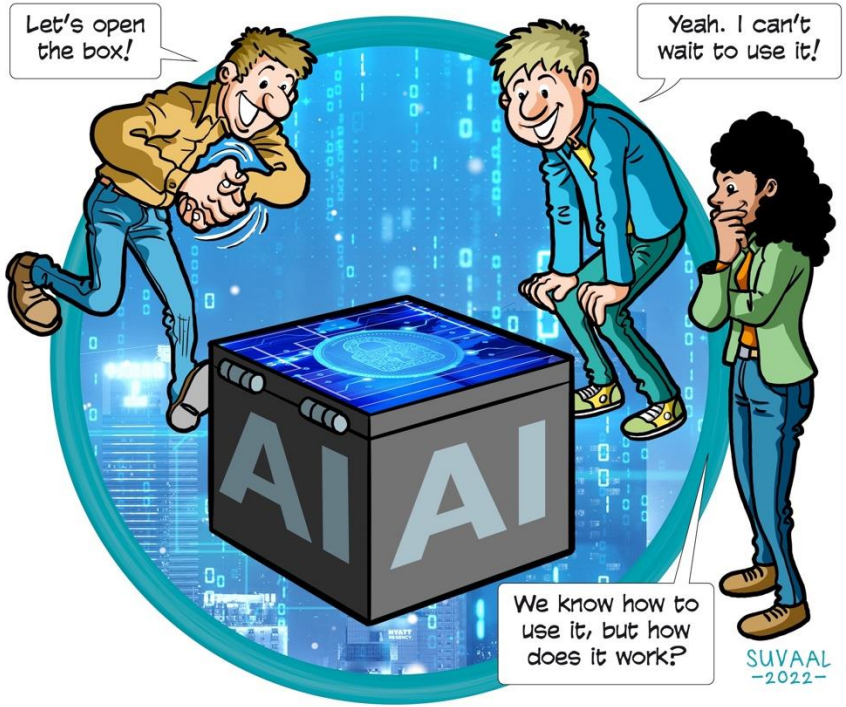
"Learn from the mistakes of others. You can't live long enough to make them all yourself." - Eleanor Roosevelt

Lesson to Action

Industry + Academia

EU Horizon Project

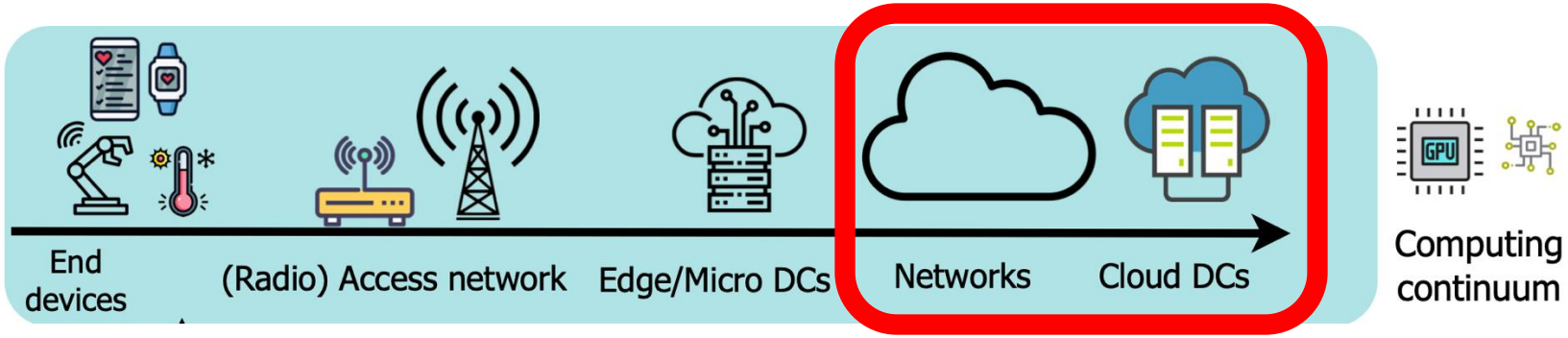
Score: 98/100 | Rate: 8%

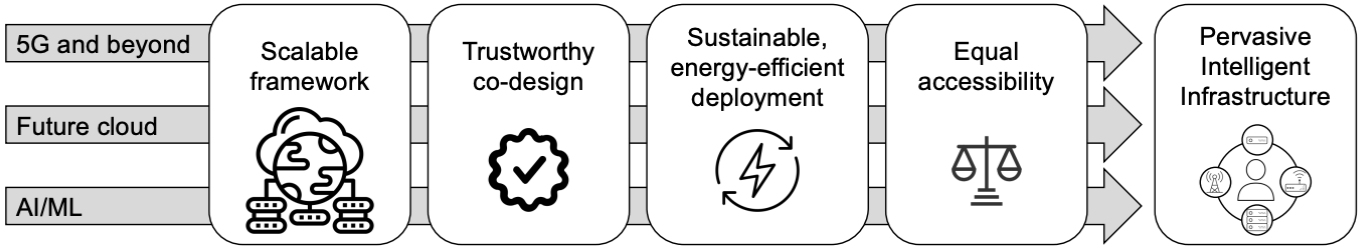


Funded by the European Union

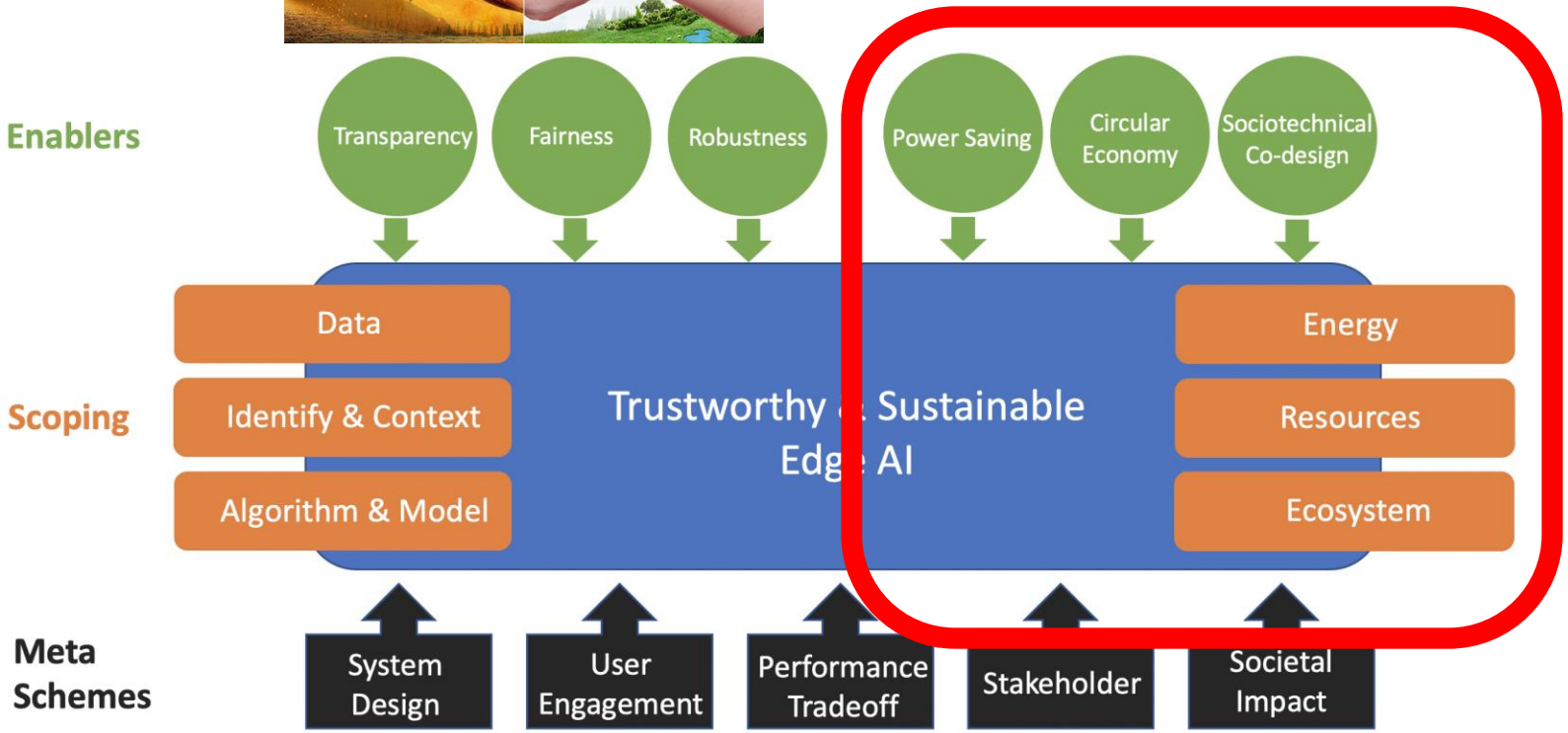
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101021808.

Huge but not Monster: Energy ?





"Roadmap of Edge AI: A Dagstuhl Perspective", ACM SIGCOMM CCR 2022

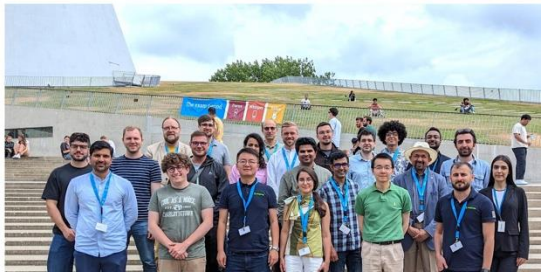


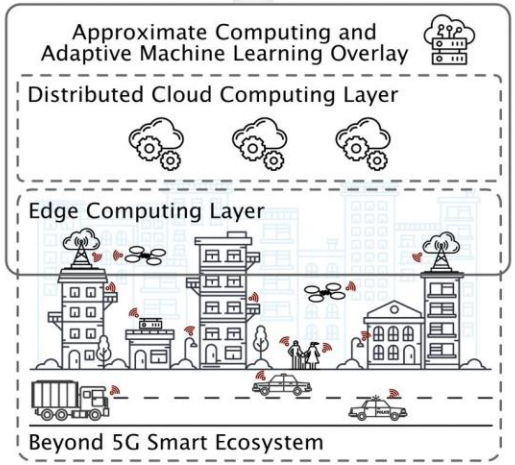
Aaron Ding, Marijn Janssen, Jon Crowcroft. "Trustworthy and Sustainable Edge AI: A Research Agenda", IEEE TPS

Sustainable Edge AI

- EU Marie Curie grant of €4M
- 20+ industrial academic partners

Approximate Computing for Power and Energy Optimisation





APROPOS Project
Sustainable AI

Score: 14.5/15 Rate: 3%

Vehicular Data

?

- Data increases
- Electric cars: battery life matters!

750MB per second, as Google's driverless car prototype reported

Autonomous Car - Sensors & Data		
Sensors	# Number	Data Volume
Camera	(8-12)	500 - 3500 Mbit/s
LiDAR	(2-4)	20-100 Mbit/s
Radar	(4-6)	0.1-15 Mbit/s
GPS		50 Kb/s
Ultrasonic	(8-16)	500-3500 Mb/s
	20 TB Car/Day	

Source: Elektrobit

Lesson #4

There is no free lunch

Safety Critical

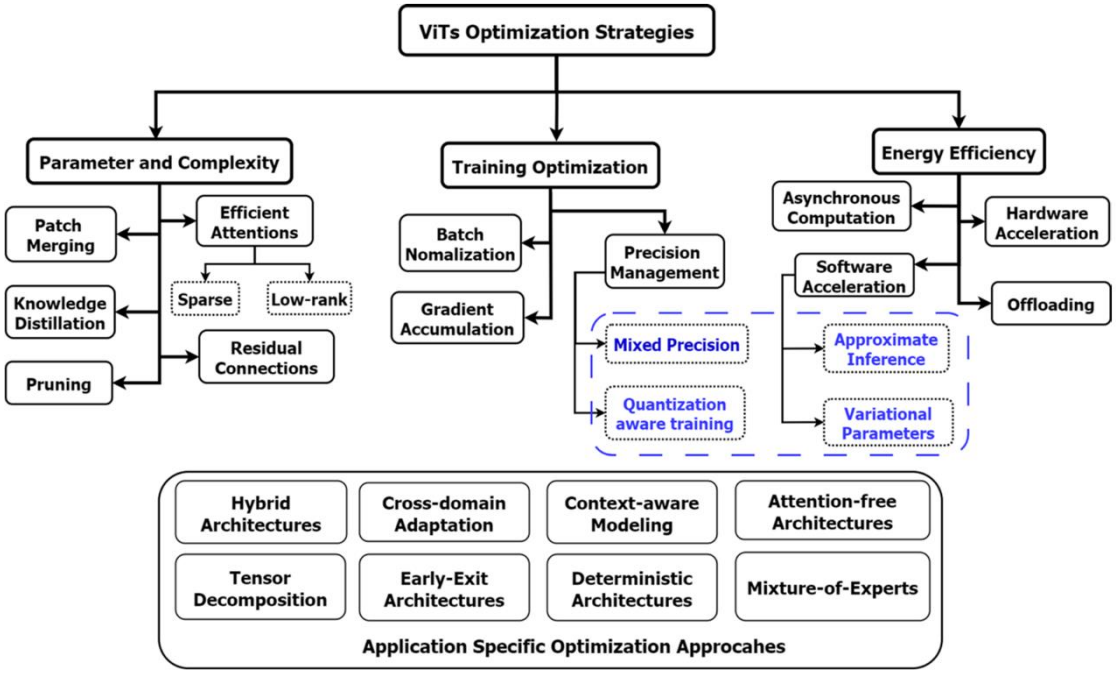
vs

Energy Saving

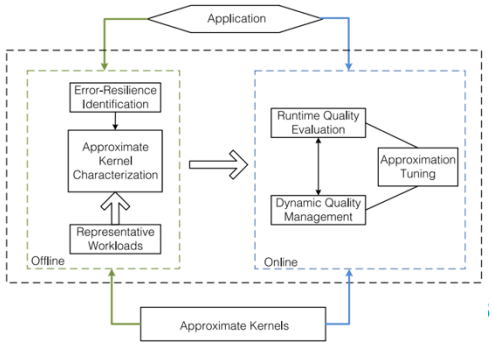
Can we strike both ?

Energy Awareness

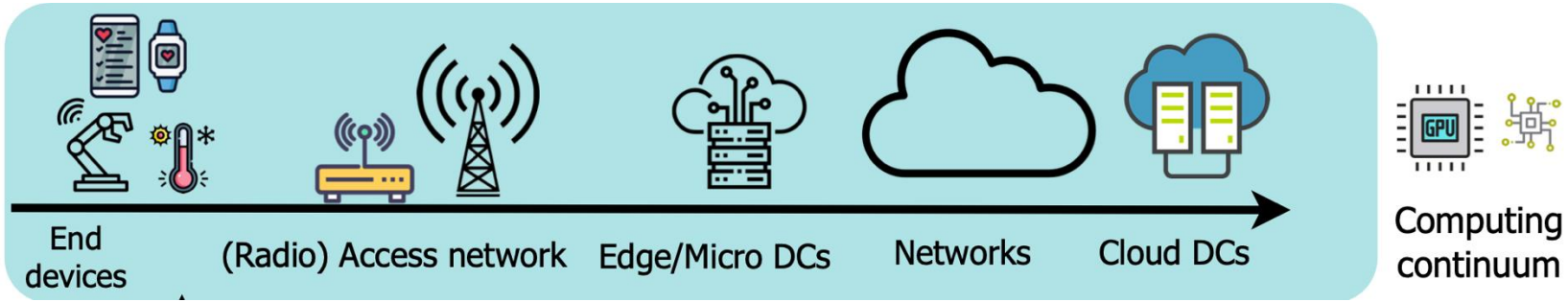
- Adaptive Approximation



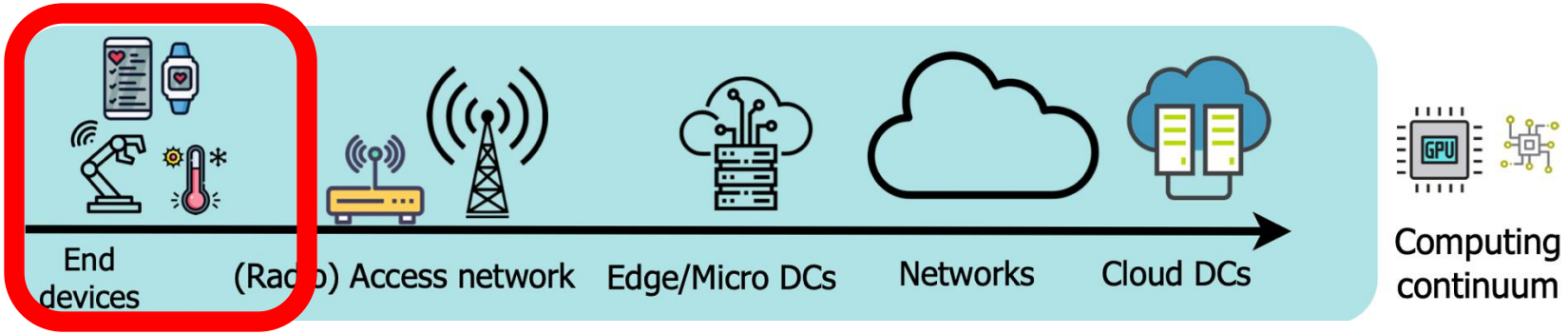
Year	Dataset	Sensors Included						
		Camera	LIDAR	Radar	GPS/GNSS	IMU	HD MAP	URL
2012 - 2022	KITTI [205]	Y	Y	N	N	Y	N	KITTI
2015 - 2019	KAIST Dataset [51]	Y	Y	N	Y	Y	N	KAIST
2016	HD1K [143]	Y	Y	N	N	N	N	HD1K
2016	CVC-14 [130]	Y	N	N	N	N	N	CVC-14
2016	Brain4Cars [120]	Y	N	Y	Y	N	N	Brain4Cars
2016	JAAD [146]	Y	N	N	N	N	N	JAAD
2016	Cityscapes [53]	Y	N	N	Y	Y	N	CITYSCAPES
2016	Udacity	Y	N	N	N	N	N	UdaCity
2016 - 2019	comma.ai driving dataset [264]	Y	N	Y	Y	Y	N	Comma datasets
2017	TRoM [192]	Y	N	N	N	N	N	TRoM
2017	Raincouver [309]	Y	N	N	N	N	N	Raincouver
2017	VPGNet [159]	Y	N	N	Y	N	N	VPGNet
2017	TuSimple	Y	N	Y	N	N	N	TuSimple
2017	TorontoCity [326]	Y	Y	N	N	N	N	TorontoCity
2017	CityPersons	Y	N	N	N	Y	N	CityPersons
2017	Mapillary Vistas [221]	Y	N	N	N	N	N	Mapillary Vistas
2017	Multi-spectral (Univ of Tokyo) [92]	Y	N	Y	N	N	N	Multi-spectral
2018	CULane [233]	Y	N	N	Y	Y	N	CULane
2018	DBNet [47]	Y	Y	Y	Y	Y	N	DBNet
2018	IDD [315]	Y	N	N	N	N	N	IDD
2018	MVSEC (U Penn) [403]	Y	Y	N	N	N	N	MVSEC
2018	NightOwls [222]	Y	N	N	N	N	N	NightOwls
2018	Road Damage [199]	Y	N	N	N	N	N	Road Damage
2018	Wilddash [377]	Y	N	N	N	N	N	wildDash
2018 - 2020	BDD-100K [370]	Y	Y	N	Y	Y	N	Berkeley
2018 - 2020	ApolloScape [113]	Y	Y	N	Y	Y	N	Apollo
2018 - 2020	Honda Driving [259]	Y	Y	N	Y	Y	N	HDD
2019	Argoverse [41]	Y	Y	N	N	N	Y	Argo
2019	Astyx HiRes [207]	Y	Y	N	N	N	N	Astyx
2019	BLVD [352]	Y	Y	N	N	N	N	BLVD
2019	Boxy Driving [24]	Y	N	N	N	N	N	BOSCH
2019	EuroCity [34]	Y	N	N	N	N	N	Eurocity Persons
2019	EU Long-term Dataset [354]	Y	Y	Y	Y	Y	N	EU Dataset
2019	IceVisionSet [240]	Y	Y	N	Y	Y	N	IceVision
2019	StreetLearn [209]	Y	N	N	N	N	N	Street Learn
2019	PandaSet	Y	Y	N	Y	N	N	PandaSet
2019	WoodScape [368]	Y	Y	N	Y	Y	N	WoodScape
2019	Unsupervised Llamas - Bosch [25]	Y	Y	N	Y	N	N	Bosch
2020	4—Seasons [336]	Y	N	N	Y	Y	N	4-Seasons
2020	A*3D [242]	Y	Y	N	N	N	N	A3TAR-3D
2020	nuScenes [37]	Y	Y	Y	Y	Y	N	nuscenes
2020	POSS [234]	Y	Y	N	N	N	N	POSS
2020	DDD20 [108]	Y	N	N	Y	Y	N	DDD20
2020	Highway Driving [139]	Y	N	N	N	N	N	Kaist
2020	Lyft Level 5 [106]	Y	Y	N	N	N	Y	lyft
2020	Brno Urban Dataset	Y	Y	Y	Y	Y	N	BRNO
2020	Ford Multi AV [4]	Y	Y	N	Y	Y	Y	Ford Seasonal
2020	A2D2 [80]	Y	Y	N	N	N	N	Audi
2020	LIBRE [38]	Y	Y	Y	Y	Y	N	LIBRE
2020	Toronto-3D	Y	Y	N	Y	Y	N	Toronto-3D
2021	NEOLIX [324]	Y	Y	Y	Y	Y	N	Neolix
2021	CADC [244]	Y	Y	N	Y	Y	N	CADC
2021	RadarScenes [266]	Y	N	Y	Y	Y	N	RadarScenes
2021	CARRADA [230]	Y	N	Y	N	N	N	CARRADA
2021	Waymo [290]	Y	Y	N	N	N	Y	Waymo Open
2021	SODA10M [93]	Y	N	N	N	N	N	SODA10M
2021	PixSet/Leddar Tech [61]	Y	Y	Y	Y	Y	N	PixSet
2021	ONCE [202]	Y	Y	N	N	N	N	ONCE
2021	Deep Route AI	Y	Y	Y	Y	Y	Y	Deep Route
2021	DurLAR[169]	Y	Y	N	Y	Y	N	DurLAR
2022	MUAD[75]	Y	N	N	N	N	N	MUAD
2022	SHIFT	Y	Y	N	N	Y	N	SHIFT
2022	Rope3D[364]	Y	Y	N	Y	N	N	Rope3D
2022	CODA[168]	Y	Y	N	Y	Y	N	CODA
2022	View-of-Delft [232]	Y	Y	Y	Y	Y	N	Delft-View
2023	LiDar-CS [70]	N	Y	N	N	N	Y	LiDar-CS
2023	ZoD [11]	Y	Y	N	Y	Y	N	ZoD
2023	Race-Car [151]	Y	Y	Y	Y	Y	N	Race-Car

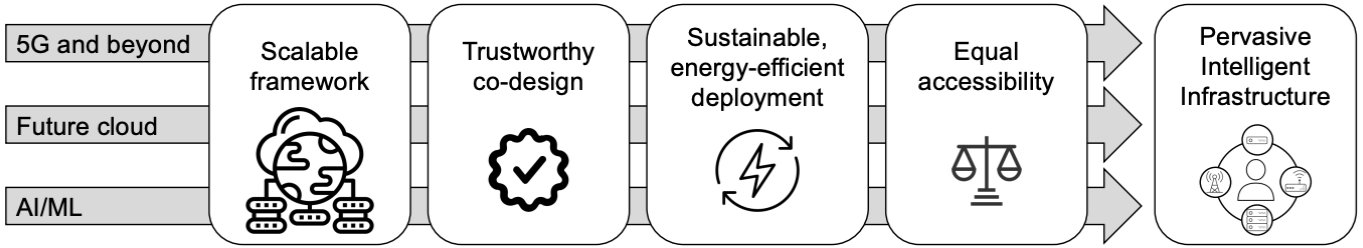


Why + Lessons + ***Directions ?***



Tiny but not Waste: Safe, Circular





"Roadmap of Edge AI: A Dagstuhl Perspective", ACM SIGCOMM CCR 2022

Tiny but not Waste: Safe, Circular

"IoT-KEEPER: Detecting Malicious IoT Network Activity using Online Traffic Analysis at the Edge" ***IEEE TNSM 2020, Q1 Journal IF 4.7***

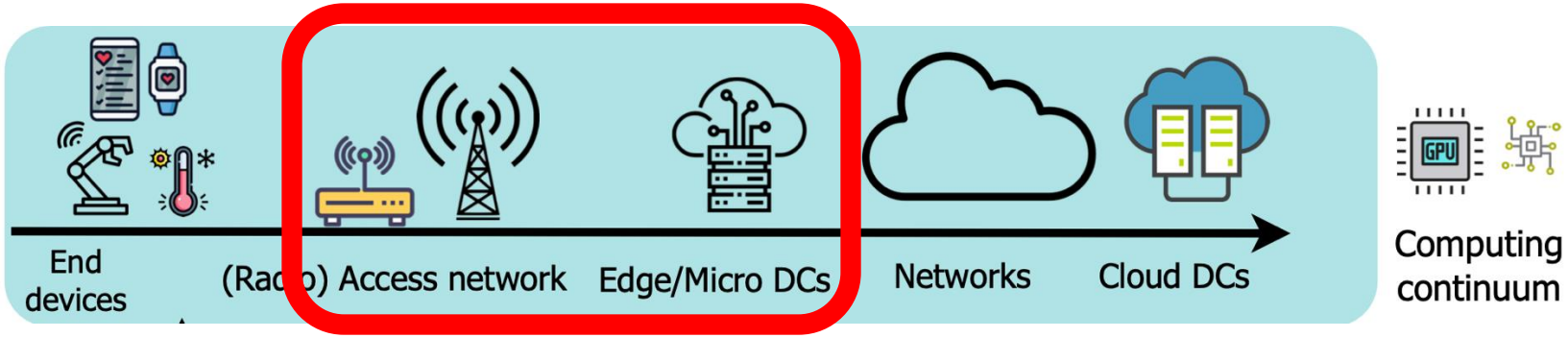
"Context-dependent Trade-offs around Platform-to-platform Openness: The Case of IoT" ***Elsevier Technovation 2021, Q1 Journal IF 11.1***

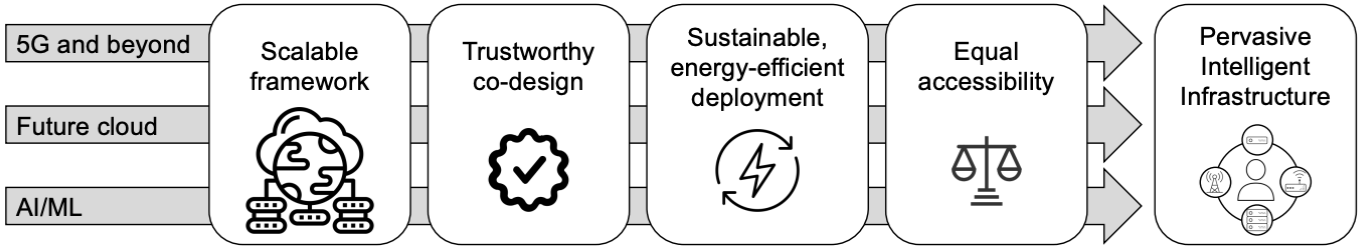
"Where Is My Tag? Unveiling Alternative Uses of the Apple FindMy Service" ***IEEE WoWMoM 2022, Core A***

"XAI for Accountable MI Detection in IoT Emergency Communication Systems" ***ACM IoT 2023, Best Paper Runner-up***

"ARASEC: Adaptive Resource Allocation and Model Training for Serverless Edge Computing" ***IEEE Internet Computing 2024, Q1 Journal IF 3.7***

Bigger but not Static: Trust + Reusable





"Roadmap of Edge AI: A Dagstuhl Perspective", ACM SIGCOMM CCR 2022

Bigger but not Static: Trust + Reusable

"Characterising the Role of Pre-Processing Parameters in Audio-based Embedded Machine Learning" ***ACM SenSys 2021, Core A****

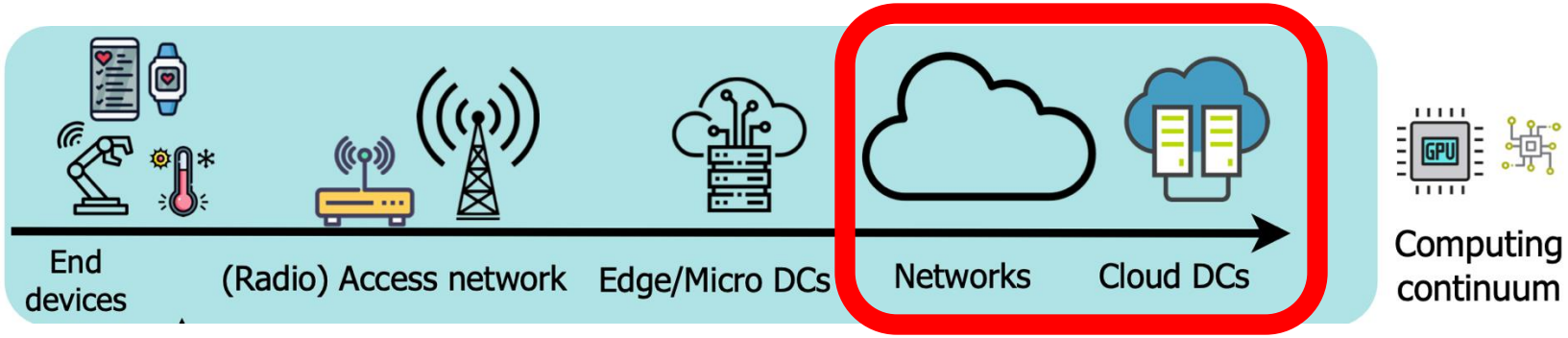
"Bias Detection and Generalization in AI Algorithms on Edge for Autonomous Driving" ***ACM/IEEE SEC 2022***

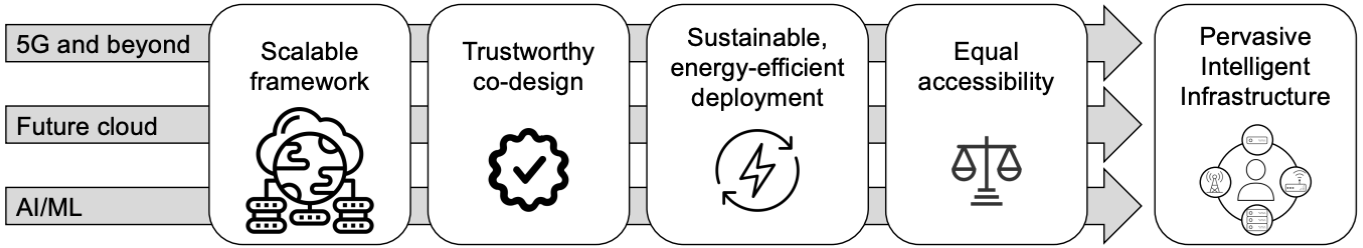
"Bias in Automated Speaker Recognition" ***ACM FAccT 2023 Best PhD Paper***

"Bias Propagation in On-device Machine Learning Workflows" ***ACM TOSEM 2023, Q1 Journal IF 6.6***

"SPATIAL Architecture: Gauging and Monitoring the AI Inference Capabilities of Modern Applications" ***IEEE ICDCS 2024, Core A****

Huge but not Monster: Approximate AI





"Roadmap of Edge AI: A Dagstuhl Perspective", ACM SIGCOMM CCR 2022

Huge but not Monster: Approximate AI

"Transfer Learning-Based Outdoor Position Recovery with Cellular Data"
IEEE TMC 2021, Q1 Journal IF 7.7

"Approximate Edge AI for Energy Efficient Autonomous Driving Services"
IEEE COMST 2023, Q1 Journal IF 35,6

"Nimbus: Towards Latency-Energy Efficient Task Offloading for AR Services"
IEEE TCC 2023, Q1 Journal IF 5,3

"Test-time Specialization of Dynamic Neural Networks"
IEEE CVPR MAT 2024, Best Paper Award

"Approximating Vision Transformers for Edge: Variational Inference and Mixed-precision for Multi-modal" ***Spring Nature Computing 2025, Q1***

Takeaway

EDITOR: Schahram Dustdar, dustdar@dsg.tuwien.ac.at

DEPARTMENT: INTERNET OF THINGS, PEOPLE, AND PROCESSES

Revisiting Edge AI: Opportunities and Challenges

Tobias Meuser, *Technical University of Darmstadt, 64283, Darmstadt, Germany*
Lauri Lovén, *University of Oulu, 90014 Oulu, Finland*
Monowar Bhuyan, *Umeå University, 90187 Umeå, Sweden*
Shishir G. Patil, *UC Berkeley, California, Berkeley, CA, 94709, USA*
Schahram Dustdar, *Vienna University of Technology, Vienna 1040, Austria*
Atakan Aral, *Umeå University, 90187 Umeå, Sweden and University of Vienna, Vienna 1090, Austria*
Suzan Bayhan, *University of Twente, 7500 AE, Enschede, The Netherlands*
Christian Becker, *University of Stuttgart, 70569, Stuttgart, Germany*
Eyal de Lara, *University of Toronto, Toronto, ON, M5S 1A1, Canada*
Aaron Yi Ding, *TU Delft, 2600 AA, Delft, The Netherlands*
Janick Edinger, *University of Hamburg, 22527, Hamburg, Germany*
James Gross, *Royal Institute of Technology Stockholm (KTH), 100 44 Stockholm, Sweden*
Nitinder Mohan, *Technical University of Munich, 80333, Munich, Germany*
Andy D. Pimentel, *University of Amsterdam, 1098 GH, Amsterdam, The Netherlands*
Etienne Rivière, *UCLouvain, B-1348, Louvain-la-Neuve, Belgium*
Henning Schulzrinne, *Columbia University, New York, NY, 10027, USA*
Pieter Simoens, *Ghent University-imec, B-9052, Gent, Belgium*
Gürkan Solmaz, *NEC Laboratories Europe, 69115, Heidelberg, Germany*
Michael Welzl, *University of Oslo, 0313, Oslo, Norway*

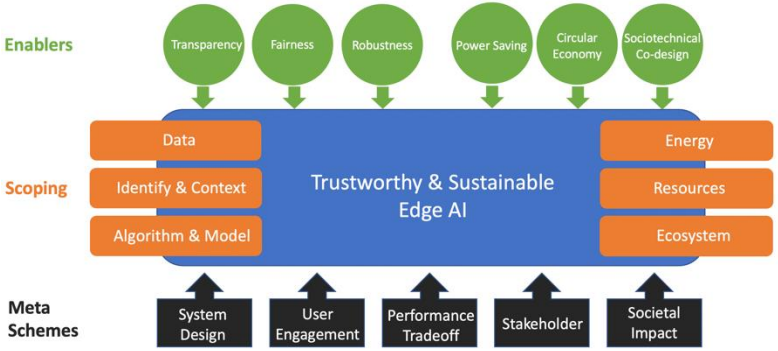
Edge artificial intelligence (AI) is an innovative computing paradigm that aims to shift the training and inference of machine learning models to the edge of the network. This paradigm offers the opportunity to significantly impact our everyday lives with new services such as autonomous driving and ubiquitous personalized health care. Nevertheless, bringing intelligence to the edge involves several major challenges, which include the need to constrain model architecture designs, the secure distribution and execution of the trained models, and the substantial network load required to distribute the models and data collected for training. In this article, we highlight key aspects in the development of edge AI in the past and connect them to current challenges. This article aims to identify research opportunities for edge AI, relevant to bring together the research in the fields of artificial intelligence and edge computing.

Roadmap for Edge AI: A Dagstuhl Perspective

Aaron Yi Ding[1]*, Ella Peltonen[2], Tobias Meuser[3], Atakan Aral[4], Christian Becker[5], Schahram Dustdar[6], Thomas Hiessl[6], Dieter Kranzlmüller[7], Madhusanka Liyanage[8], Setareh Magshudi[9], Nitinder Mohan[10], Jörg Ott[10], Jan S. Rellermeyer[11,1], Stefan Schulte[12], Henning Schulzrinne[13], Gürkan Solmaz[14], Sasu Tarkoma[15], Blesson Varghese[16], Lars Wolf[17]

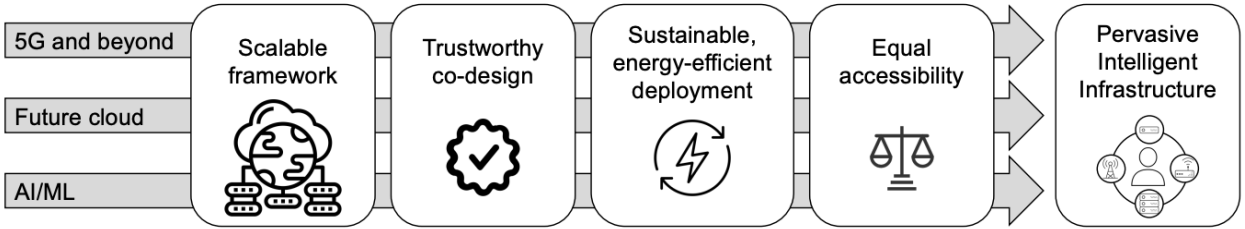
[1]TU Delft, [2]University of Oulu, [3]TU Darmstadt, [4]University of Vienna, [5]University of Mannheim, [6]TU Wien, [7]LMU Munich, [8]University College Dublin, [9]University of Tübingen, [10]TU Munich, [11]Leibniz University Hannover, [12]Hamburg University of Technology, [13]Columbia University, [14]NEC Lab, [15]University of Helsinki, [16]Queen's University Belfast, [17]TU Braunschweig

* Corresponding author: Aaron Ding (aaron.ding@tudelft.nl)



"Trustworthy and Sustainable Edge AI: A Research Agenda"

Outlook



acm sigcomm
ACM SIGCOMM CCR, Vol. 52, No.1
COMPUTER COMMUNICATION REVIEW