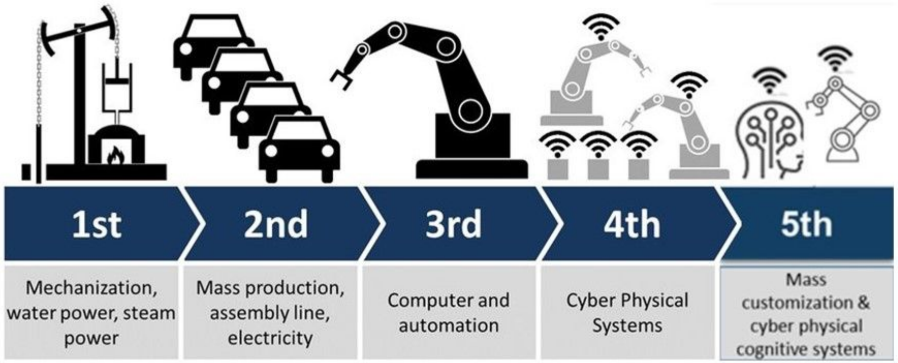
Industry 5.0, Edge Computing (Jetson Nano)

Marcelo Pias, FURG-Brazil
mpias@furg.br

Evolution: “industry of generations”



Factfile

History of industrial revolution

1.0 — **1780 – Mechanisation**
Industrial production based on machines powered by water and steam

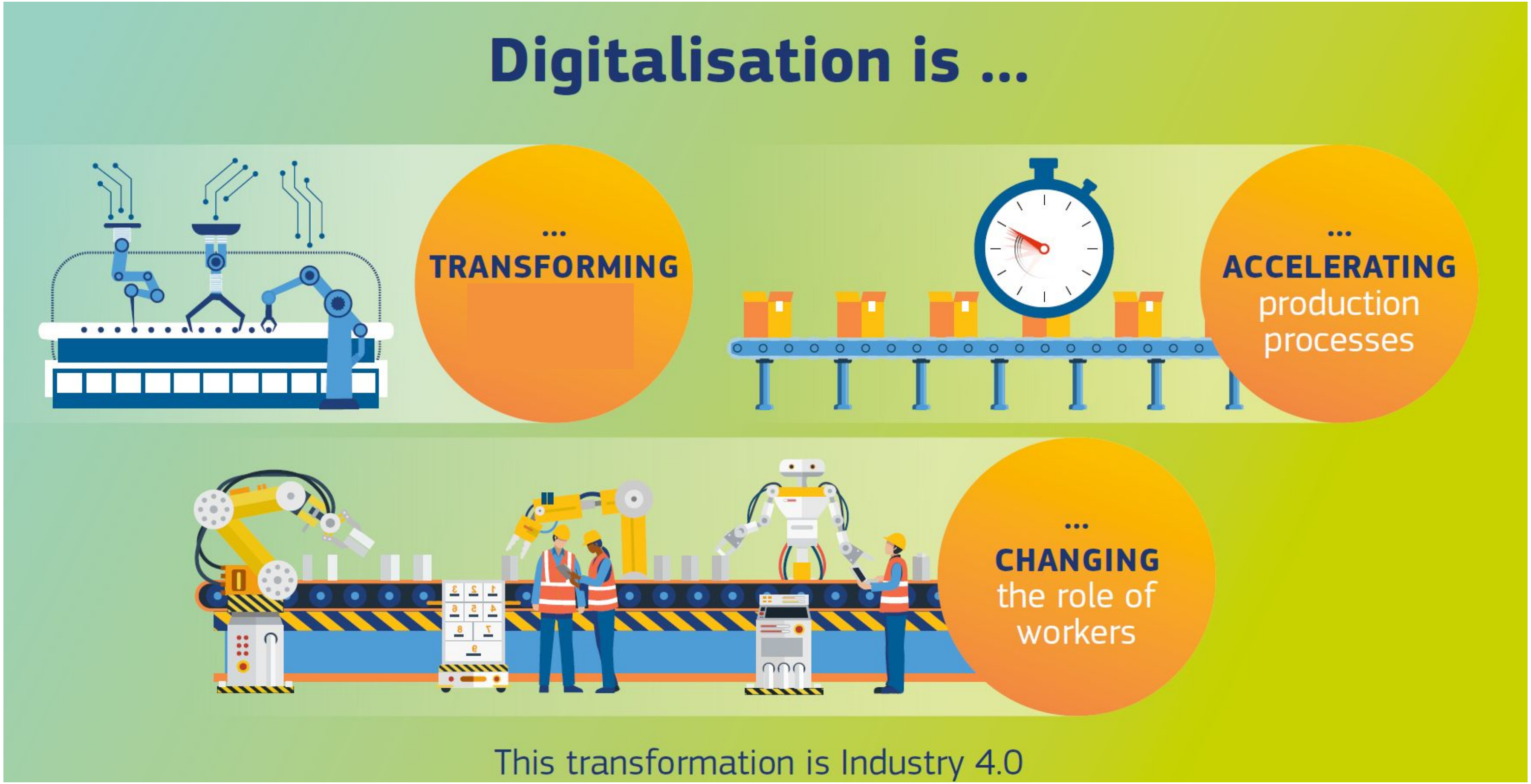
2.0 — **1870 – Electrification**
Mass-production using assembly lines

3.0 — **1970 – Automation**
Automation using electronics and computers

3.5 — **1980 – Globalisation**
Offshoring of production to low-cost economies

4.0 — **Today – Digitalisation**
Introduction of connected devices, data analytics and artificial intelligence technologies to automate processes further

5.0 — **Future – Personalisation**
The fifth industrial revolution, or Industry 5.0, will be focused on the co-operation between man and machine, as human intelligence works in harmony with cognitive computing. By putting humans back into industrial production with collaborative robots, workers will be upskilled to provide value-added tasks in production, leading to mass customisation and personalisation for customers



European Commission, Directorate-General for Research and Innovation, Industry 5.0 : human-centric, sustainable and resilient.


Industry 5.0 ...
HUMAN-
CENTRIC
... promotes talents, diversity
and empowerment
RESILIENT
... is agile and resilient
with flexible and adaptable technologies
SUSTAINABLE
... leads action on sustainability
and respects planetary boundaries

Enablers

INDUSTRY 5.0 recognises the power of industry to achieve societal goals beyond jobs and growth to become a RESILIENT PROVIDER OF PROSPERITY, by making production RESPECT THE BOUNDARIES OF OUR PLANET and placing the WELLBEING OF THE INDUSTRY WORKER at the centre of the production process.

Link humans with technologies
- Multilingual speech, human intention prediction
- Tracking technologies for mental and physical strain and stress of employees
- Collaborative robots ('cobots'), which work together with humans

Augmented, virtual or mixed reality technologies, esp.for training and inclusiveness

Individualised

Causality-based and not only correlation-based artificial intelligence
Brain-machine interfaces
Individual, person-centric Artificial Intelligence
Informed deep learning (expert knowledge combined with Artificial Intelligence)
Skill matching of humans
Secure and energy-efficient Artificial Intelligence

Artificial Intelligence

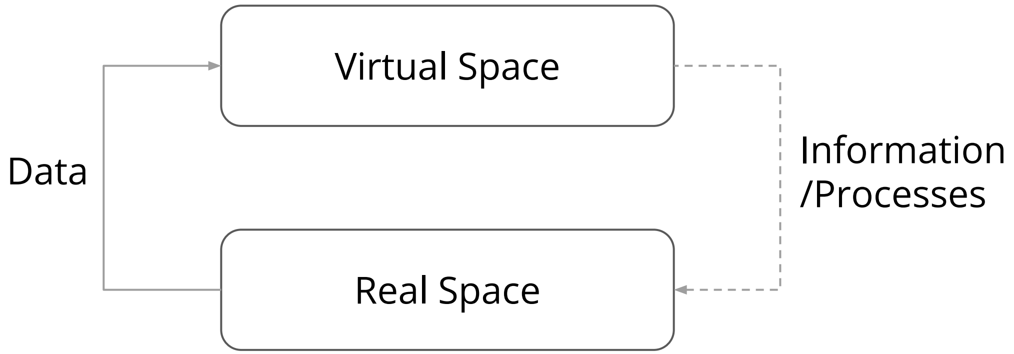
Digital twins and simulation of products and processes
Multi-scale dynamic modelling and simulation
Simulation and measurement of environmental and social impact
Cyber-physical systems and digital twins of entire systems
Planned maintenance

Digital Twins

QUALITY CONTROL AND PRODUCT TRACEABILITY OF AN AUTOMATED PRODUCTION CELL

Project-based experience

Digital Twin: Definition (Industry 5.0)



- Grieves e Vickers (life cycle of products)
- Virtual and physical entities (Twins) are connected through **data** and **processes**
- The goal is improving the performance of the physical entity through **computational techniques** applied on the virtual Twin

Digital Twins and Industry 4.0/5.0

- Development of smart products and services
  - Embedded systems
  - IoT (Internet of Things)
  - CPS (Cyber-Physical Systems)

- The use of Digital Twins in the life cycle of automotive products is still little explored in Brazil

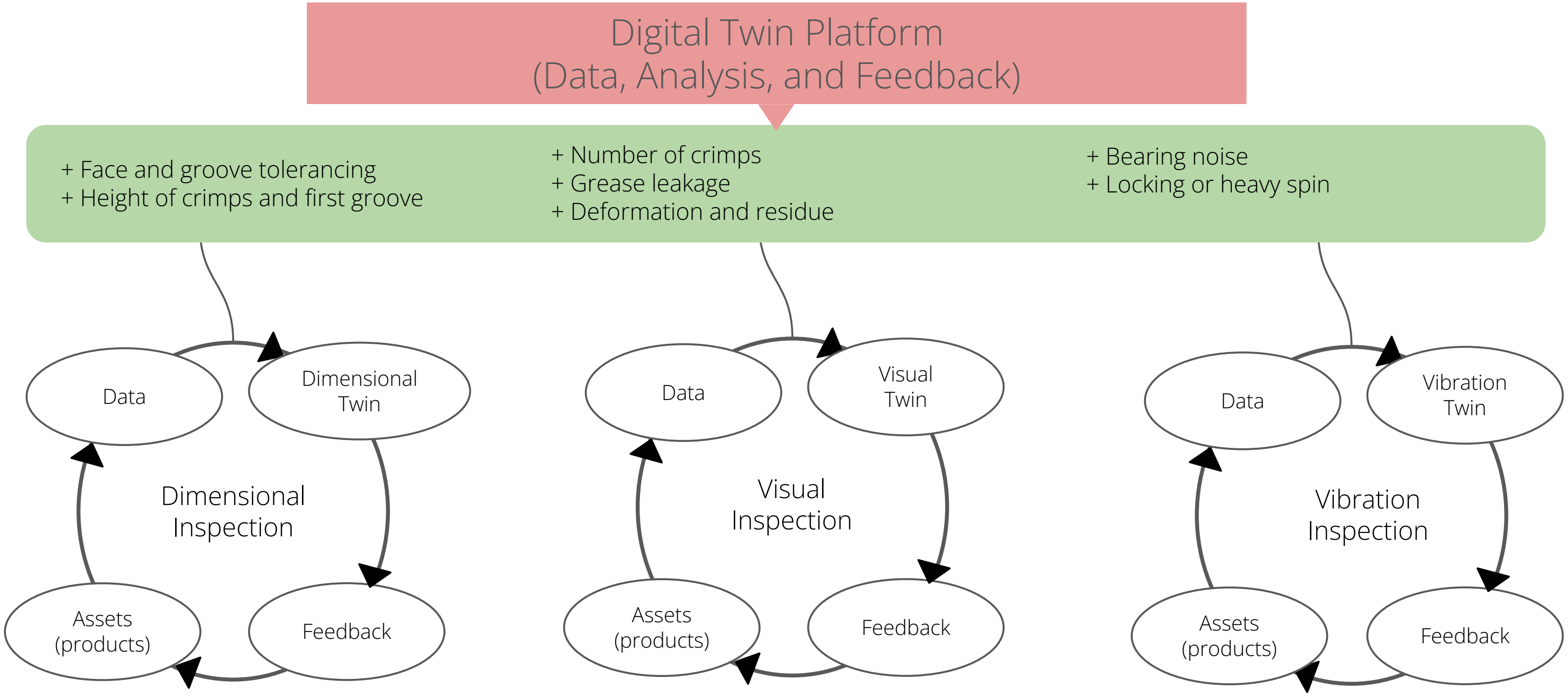
General View of the Project

- Subject: Production cell of parts for automotive air conditioning
- Intelligent inspection systems, with a Digital Twin approach

Goals:

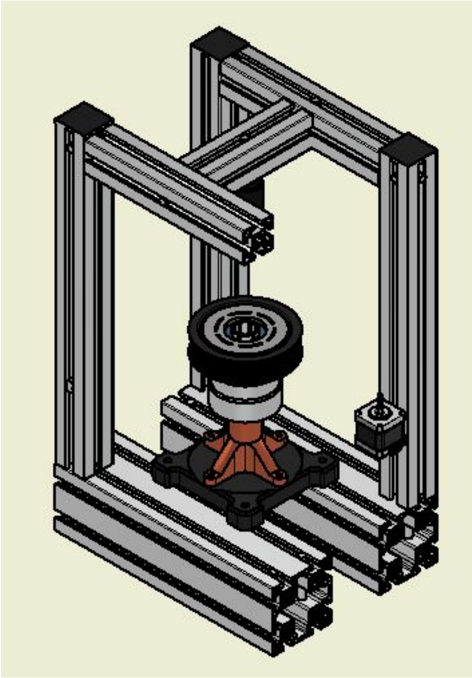
- Ensure critical quality control processes
- Provide real-time management and analytics through a platform
- Minimize cycle time of the production line

Digital Twin Approach

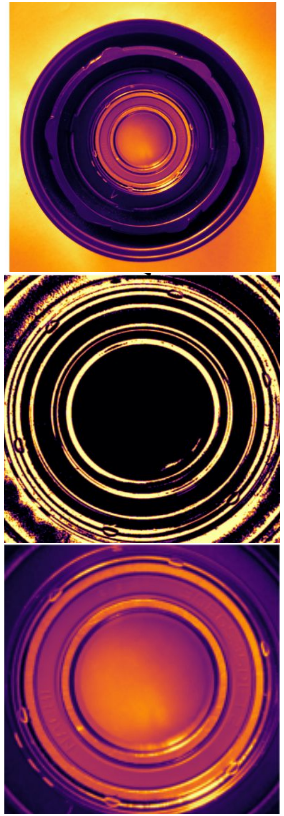


DIGITAL TWIN LOOP
real-time, automated, and low latency

Digital Twinning

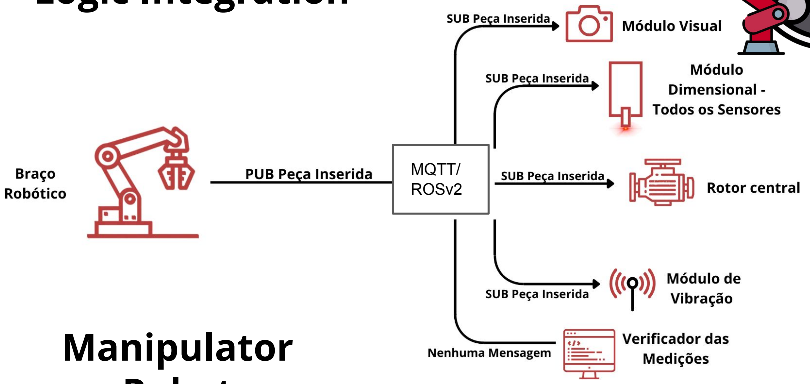


Physical Integration

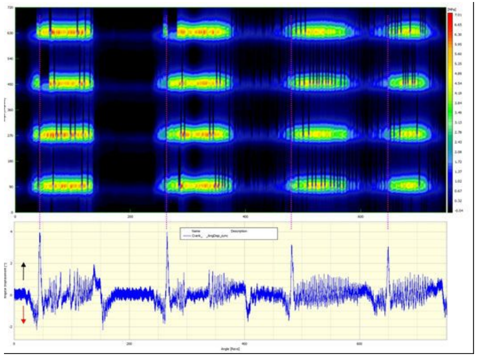


Visual and Dimensional Inspection

Logic Integration



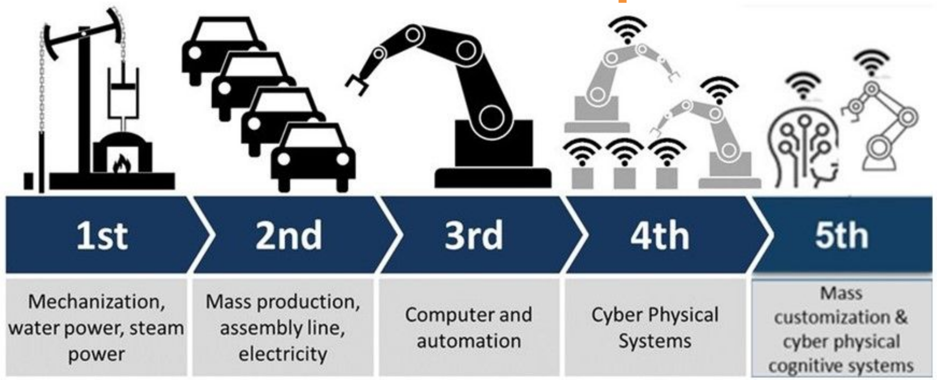
Manipulator Robot



Vibration Inspection

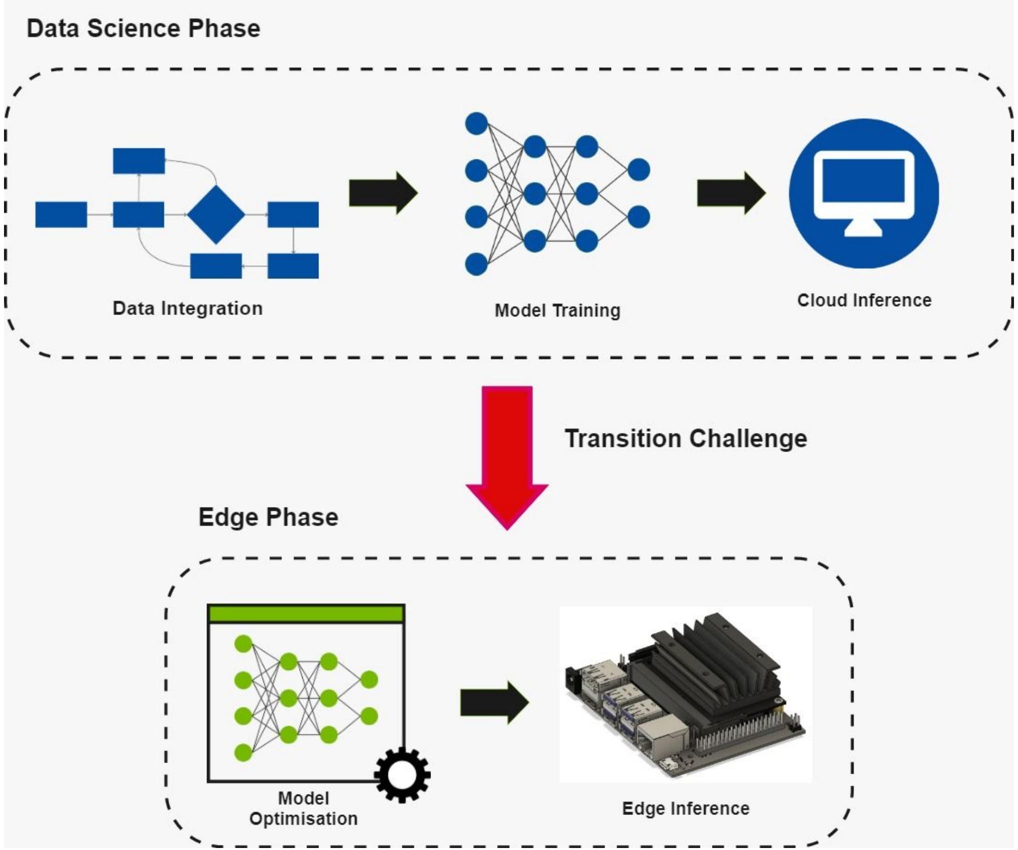
How can we keep the user-in-the-loop?

15 seconds
Full automated inspection



Failure analysis
“Debug mode”
Cobot - Collaborative Robot

Transition Challenge

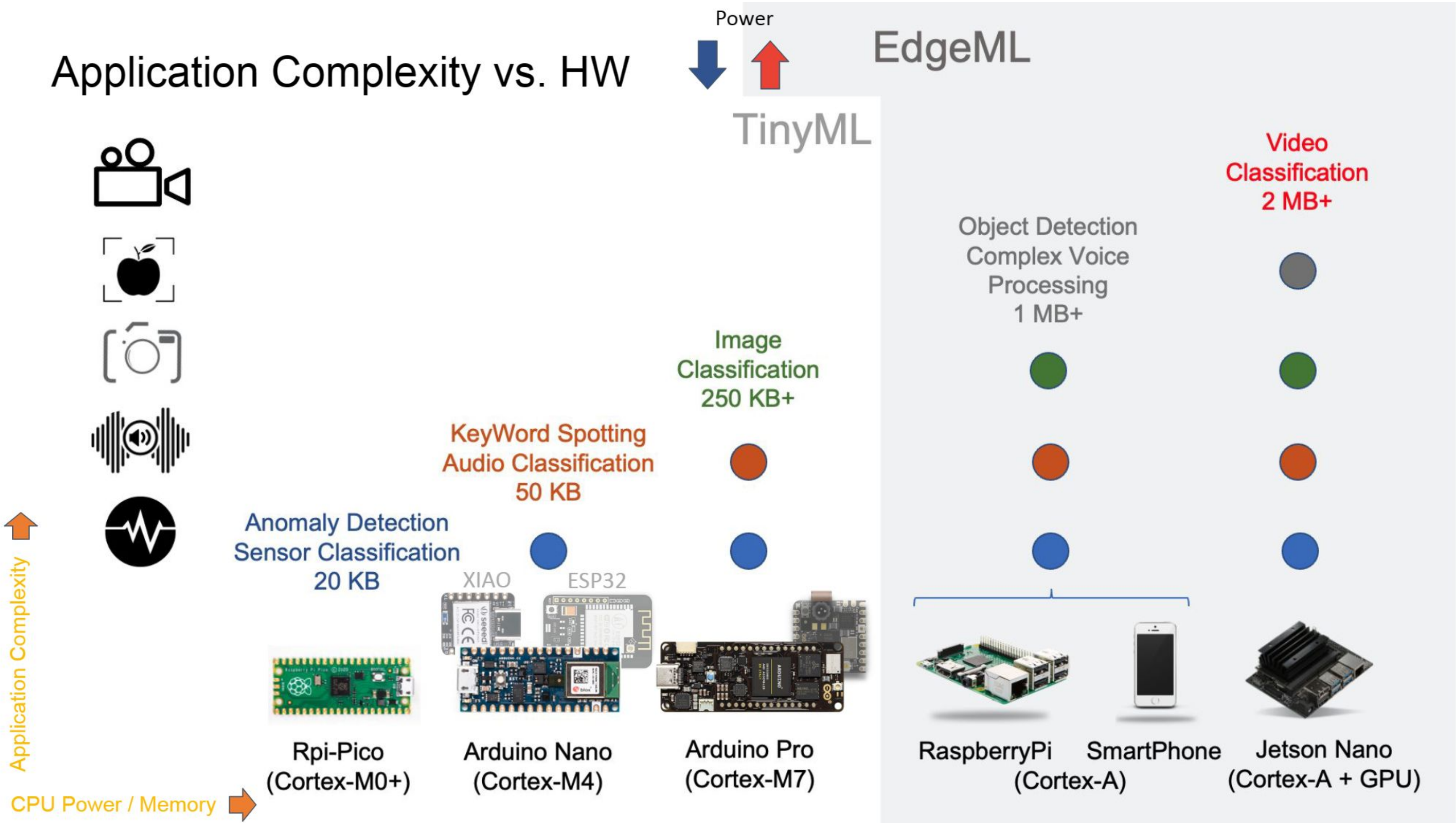


Understand system limitations
Transition challenge: taking models trained in the cloud (**Data Science Phase**) into resource-limited edge devices **(Edge Phase**).

RQ 1 - **Performance and accuracy impact** of taking cloud-based models to resource-constrained devices at the network edge?

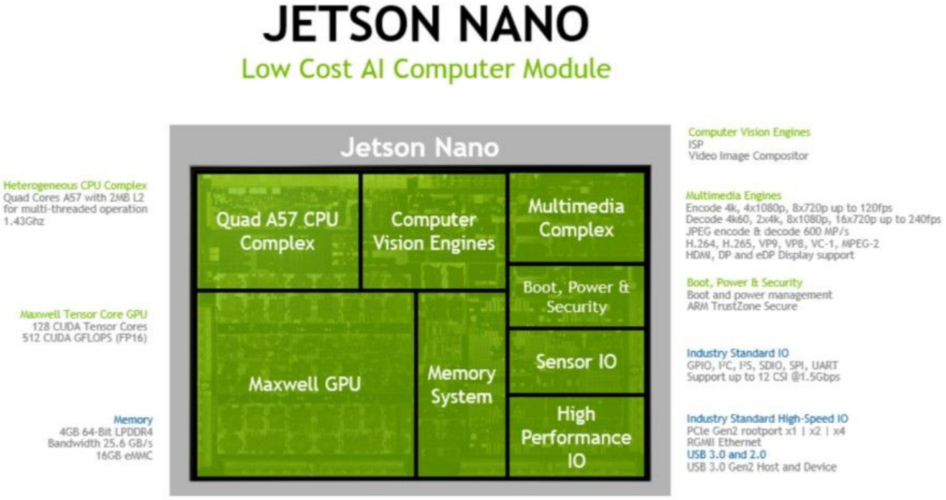
RQ 2 - **Power footprint** in running image-based ML in edge devices. **TensorRT** impact on deployed models in terms of **performance-watt**

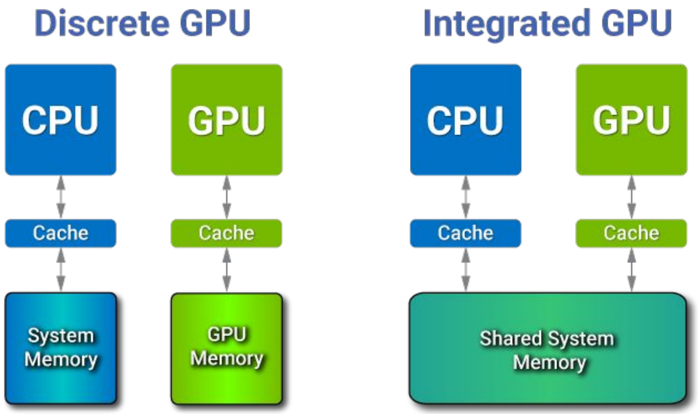
Meireles, L. et al. The not-so-easy task of taking heavy-lift ML models to the edge: a performance-watt perspective. In Proc. ACM SAC 2023.



Marcelo Rovai, 18/04/2023, @12.04PM Brazil time

Jetson Nano Architecture



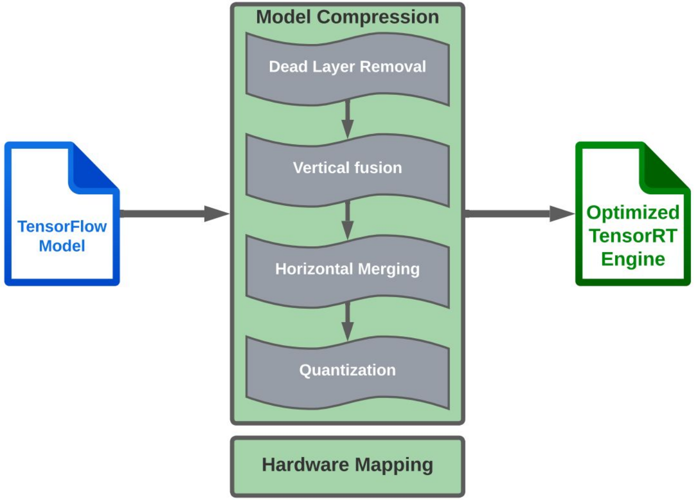


System Profiling to the Transition Challenge
Experimental Design

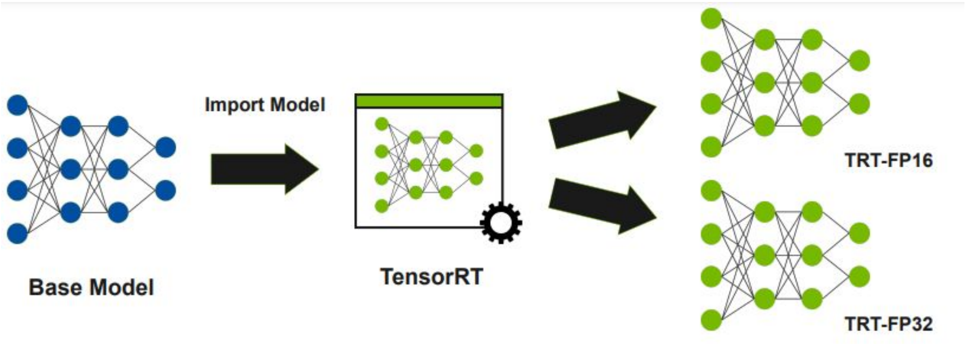
- Jetson Nano is an entry-level Edge device by NVIDIA with parallel cores for AI applications
- TensorRT is a high-performance library that interfaces deep learning applications with production environments

Table 2: NVIDIA JETSON Nano Specifications

CPU	ARM Cortex-A57 (quadcore)	@1.73GHz
GPU	256-core Maxwell	@998MHz
Memory	4GB 64-bit LPDDR4	@1600MHz — 25.6 GB/s
Storage	16GB eMMC 5.1	-
Power	10W	-
Jetpack	4.6	[L4T 32.6.1]
CUDA	10.2.300	-
cuDNN	8.2.1.32	-
TensorRT	8.0.1.6	-



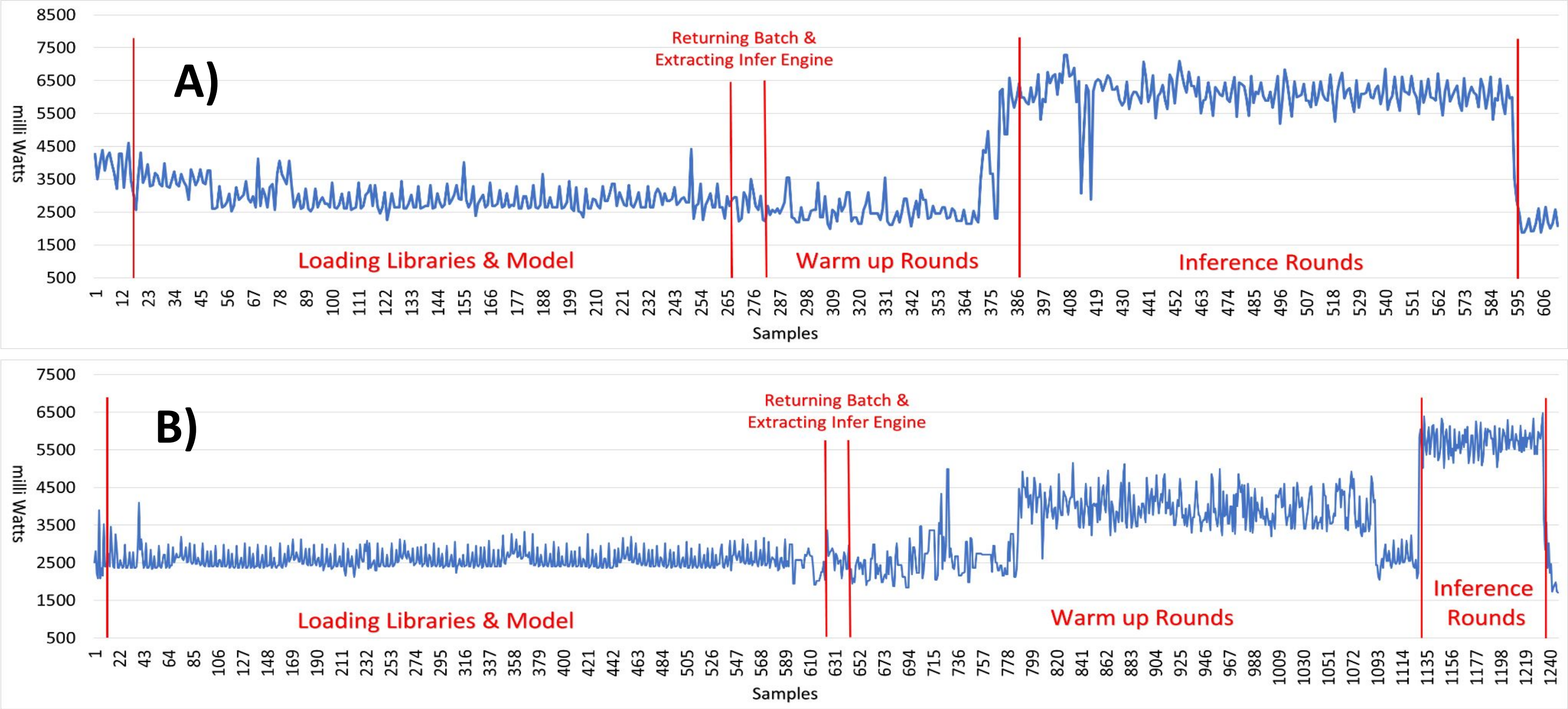
System Profiling to the Transition Challenge
Experimental Design



- Cloud-based or base model
- TRT-FP16 model
- TRT-FP32 model

- **Loading Model** - Initial processes (e.g. loading packages, initializing variables)
- **Extracting Infer Engine & Returning Batch** - manual operation to provide the code with the data and structure capable of passing it through the model.
- **Warm Up Rounds** - in the first rounds, it is still necessary to cache data and other procedures. So **warm-up rounds** are important to avoid "cold start" problems.
- **Real Rounds** - inference rounds when energy profile and resource usage are assessed.

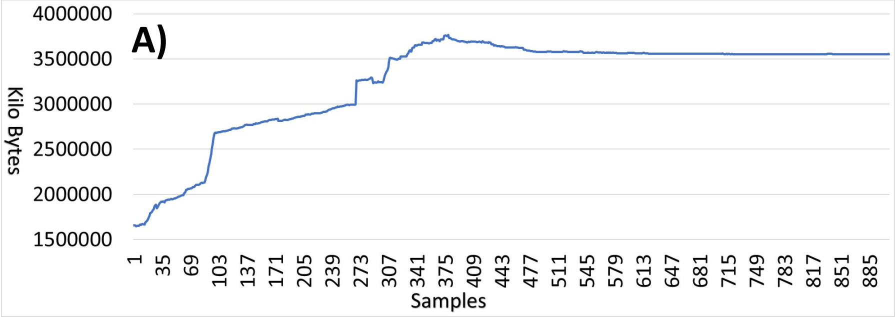
Power Consumption profiling

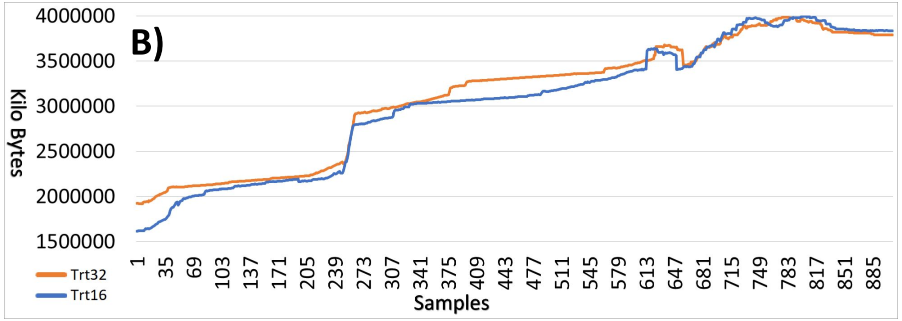


Power consumption: MobileNetV2 on the Jetson Nano platform (A) and its TRT-FP32 model (B) . The y axis is the instantaneous power (in milli Watts); x axis is the time (in samples)

Edge results

	Cloud-based	TRT-FP32	TRT-FP16
Accuracy (%)	97	97	97
Throughput (FPS)	41	82	84
Memory use (%)	93,2	99,5	99,8
Total Avg Power (mW)	3386	3185	3229
Inference Avg Power 2 (mW)	6125	5711	5860
Total Performance-watt (FPS/W)	12,10	25,74	26,01
Inference Performance-watt (FPS/W) 2	6,69	14,35	14,33





RAM memory consumption for baseline model (A), TRT-FP16 and TRT-FP32 models (B).

Take-away message

- Industry 5.0:
  - Human-centric, sustainable and resilient
- Automation versus human-in-the-loop (4.0 -> 5.0)
  - Quality control of automotive parts in a production cell
  - Achieved automation (15 seconds) of manual processes
  - Created a new process of part failure analysis, “Debug Mode”, collaborate with the machine, press enter, analyse, press enter, etc.
- Transition challenge: cloud trained model to edge devices
  - Power consumption of Jetson Nano
  - Performance improvement of optimised models (FPS)
  - Performance-watt improvement

Acknowledgements

Eder Gonçalves, Bruno Oliveira, Gabriel Souza, Marcelo Malheiros, Nelson Duarte Filho, Johann Pires, Rafaella Quaresma, Eduardo Borges, Nicolas Nobre, Vinícius Menezes, Victor Coch, Mateus Borges, Thássio Silva, Anajara Martins, Bruna Guterres, Lucas Meirelles, Sílvia Botelho, Paulo Drews