The Future of AI is on the Edge

Dr Varun Ojha

Associate Professor in Artificial Intelligence

School of Computing Newcastle University

AI Theme Leader, National Edge AI Hub

varun.ojha@newcastle.ac.uk

ojhavk.github.io

edgeaihub.co.uk

at

NCL Action Event, 27 Feb 2025

Newcastle University

National Edge AI Hub

UKRI Engineering and Physical Sciences Research Council

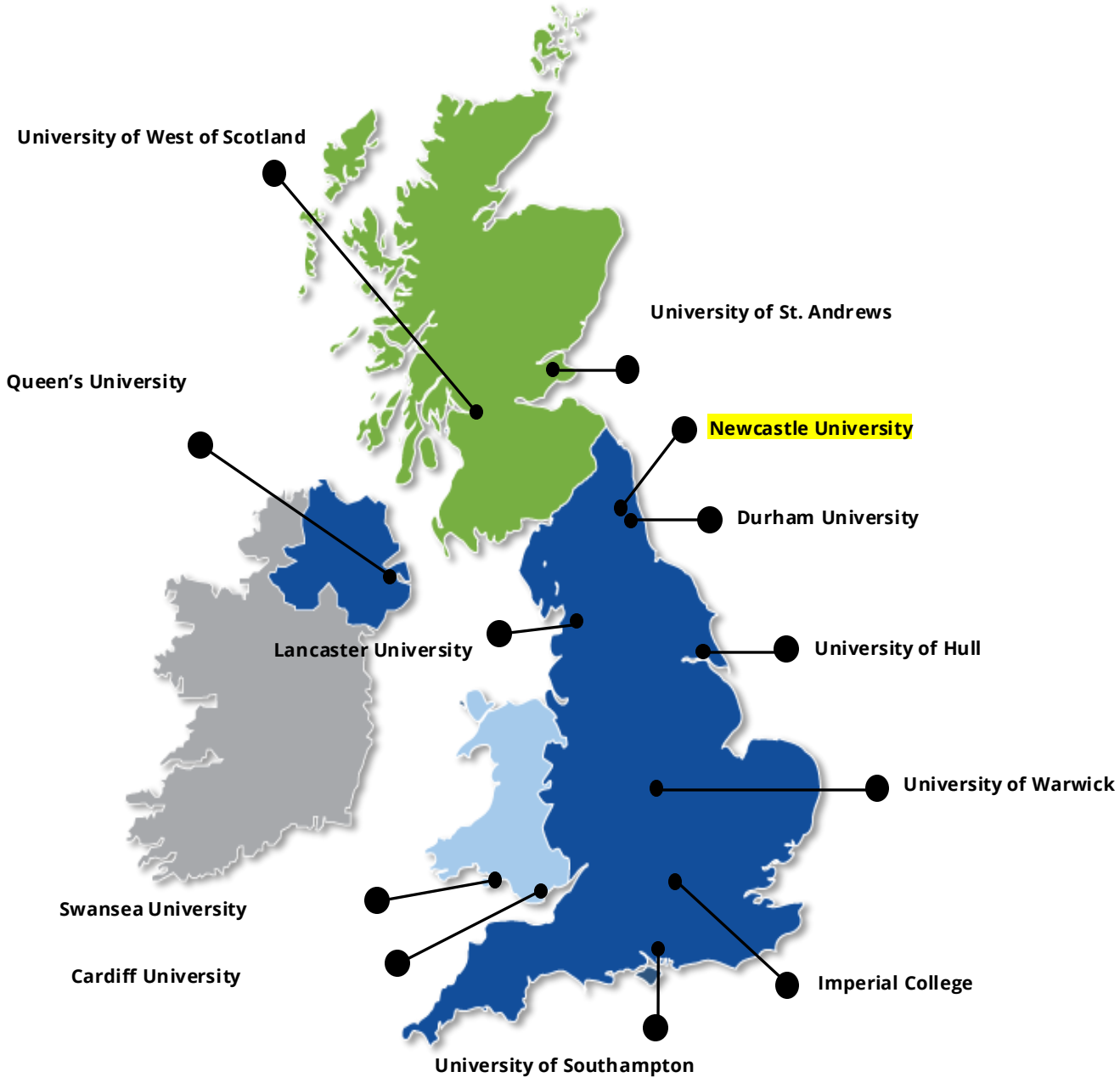
National Edge AI Hub

Engineering and Physical Sciences Research Council

Newcastle University

EPSRC £12 million investment

edgeaihub.co.uk



Artificial Intelligence (AI)

- **General-purpose AI** like the robots of science fiction is incredibly hard.
  - Human brain appears to have lots of special and general functions, integrated in some amazing way that we really do not understand at all (yet)

how far

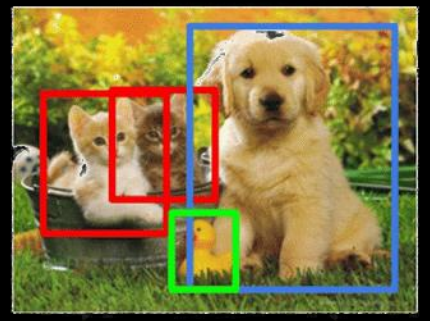
Grok runway

Gemini ChatGPT

how close

- **Special-purpose AI** is more doable (nontrivial)
  - E.g., chess/poker playing programs, logistics planning, automated translation, voice recognition, web search, data mining, medical diagnosis, keeping a car on the road

CAT



CAT, DOG, DUCK


1


2

Generative AI

Trained on AVA: A Large-Scale Database for Aesthetic Visual Analysis (255,000 images)

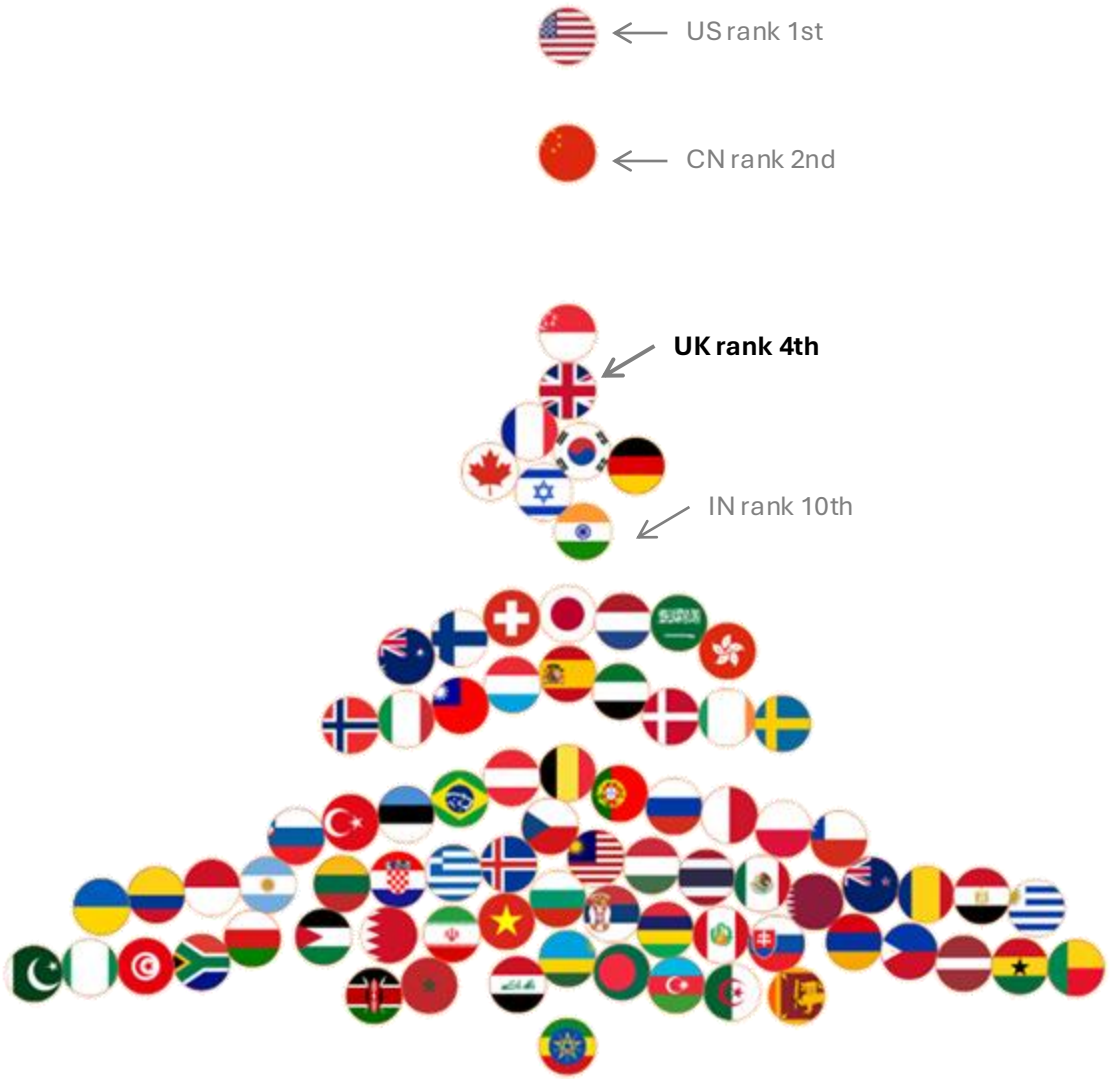
a dolphin in an astronaut suit on saturn, artstation

a propaganda poster depicting a cat dressed as french emperor napoleon holding a piece of cheese

a teddy bear on a skateboard in times square

DALL·E 2 Outputs

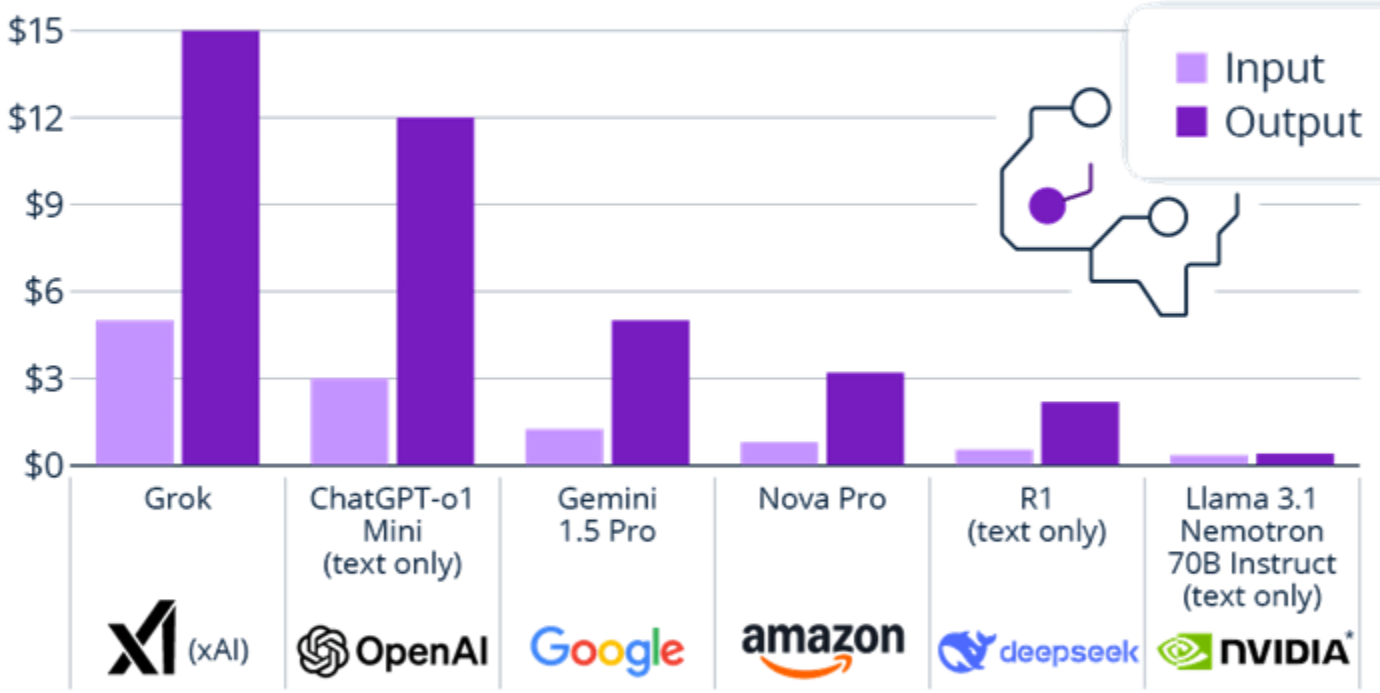
AI Arms Race



https://www.tortoisemedia.com/intelligence/global-ai

DeepSeek-R1 Upsets AI Market With Low Prices

Estimated price for processing one million input/output tokens on different AI models

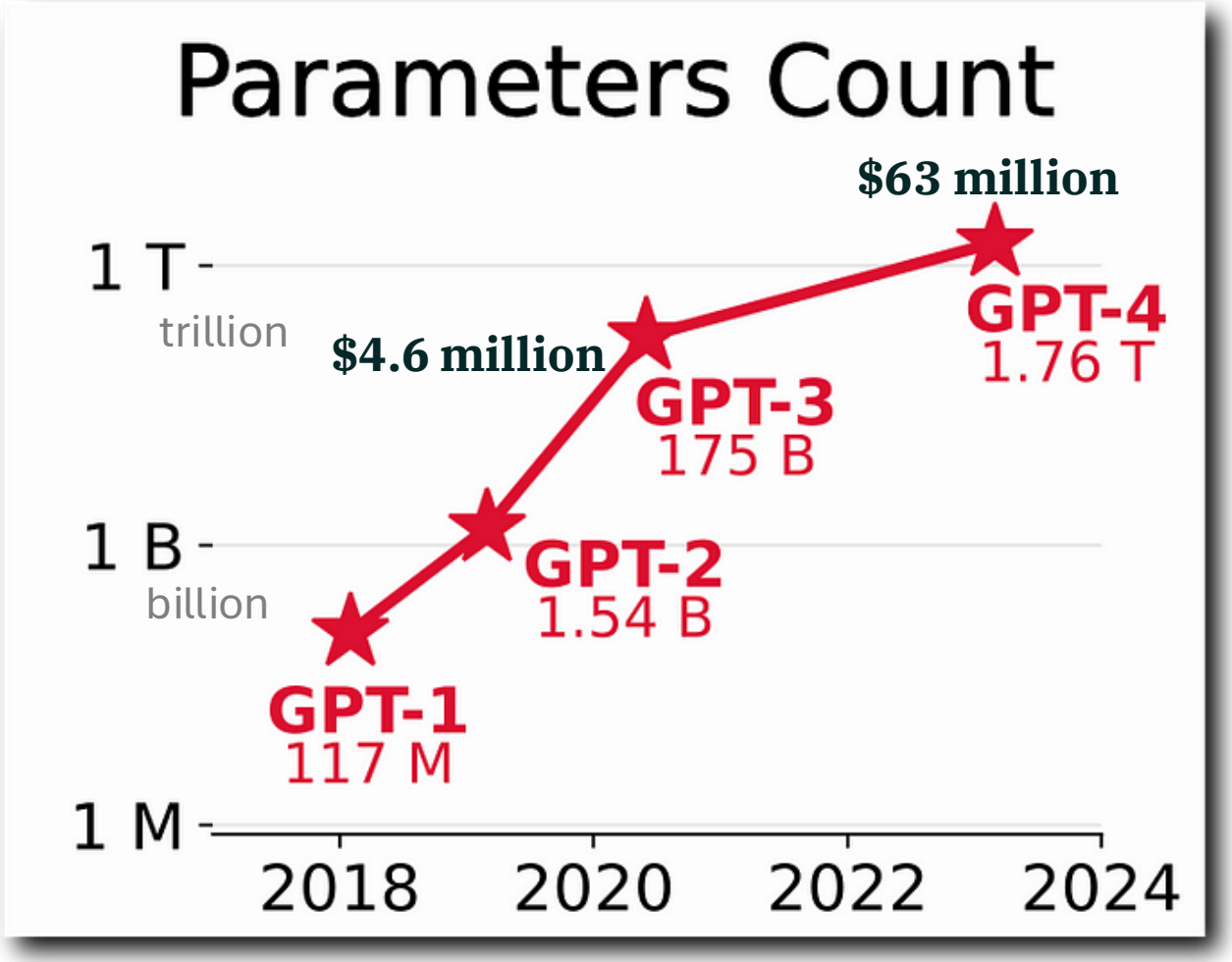


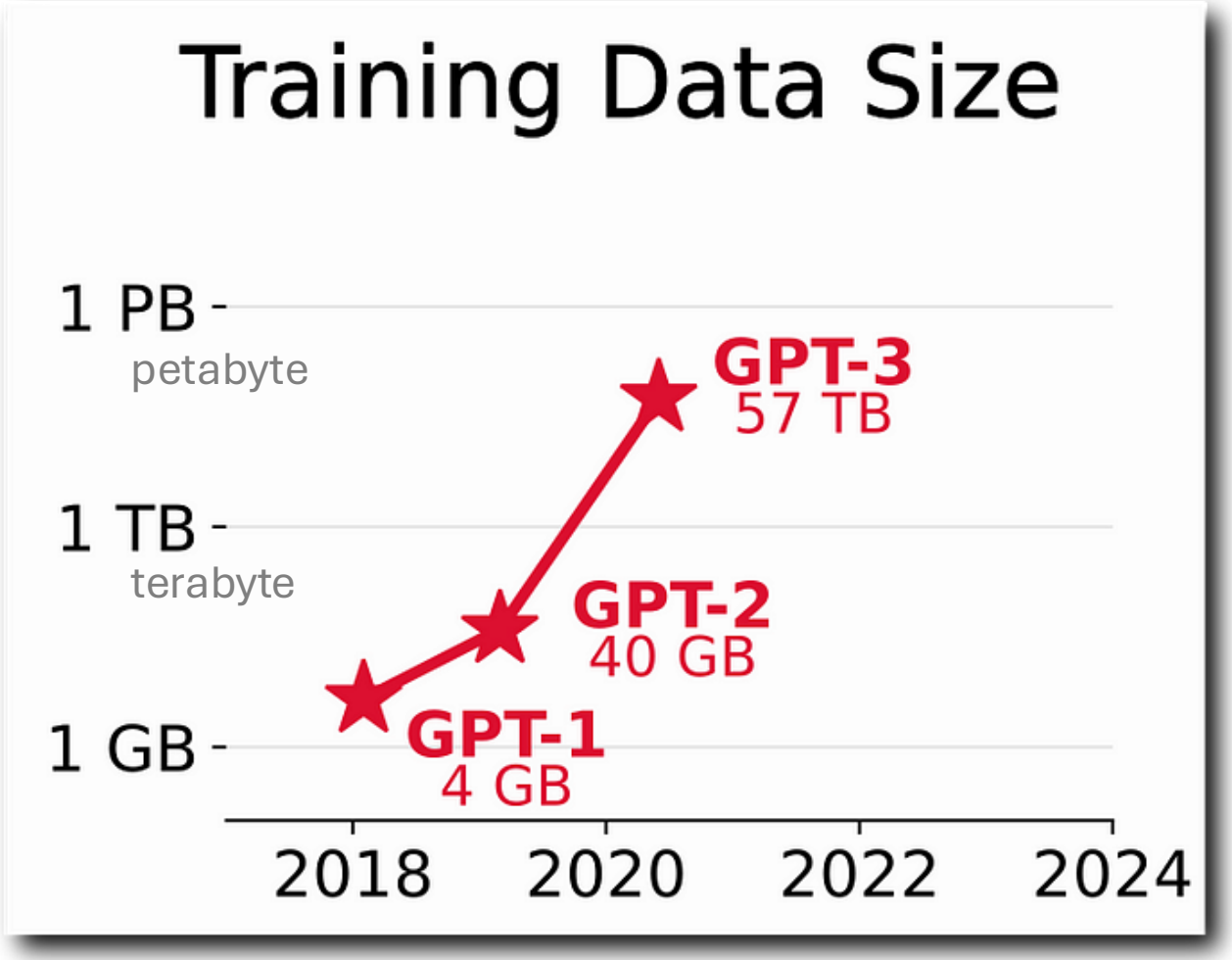
A token is the smallest unit of AI model processing (~4 characters).
o1 is ChatGPT's latest model. List includes most comparable model per company
* Uses Meta's open-source Llama AI

Source: DocsBot

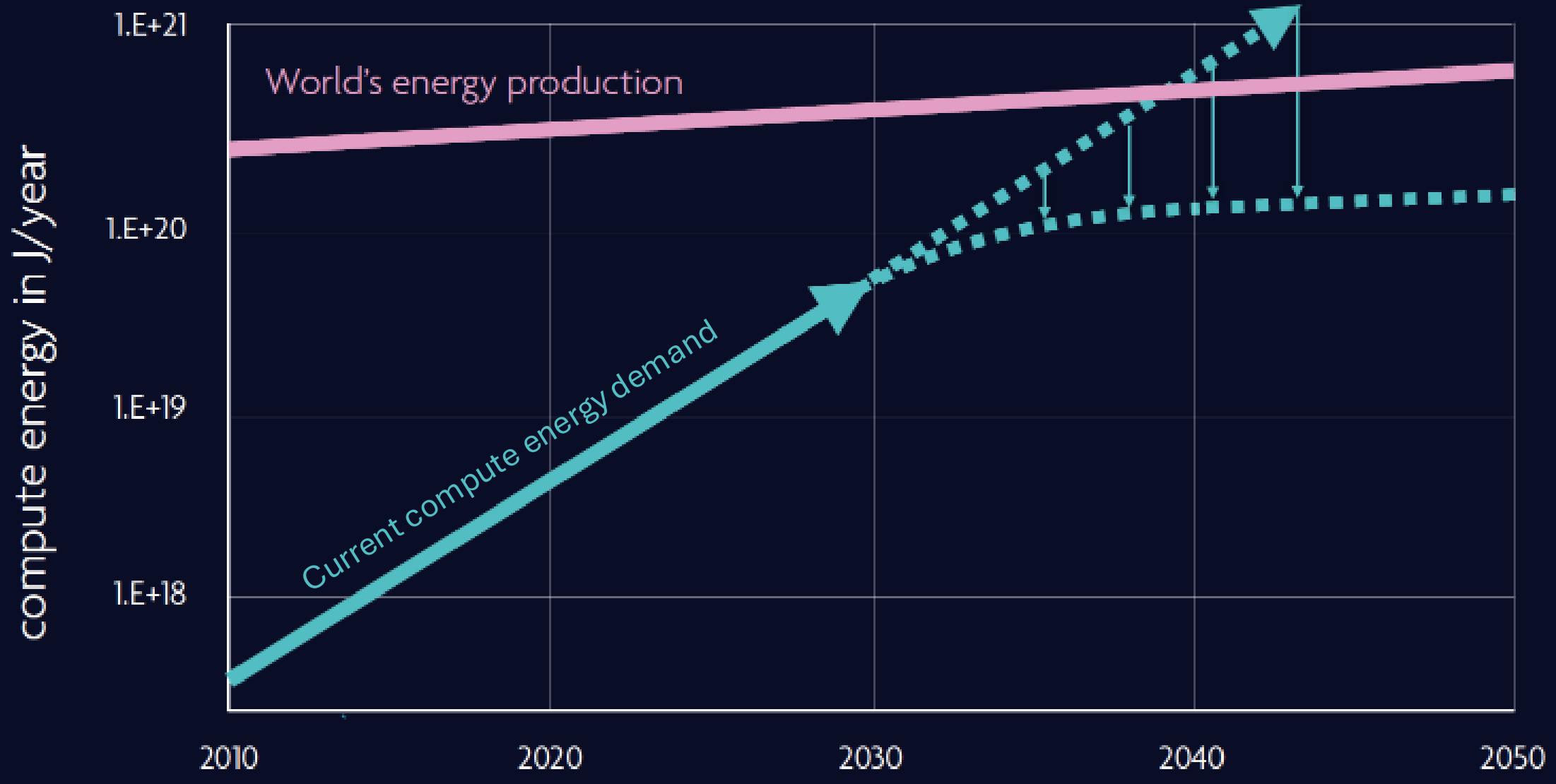
statista

Massive AI Model and Data Size





Source: Francesco Casalegno, ChatGPT Unveiled: What's the ML Model Inside it, from GPT-1 to GPT-4

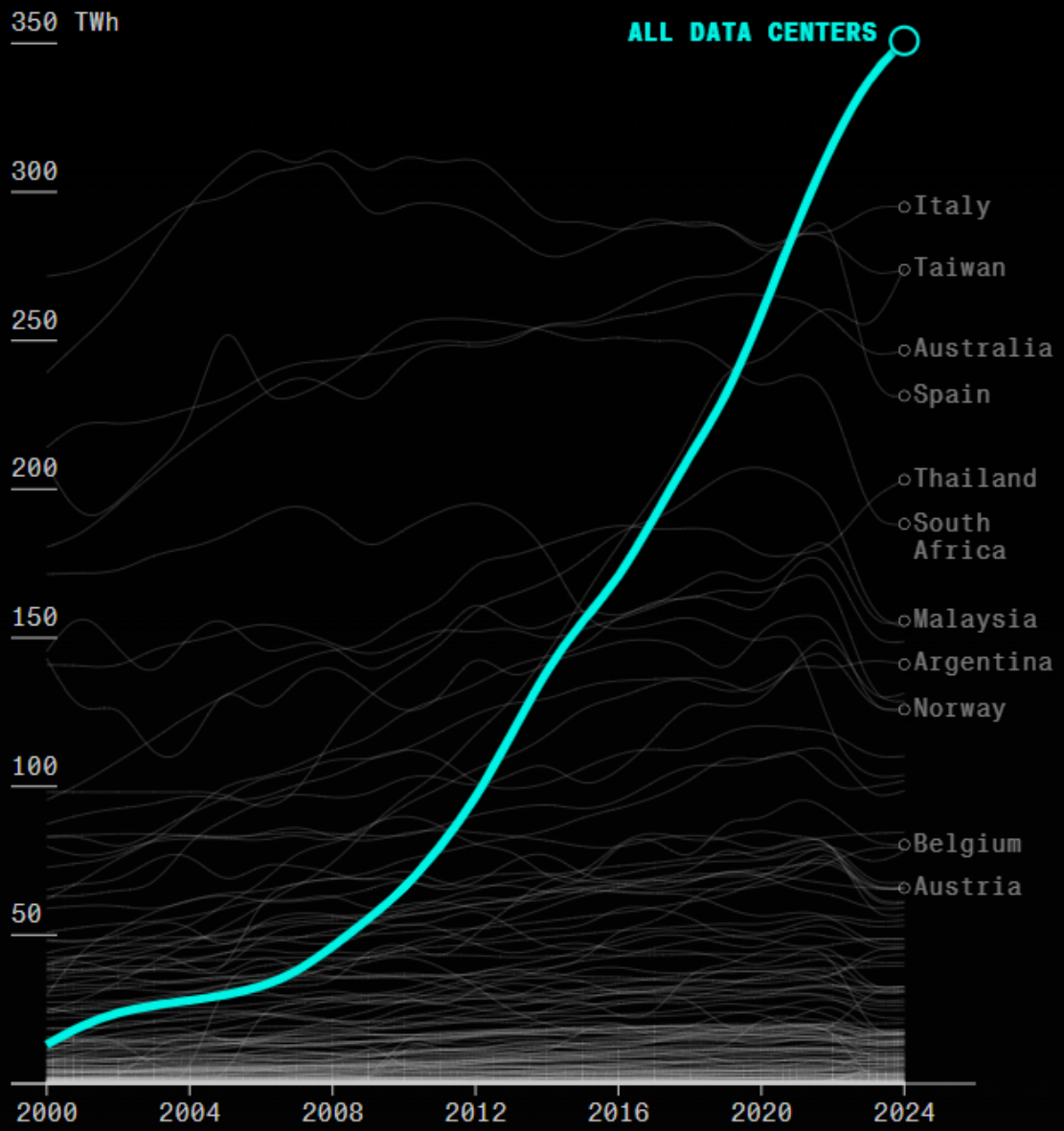


Source: SRC decadal plan 2020

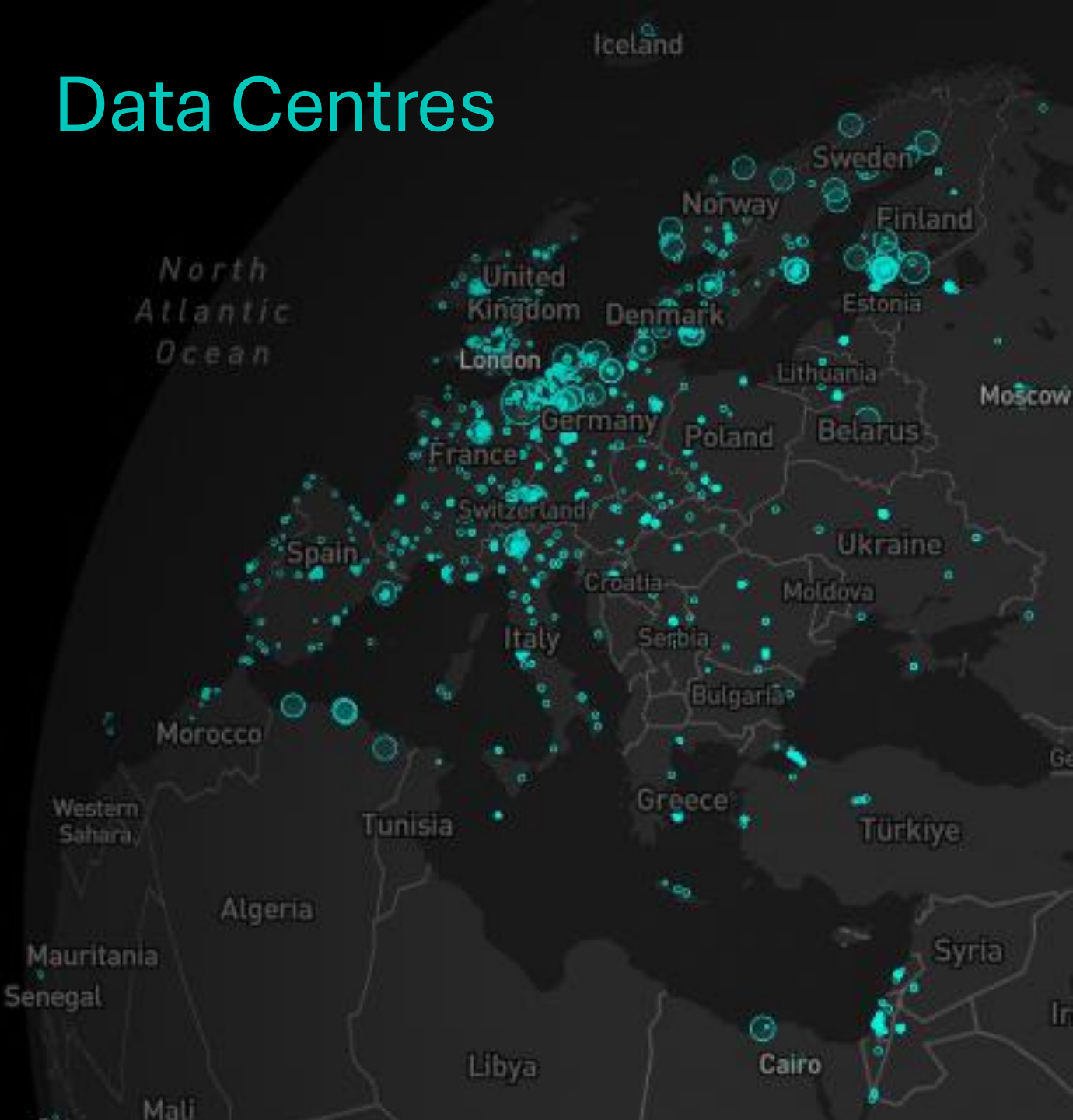
Altogether, data centers use more electricity than most countries

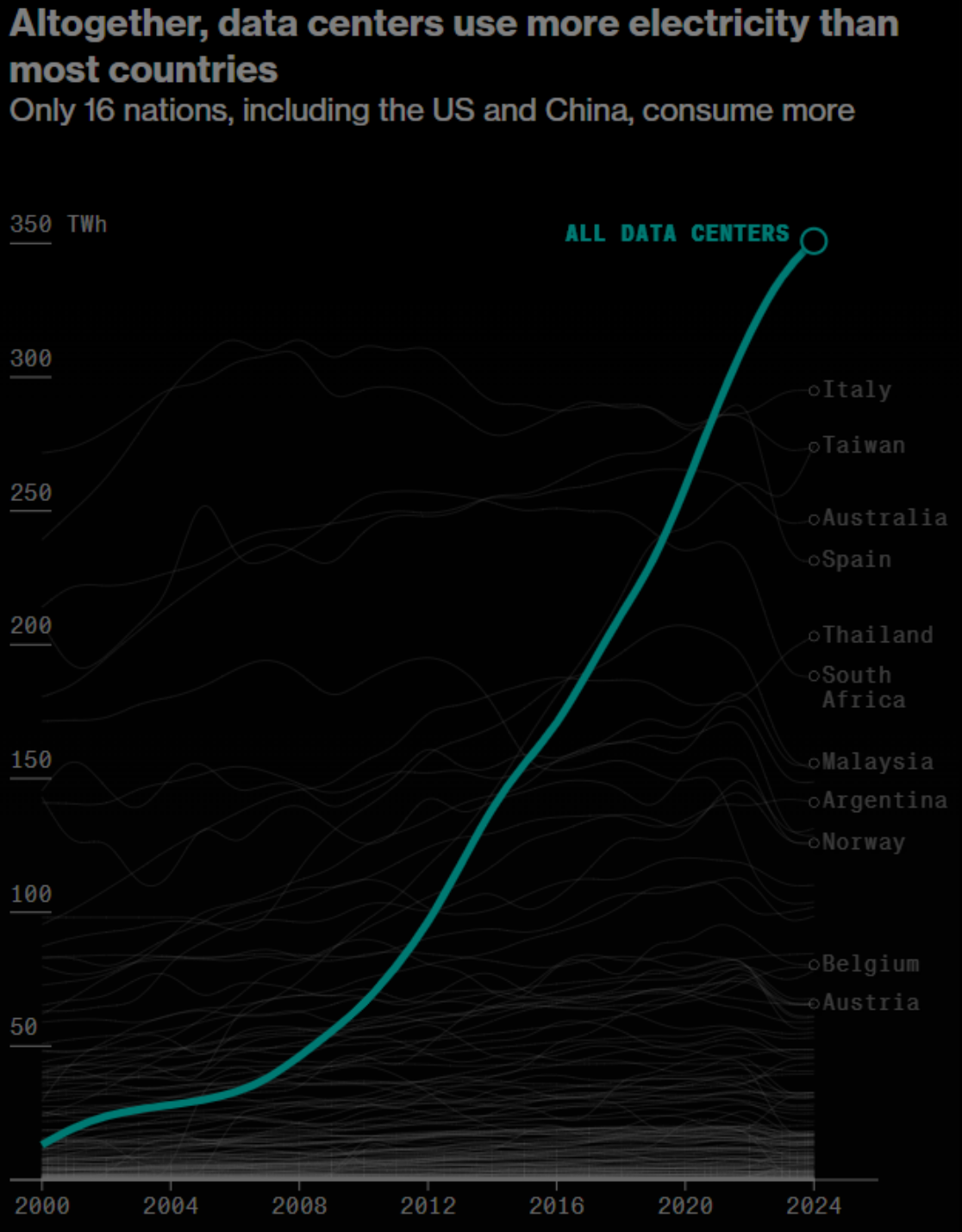
Only 16 nations, including the US and China, consume more

Source: Bloomberg



Data Centres





Potential solutions to energy problem: Cloud AI → Edge AI

- Moving AI applications from the cloud to the edge
- AI model simplification techniques to reduce power consumption
- AI Model quantization, etc.
- On device (Edge) AI model training
- Federated Learning

Edge AI

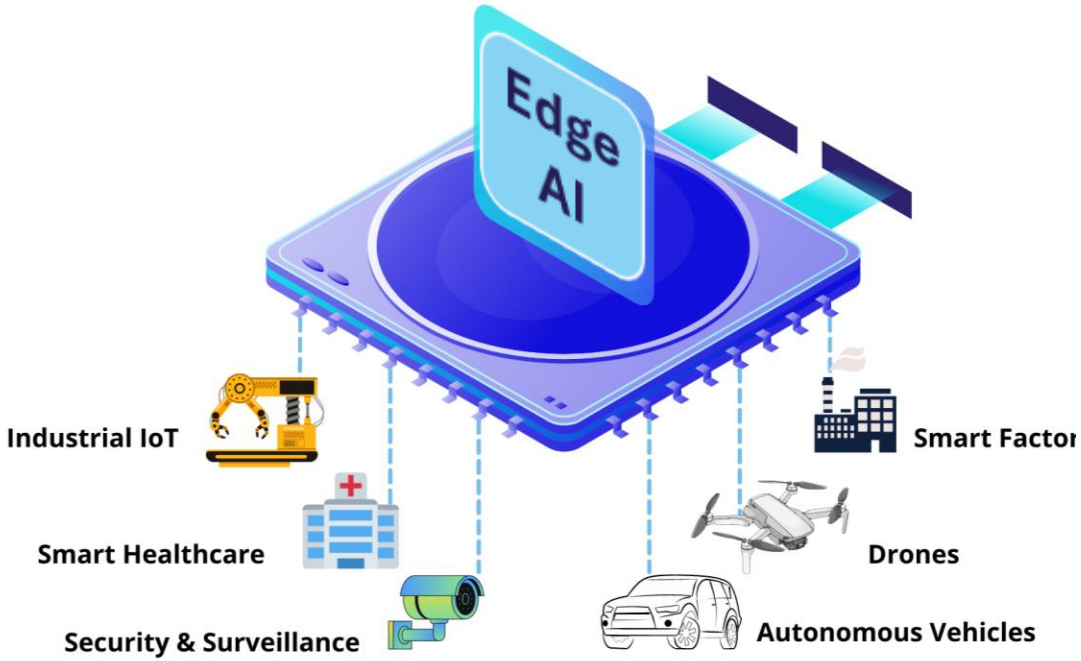
- Unlike Cloud AI (e.g., ChatGPT that runs in data centers), edge AI runs at the edge computing devices such smartphones, cameras, cars, medical devices, ensuring quality of data for inference
- Reduces latency, cost, and power consumption
- Protects data privacy and reduce improve data security and cybersecurity
- Reduces risk of inference failure in critical systems (e.g., autonomous vehicles, healthcare devices) that may endanger lives

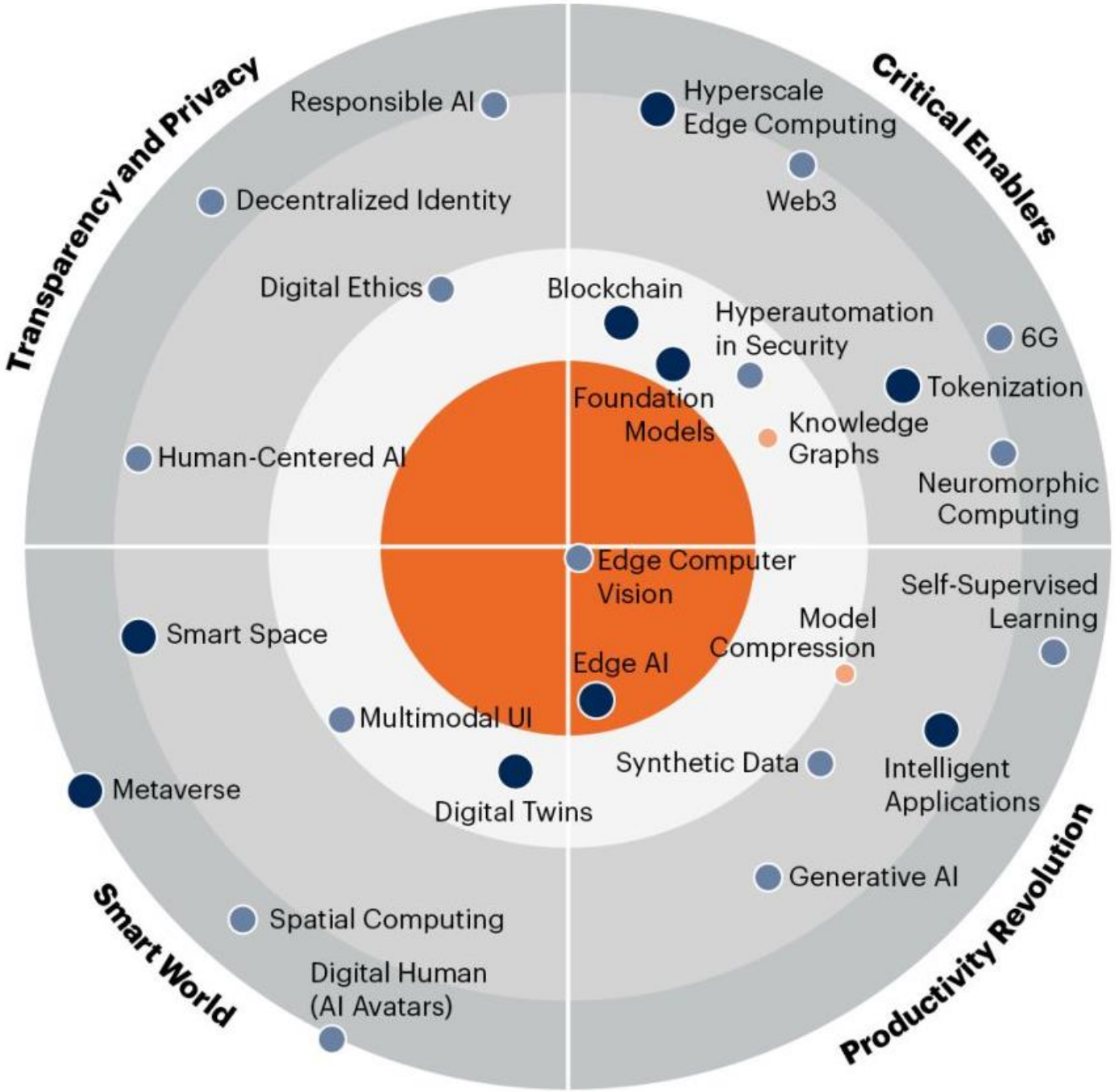
Everyone is a walking AI computer

- **Chip**
- A18 Bionic chip
- 6-core CPU with 2 performance
- 4 efficiency cores
- 5-core GPU
- 16-core Neural Engine
- **Capacity**
- 512GB
- **Multiple Sensors**
- 48MP camera
- Satellite and GPS

The Emergence of Edge AI: a game changer for industries

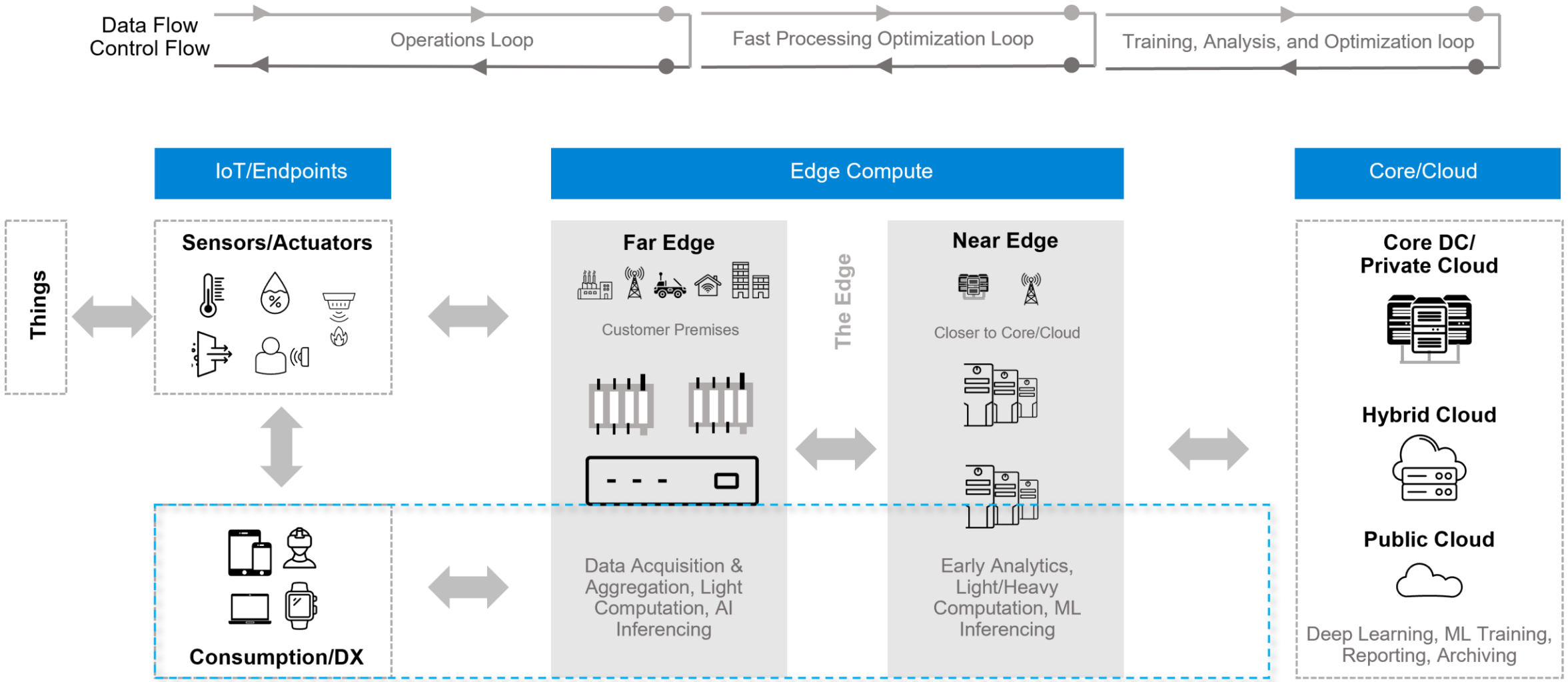
(Gartner 2023)





The Rise of Generative AI at the Edge

image source: Arrow Intelligent Solutions



Small Language Models Could Redefine The AI Race,

Forbes

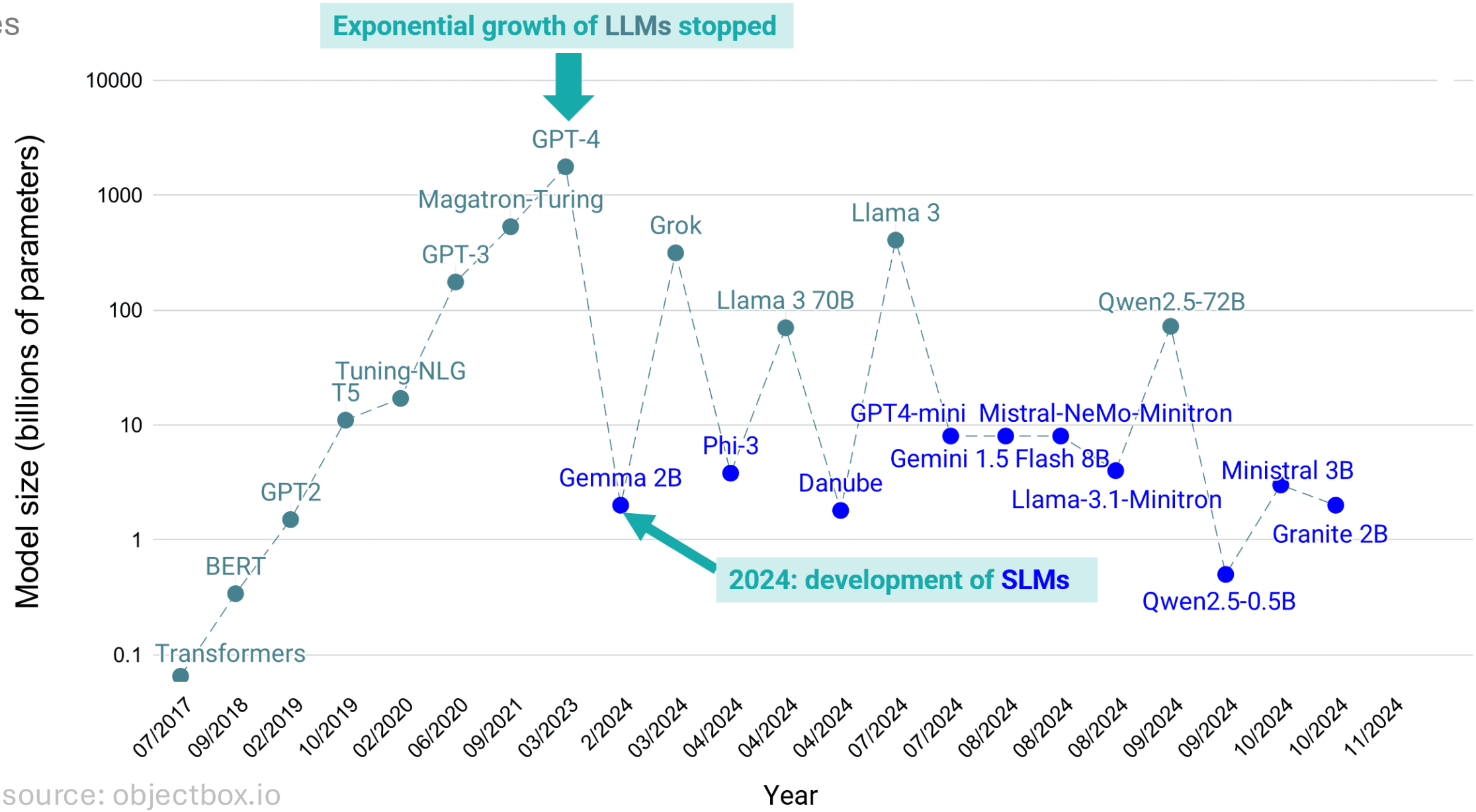
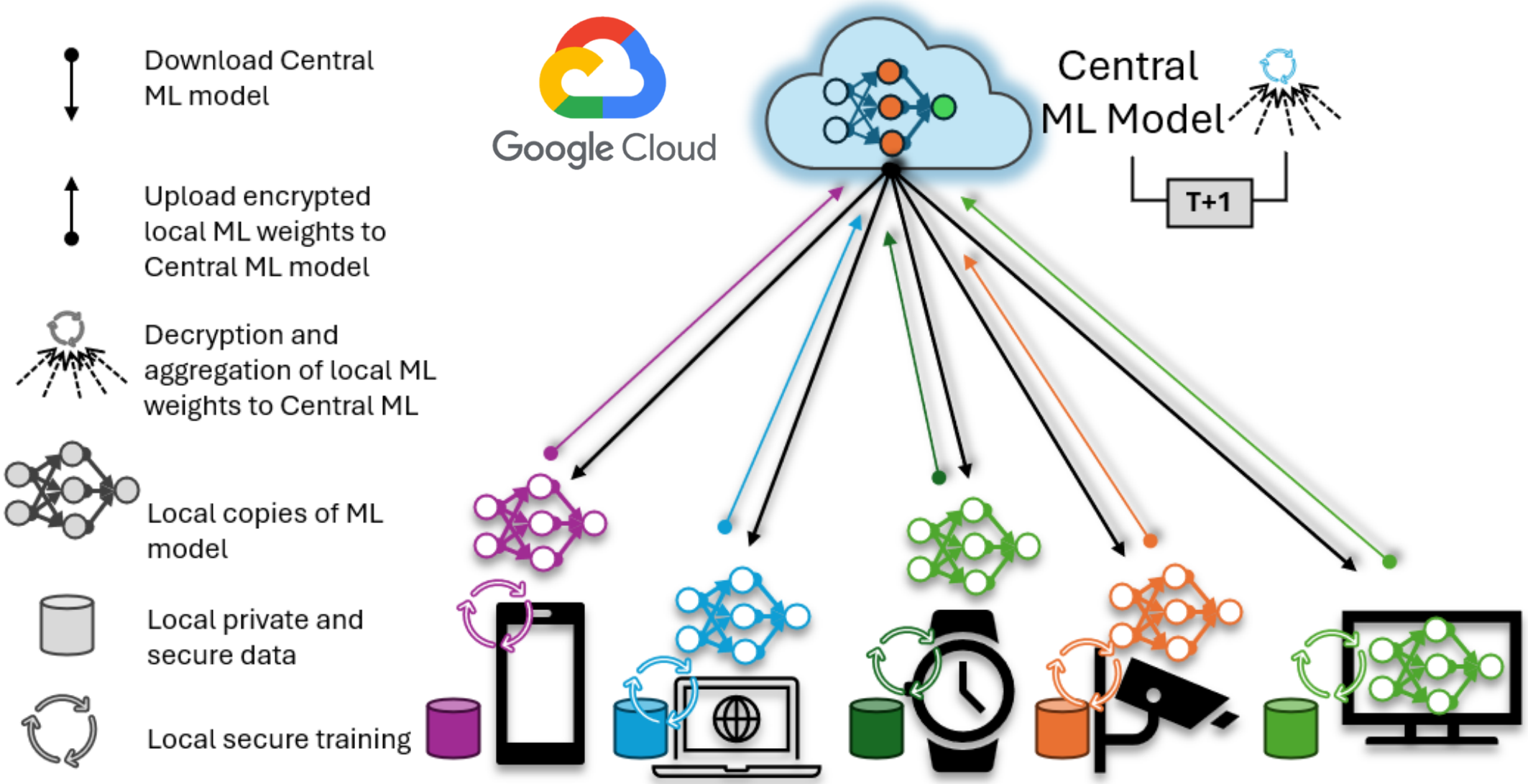


Image source: objectbox.io

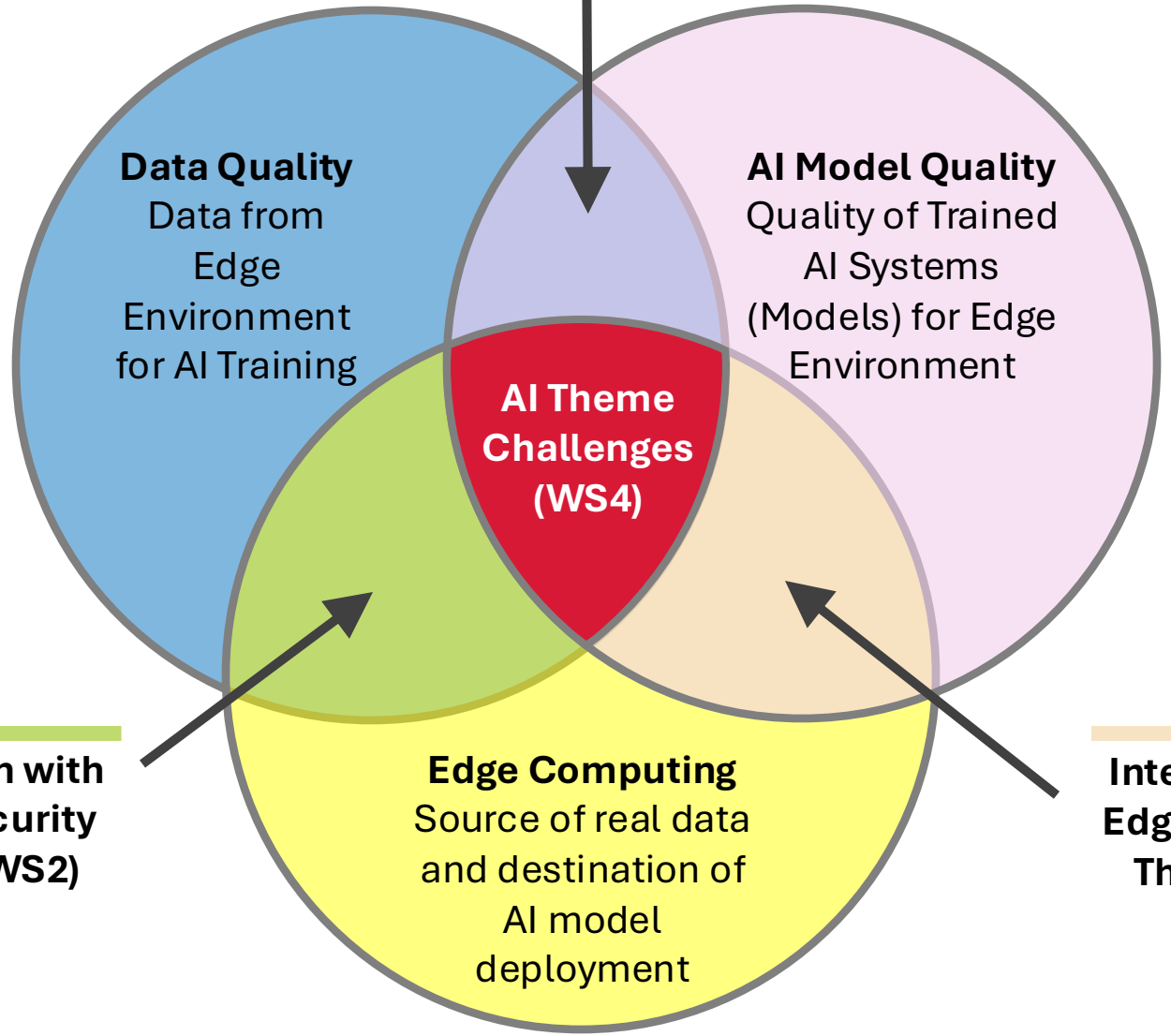
We are the data producer and AI model trainers

Concept of Federated learning



National Edge AI Hub

Interaction with Data Sensitive Applications of AI (WS5)



Interaction with Cyber Security Theme (WS2)

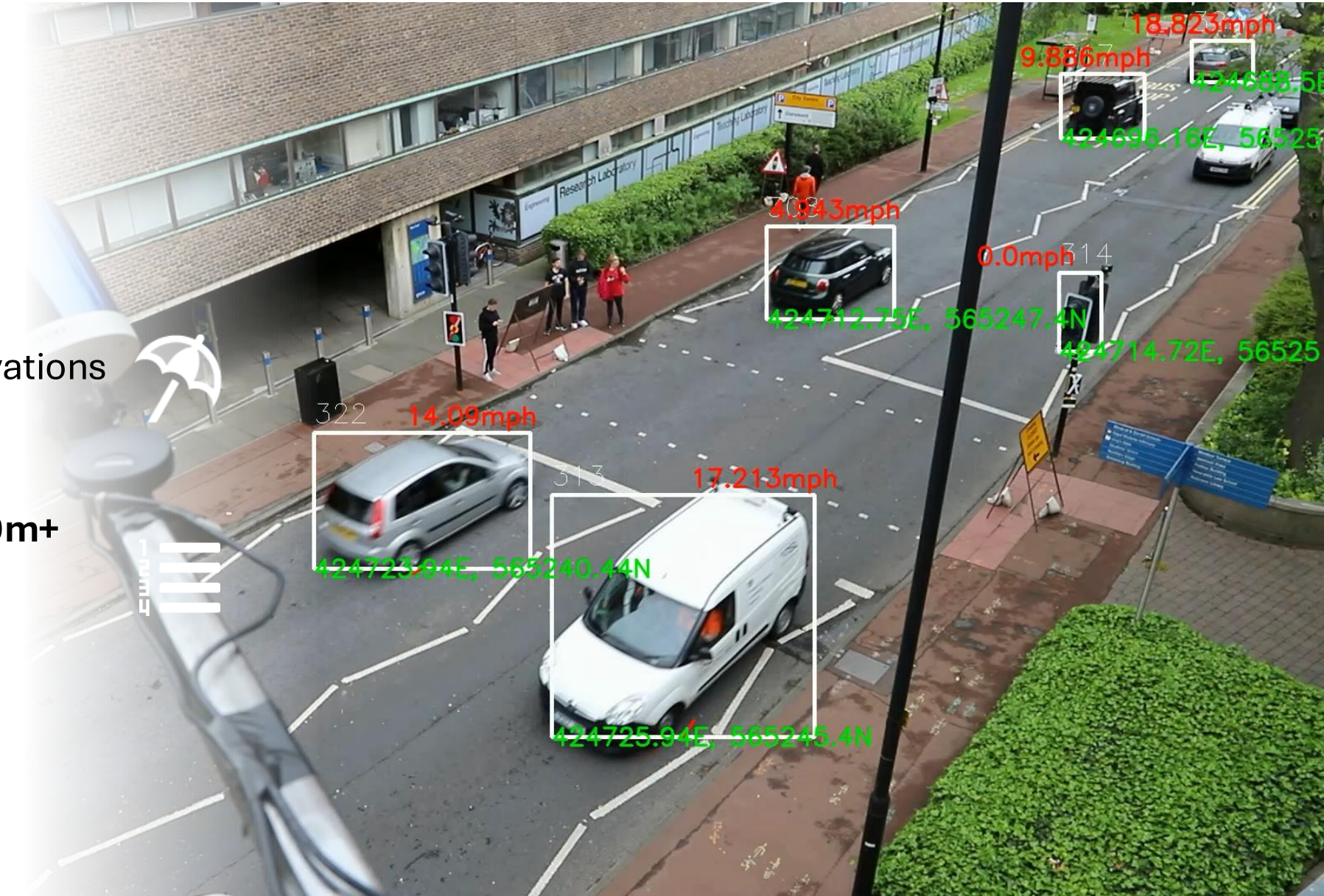
Interaction with Edge Computing Theme (WS3)

Newcastle University's Urban Observatory

>8 million pounds
(Capital investment)

10 billion city observations
10,000 a minute

CCTV: **500** views, **500m+** images, 24 real-time feeds



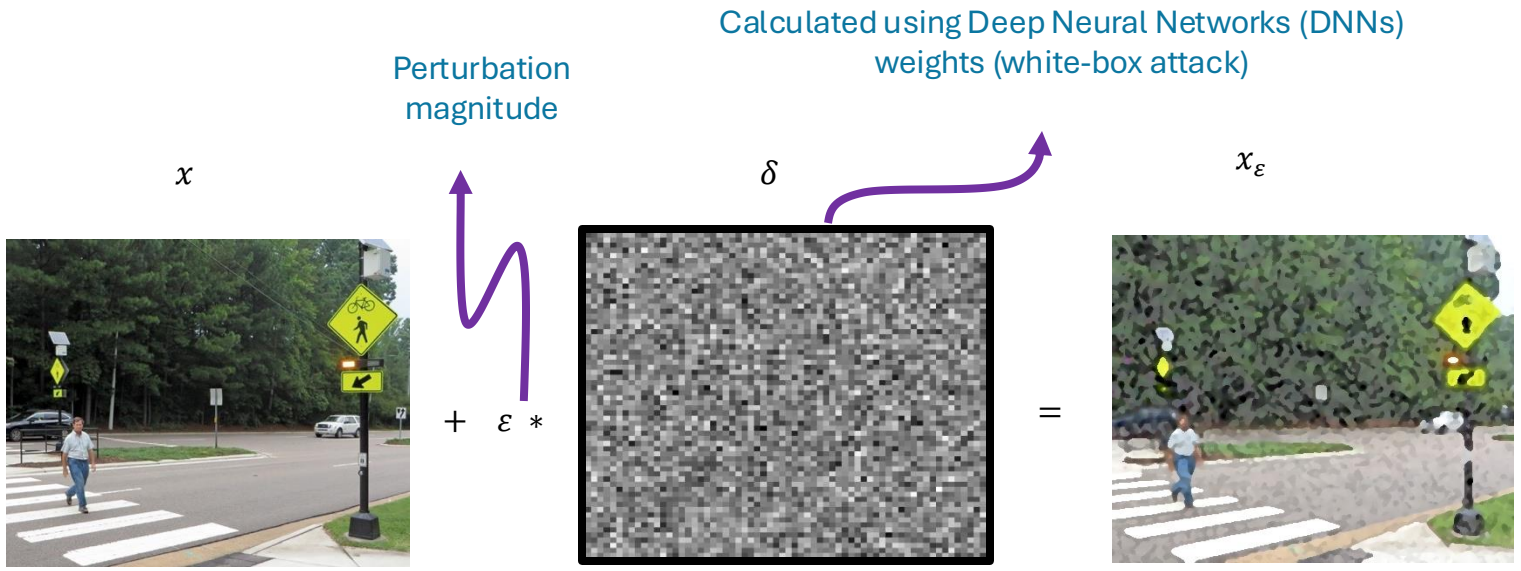
Source: Phil et al (Newcastle)

Edge AI Data Quality Challenges

- **Data quality**
  - degradation of sensors over time
  - data out of range, out distribution, uncertainty
- **Data stream issues**
  - data retrieval - source API failure
  - network failure, network overload
- **Cyber security**
  - adversarial attacks
  - denial of services, spoofing
- **Failure**
  - hardware failure at sensor

Source: Phil et al (Newcastle)

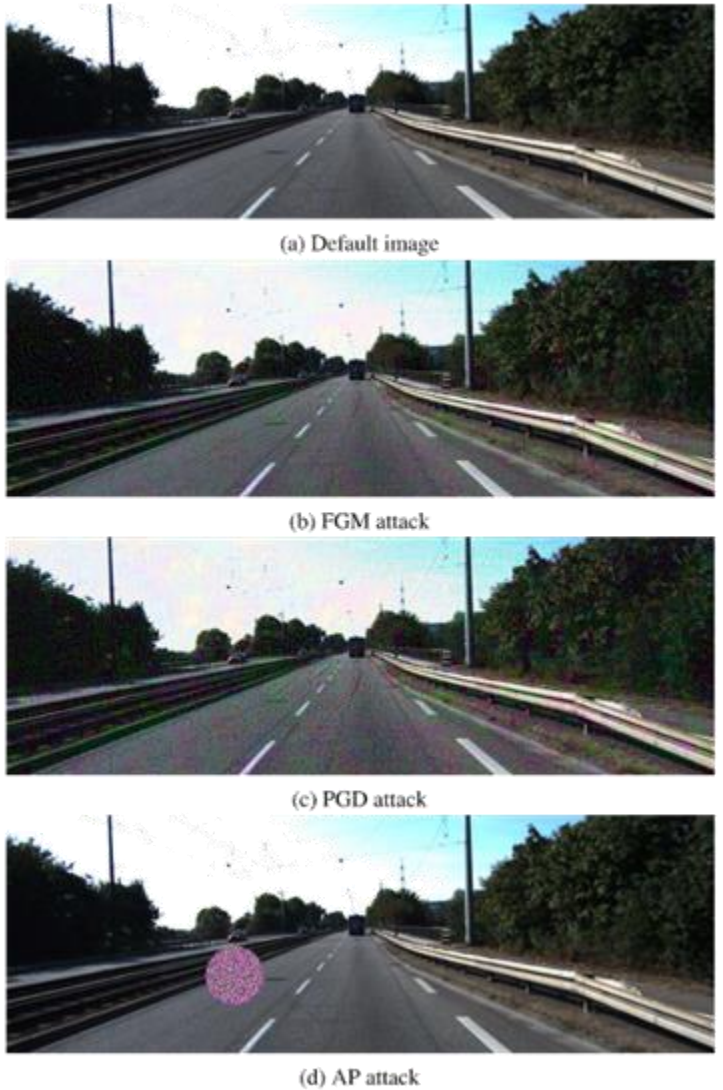
Edge AI Model Quality Challenges



Input example
Predicted as '***Pedestrian***'

Adversarial perturbation
('***Plane***' class)

Adversarial example
Predicted as
'***Speed limit change***'

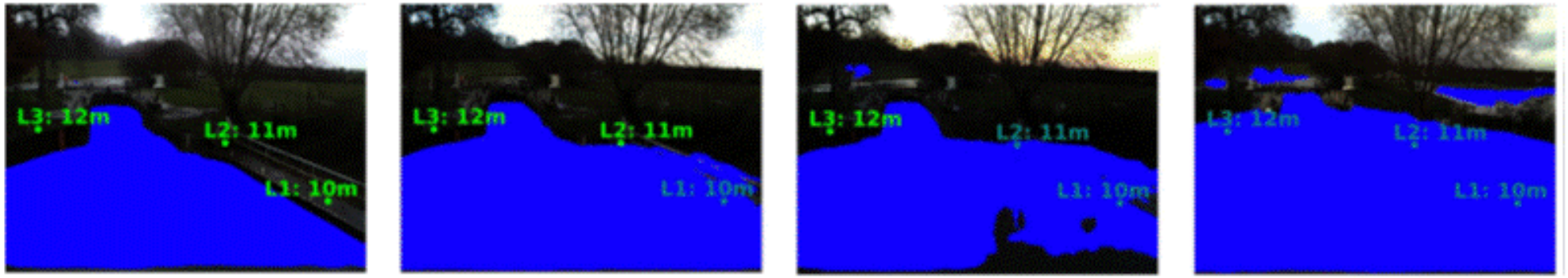


One of objectives of the AI Model Quality analysis is to subject AI model to the '**worst case conditions**' (such as adversarial cyber/attacks) and evaluate the ***ability for a model to remain invariant*** under such settings.

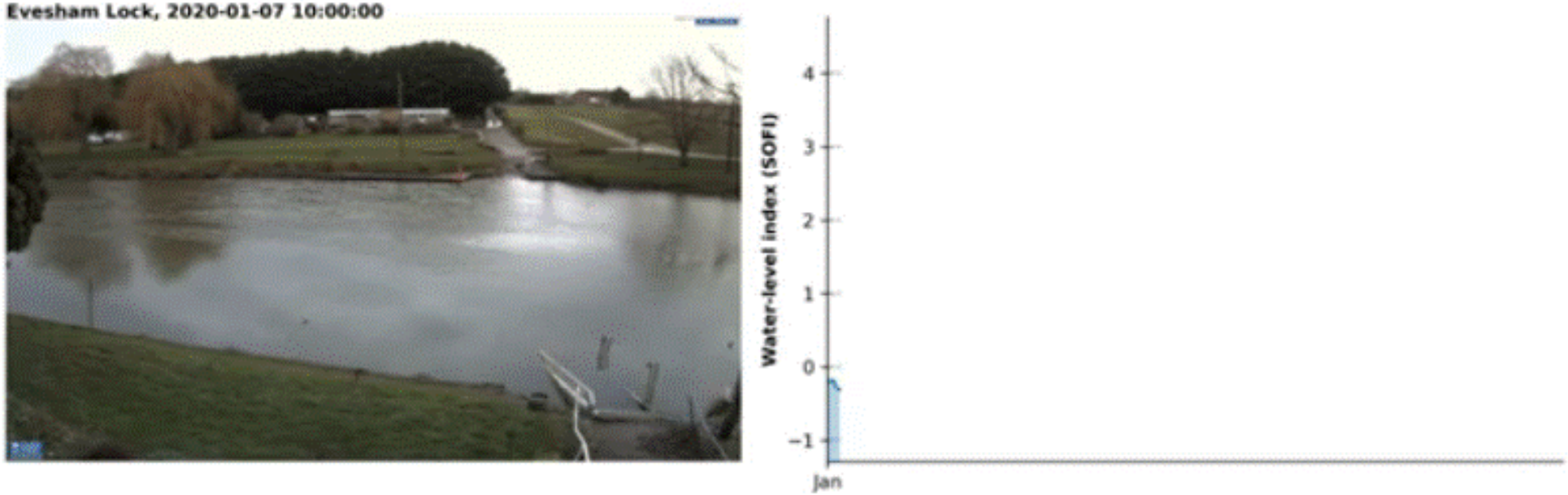
Source: Ojha et al (Newcastle)

Edge AI Powered Environment Monitoring

Example Flood Monitoring



Time-series sequence of images of river. Blue pixels are water segmentation by using deep learning models



Vandaele, Dance, and Ojha, (2021) *Hydrology and Earth System Sciences*

Edge AI Powered Infrastructure Monitoring

Example Structural Engineering

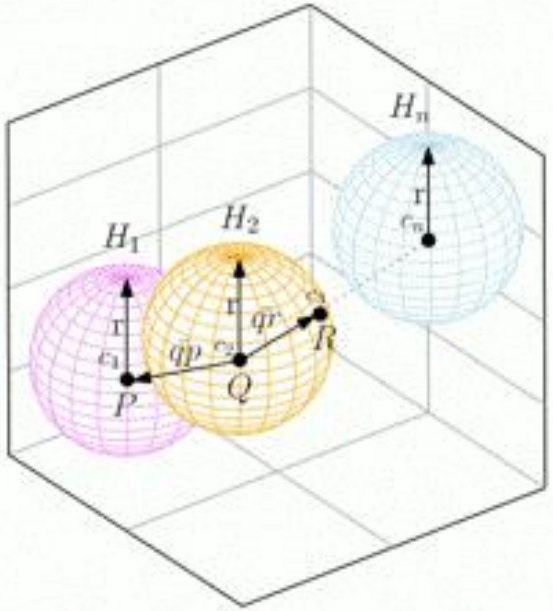


Fig A. Adaptive Hypersphere Search Algorithm for Structural Static Analysis

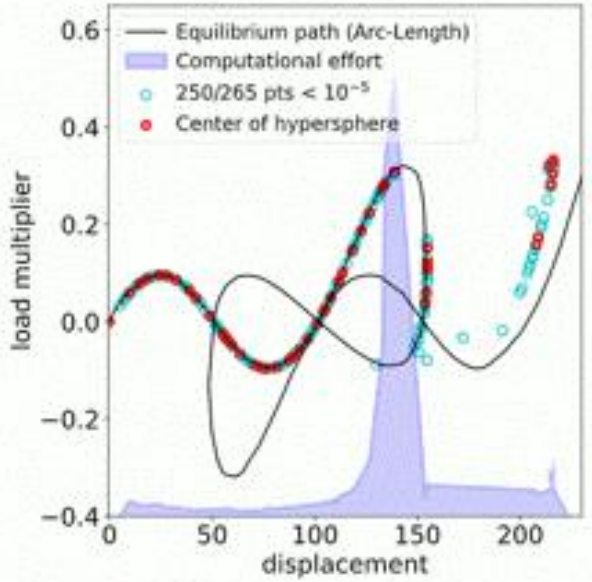
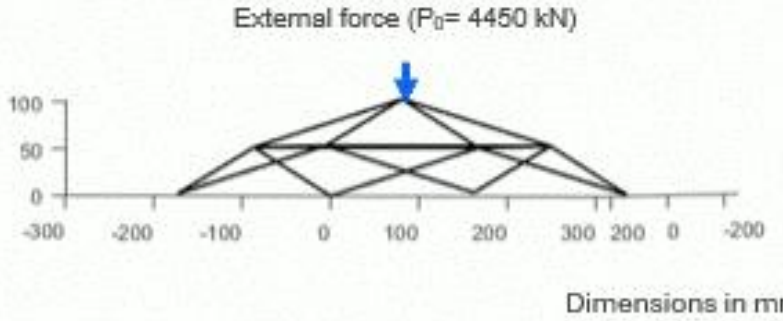


Fig B. Equilibrium Path traced using Adaptive Hypersphere Search Algorithm



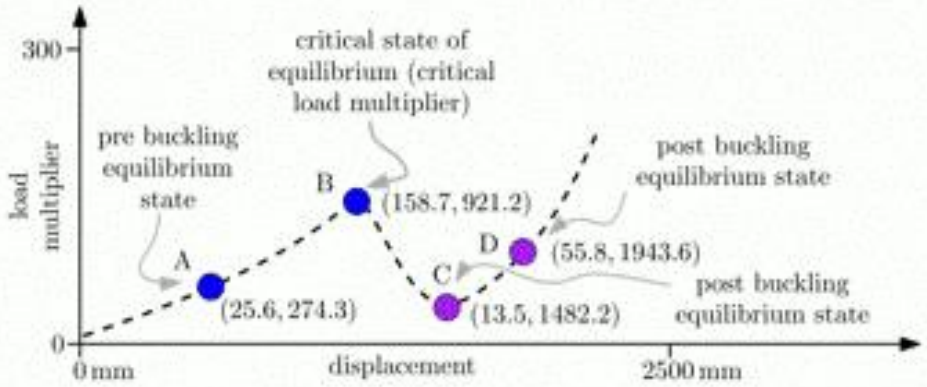
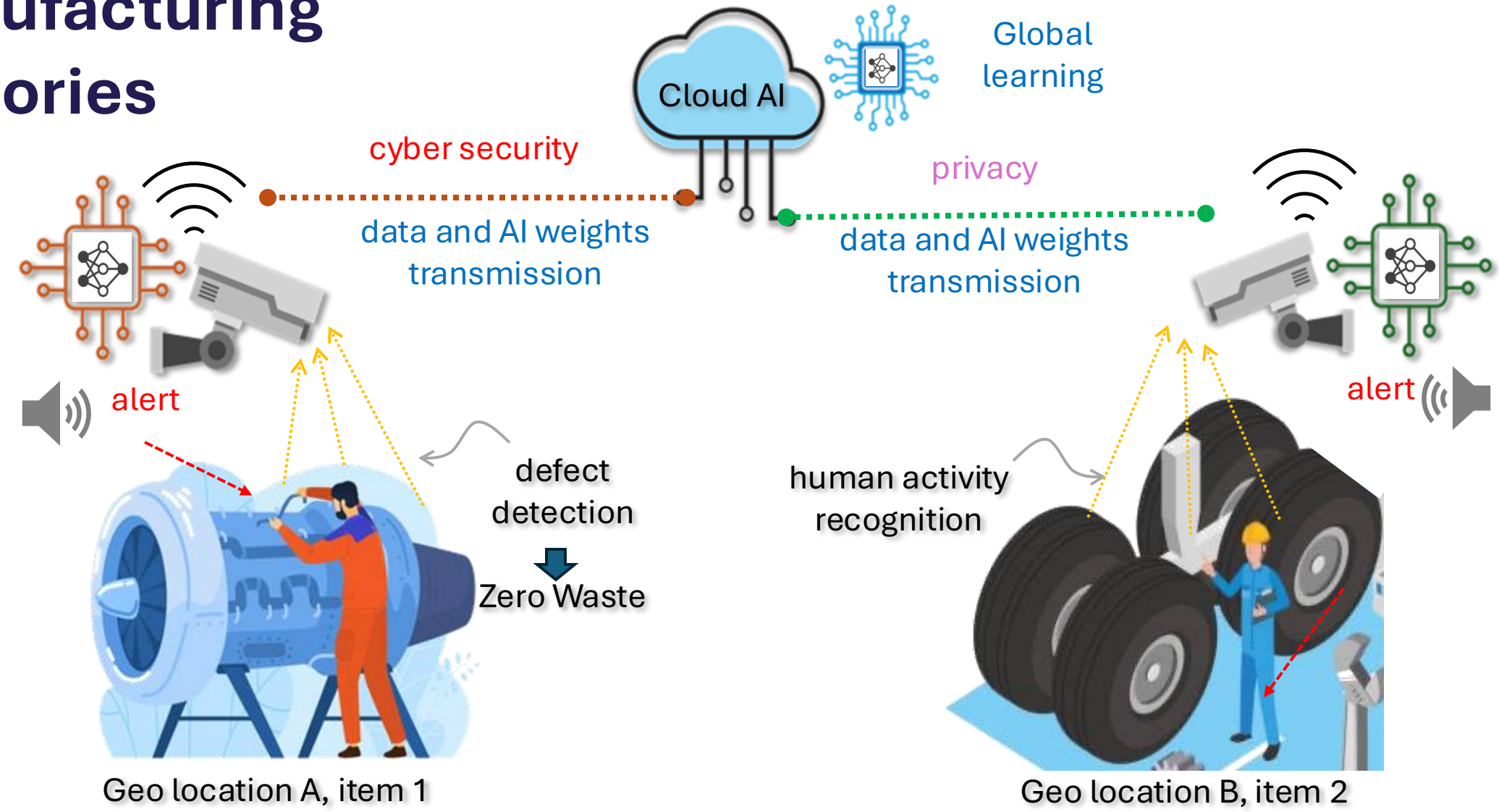
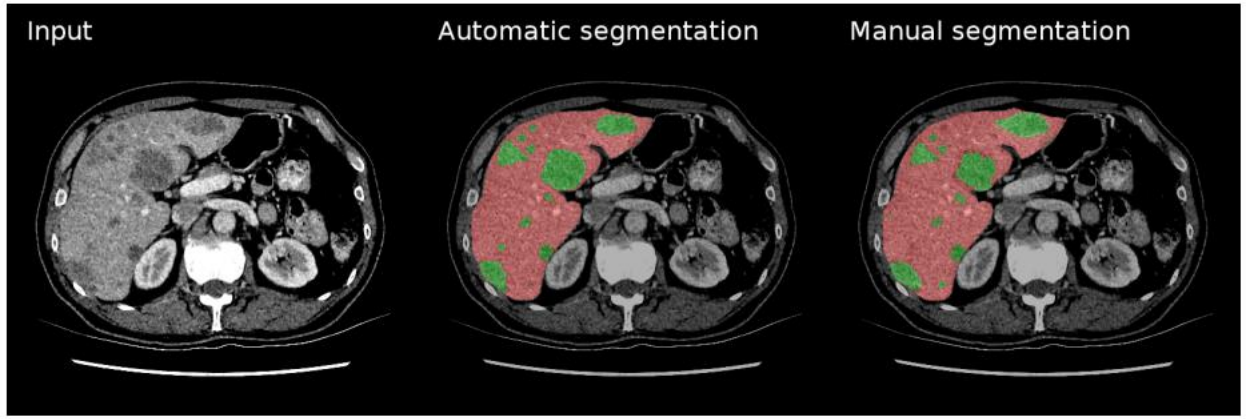


Fig C. Prediction of the Collapse (buckling Analysis)

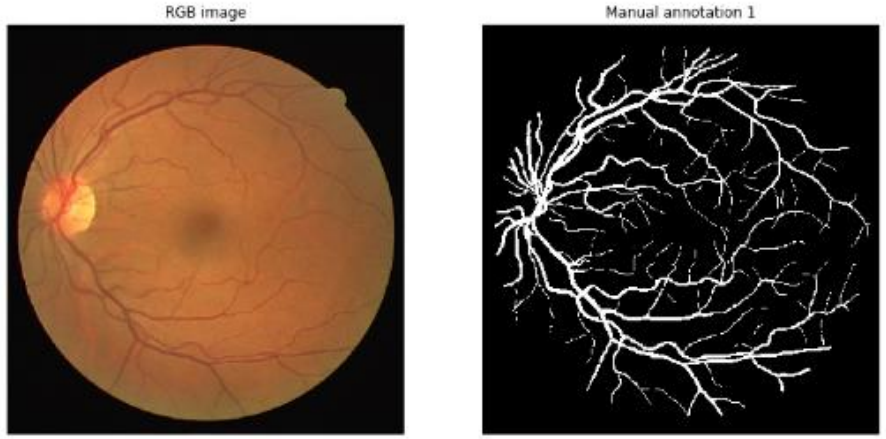
Edge AI Powered Manufacturing Factories



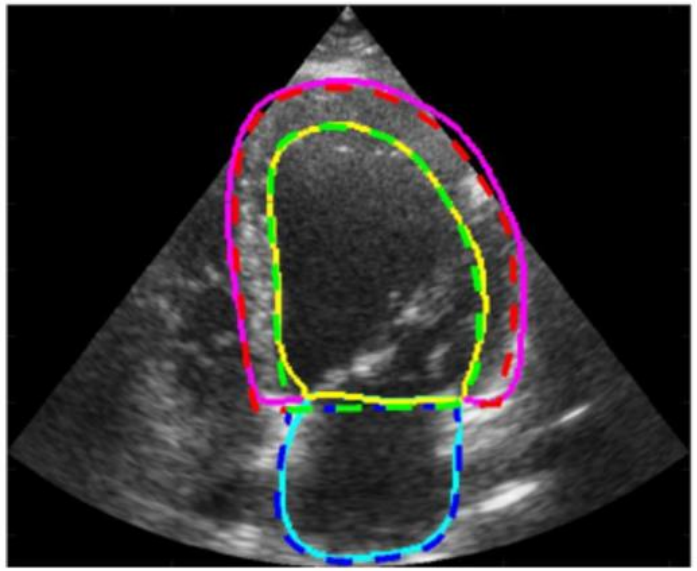
AI Powered Medical Diagnosis

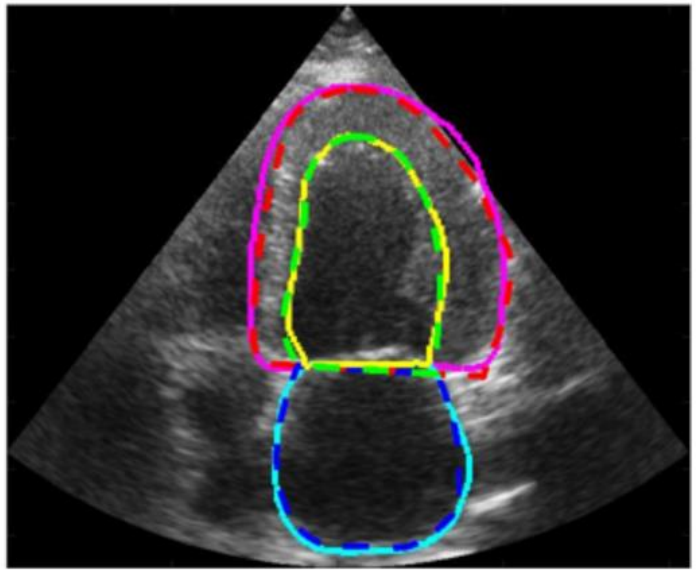


Liver Tumor Segmentation from CT scans

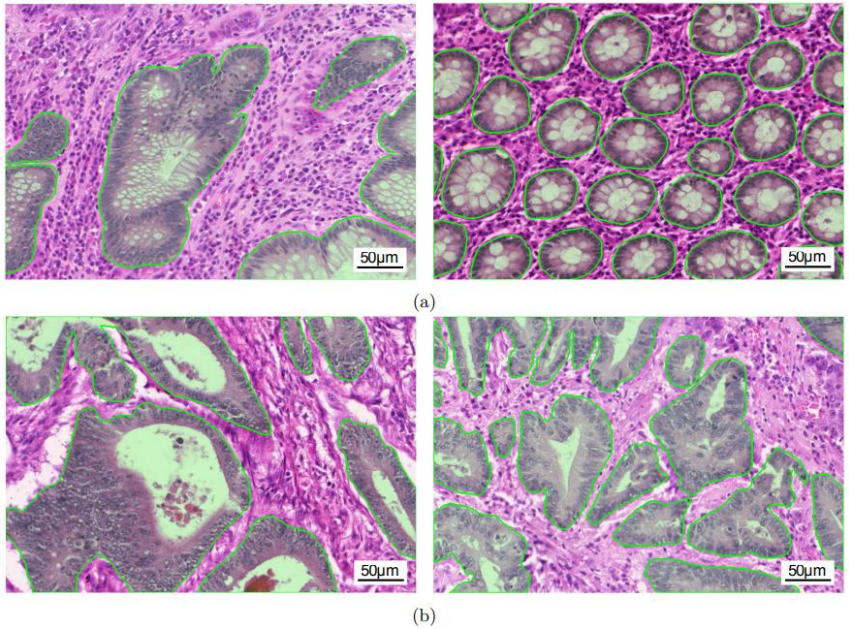


Digital Retinal Images for Vessel Extraction



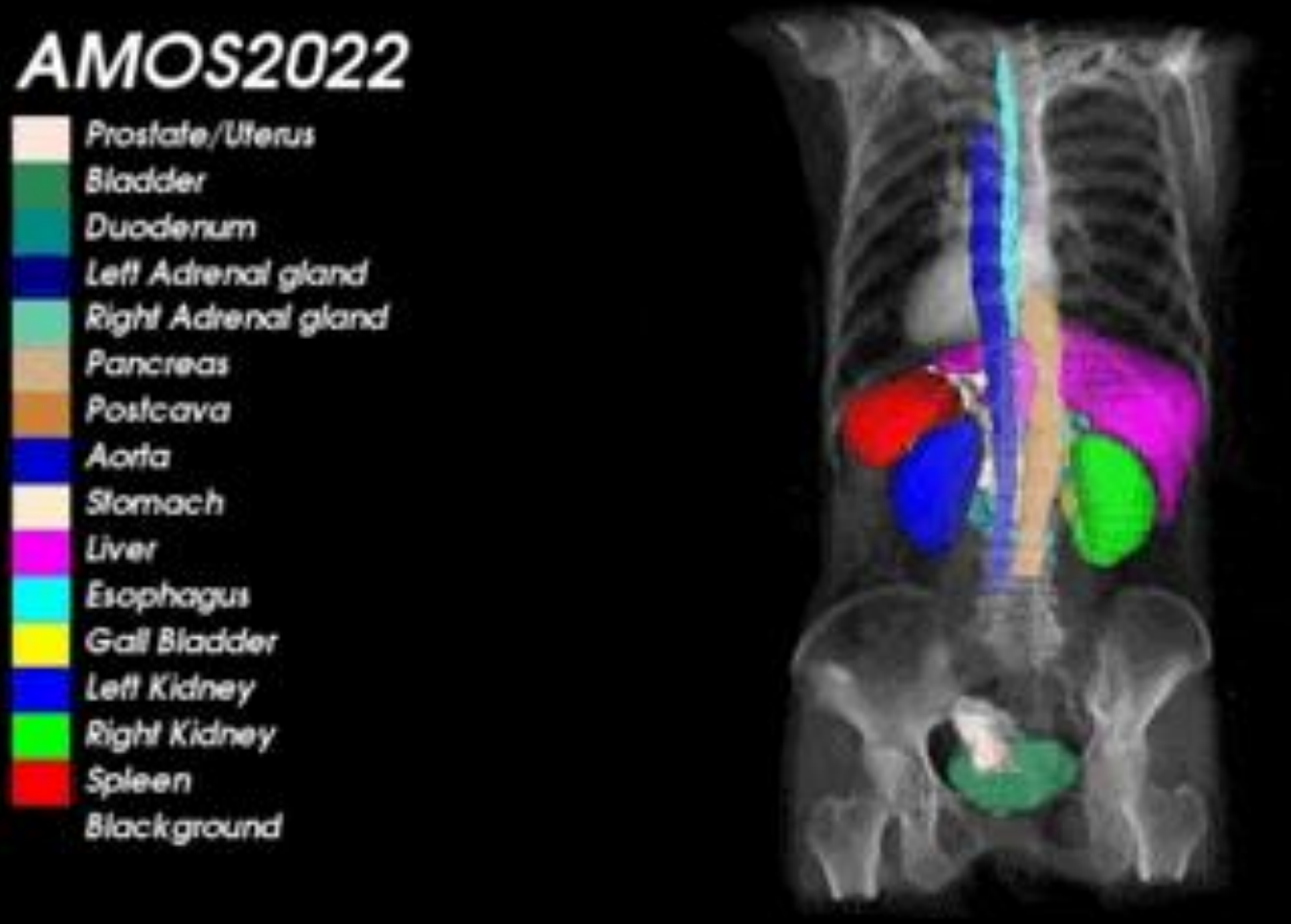


Cardiac Acquisitions for Multi-structure Ultrasound Segmentation



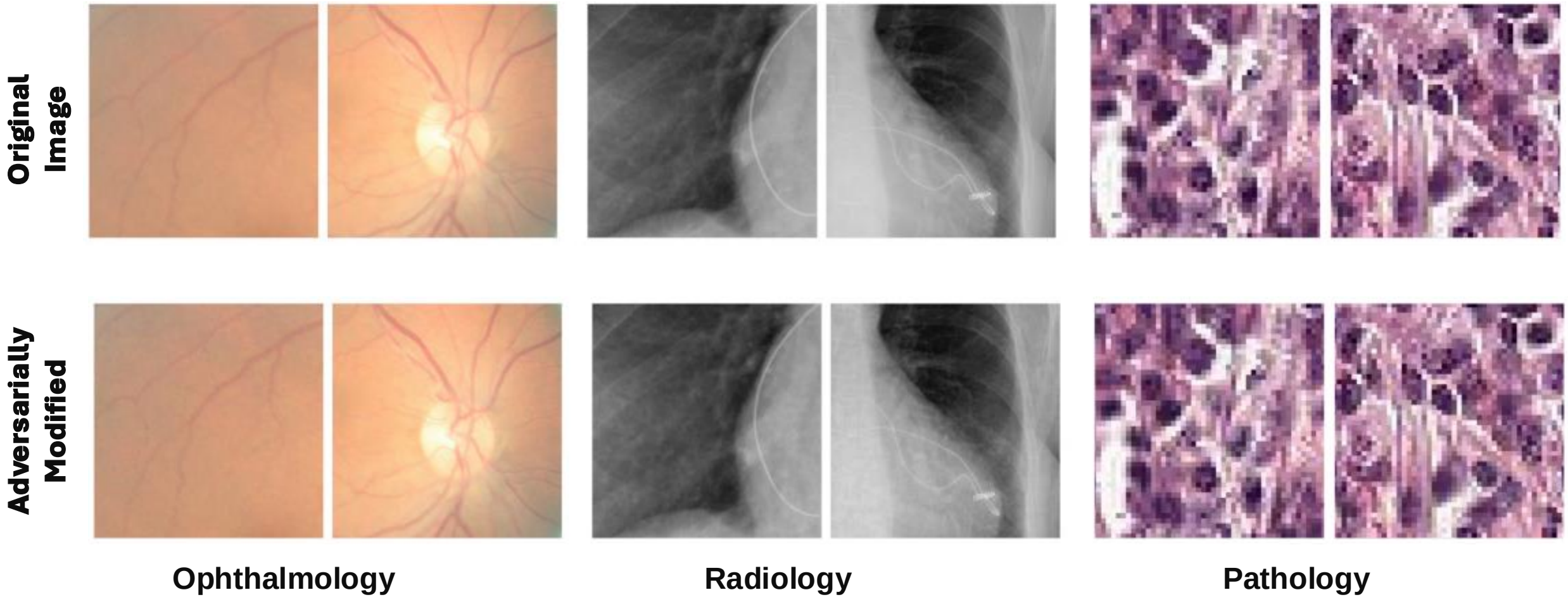
Gland Segmentation in Colon Histology Images

MRI Scans: Multi-Modality Abdominal Multi-Organ Segmentation



https://amos22.grand-challenge.org/

AI Quality (Safety) Concern in Medical Image Analysis

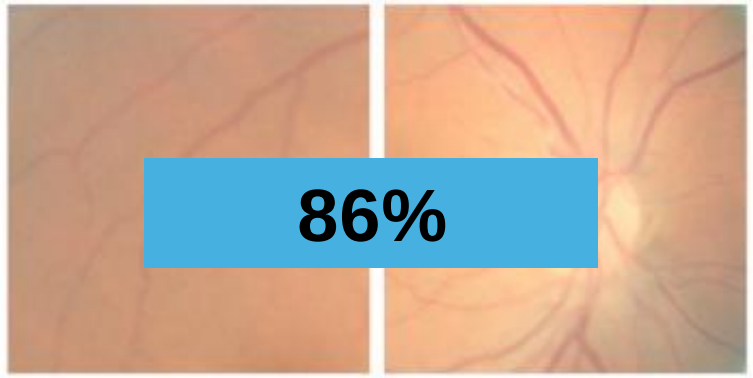


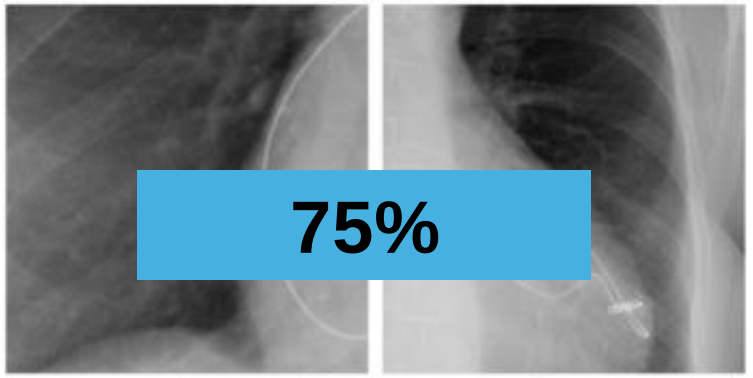
Unperceived changes in images make misclassification

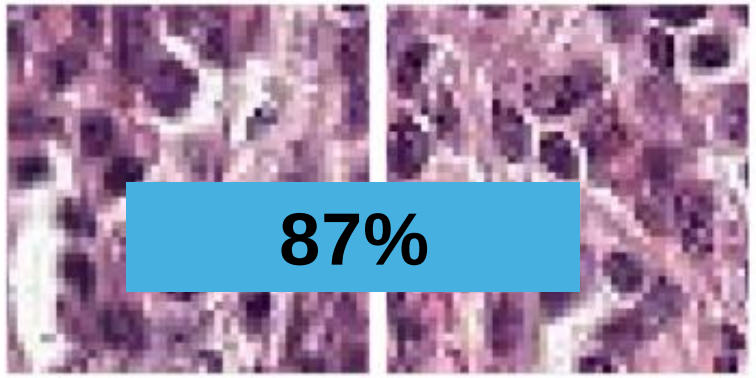
Bortsova et al. (2021). Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. *Medical Image Analysis*,

AI Quality (Safety) Concern in Medical Image Analysis

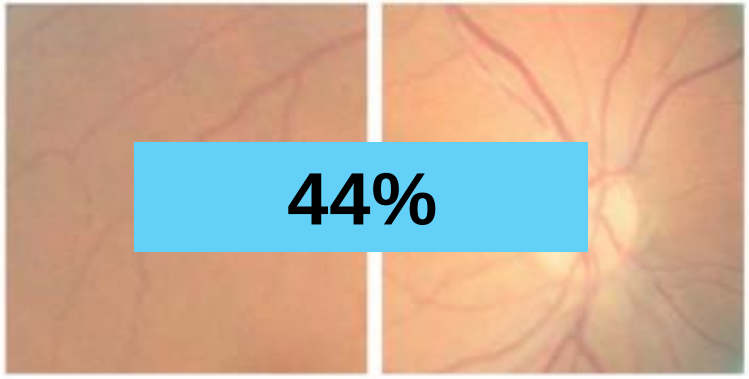
Original Image

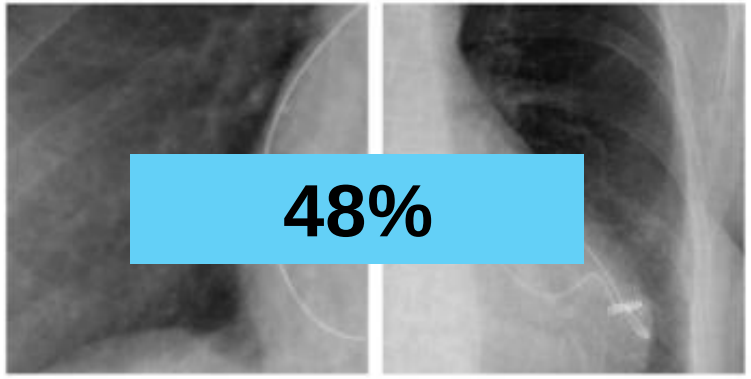


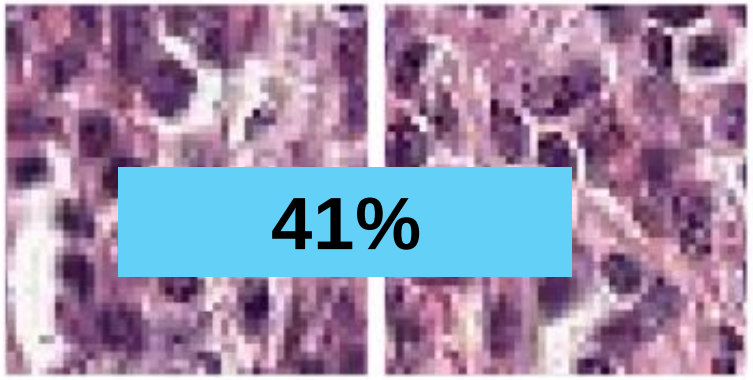




Adversarially Modified







Ophthalmology

Radiology

Pathology

Accuracy of detection decreases even on an unperceived modification

Bortsova et al. (2021). Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. *Medical Image Analysis*,

Edge AI ↔Agentic AI (Next Big Wave of AI)

Edge AI Enables Agentic AI, a complex AI system where AI act autonomously, learn, and adapt to changing situations with minimal human interventions

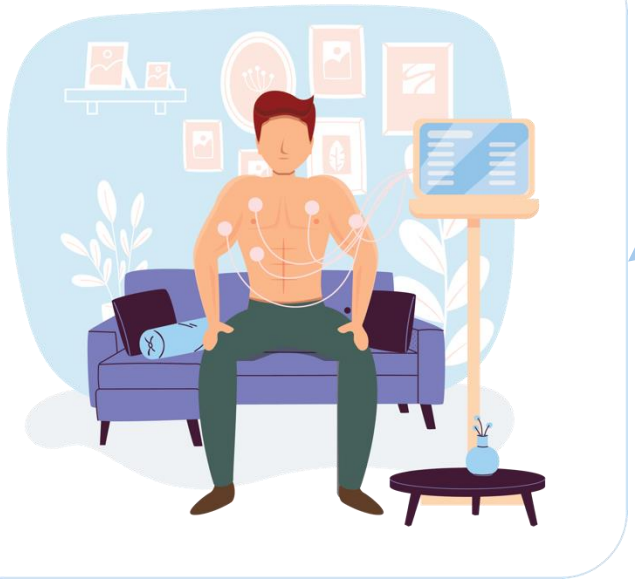
Cloud AI

Secure data communication

Real time processing

Adversarial attacks

Patient's Home

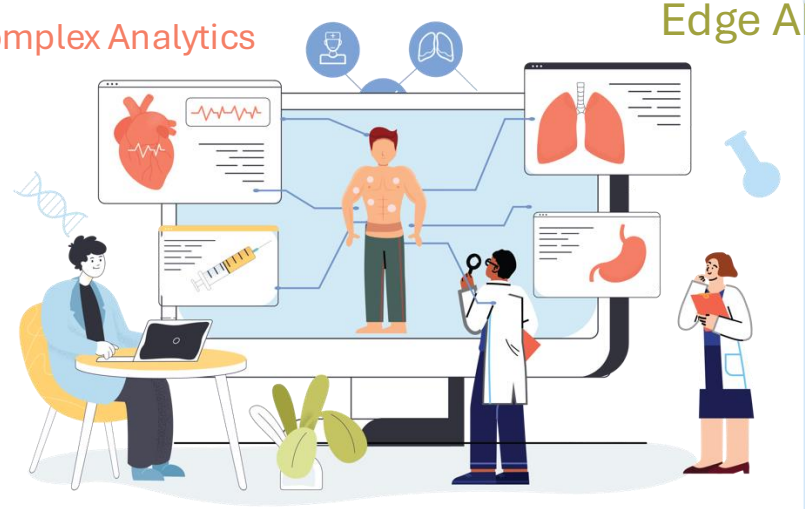


Patient home sensors and health monitoring devices

Patient's AI Models at Clinic

Complex Analytics

Edge AI



Hospital / Clinic remote monitoring, data processing, and AI powered data analytics and prescription

Newcastle University

Computing

Security, Cyberattacks mitigation, Model accuracy

Medical

Reliability, Explainability, Transparency

Social Sciences

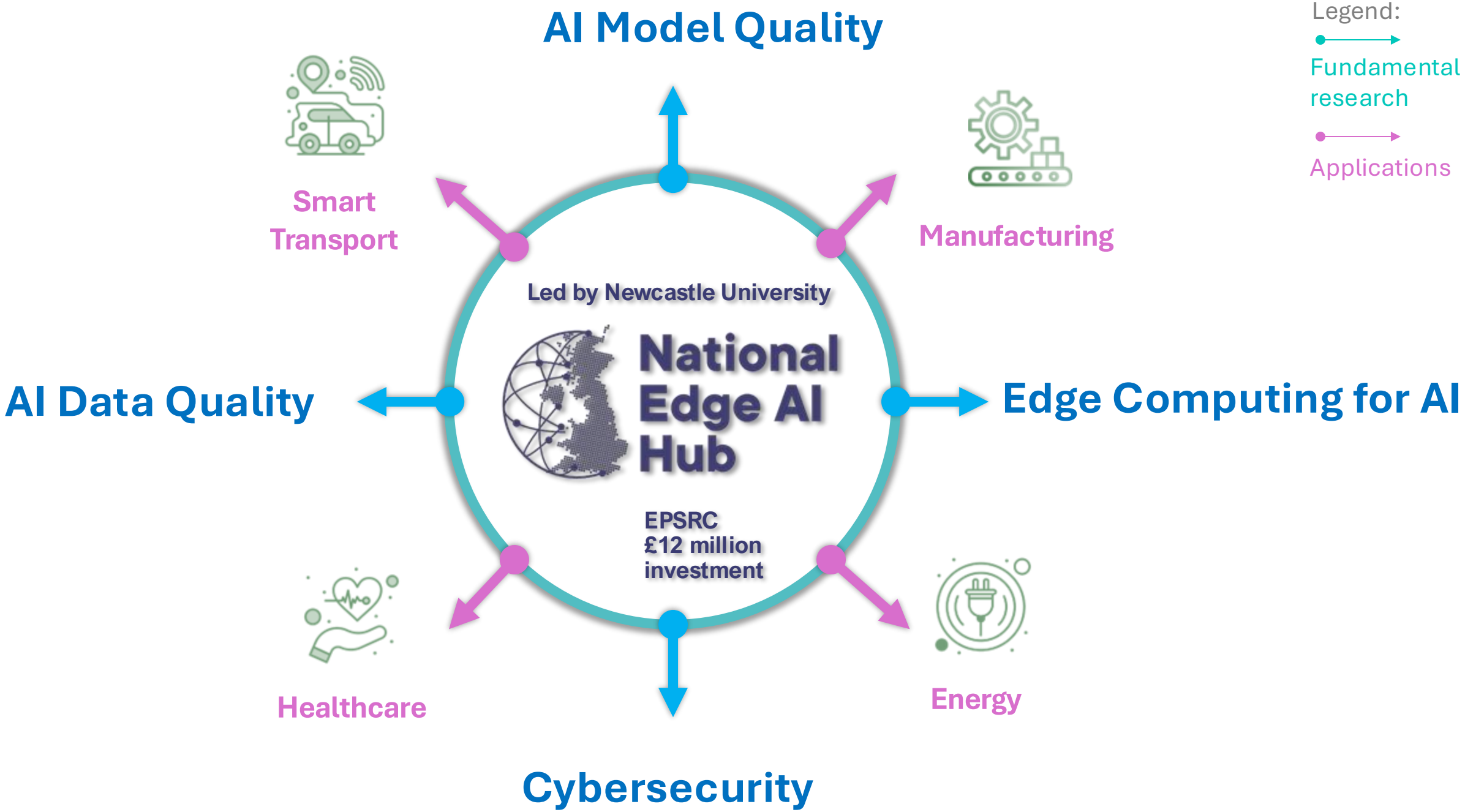
Privacy, Law, Ethics, Business

UK Research and Innovation

Engineering and Physical Sciences Research Council

AISI | AI SAFETY INSTITUTE

National AI Strategy


AI Model Quality
Legend:
Fundamental research
Applications
Smart Transport
Manufacturing
Led by Newcastle University
National Edge AI Hub
EPSRC
£12 million
investment
AI Data Quality
Edge Computing for AI
Healthcare
Energy
Cybersecurity

National Edge AI Hub

Study PhD with us

https://edgeaihub.co.uk/phd-in-edge-ai/

Pathways to PhD. Send us you research proposal and plan; we will work with you refining your PhD agenda and find suitable supervisor or institution for your PhD enrollment. We can help explore PhD funding sources if you do not have your own Govt. or self funding provisions

Work with us

https://edgeaihub.co.uk/

Pathways to Industry engagement. We have several pathways for engagement, for example, pump-priming funding, 50-50 matching funding, and other. We will work with you and explore external industrial-academic partnership funding sources with you.

Contact

varun.ojha@newcastle.ac.uk

ojhavk.github.io