AGH University of Krakow

Wprowadzenie do…

Wbudowanych systemów AI

Prowadzący: mgr inż. Piotr Wzorek
Wykładowca: dr inż. Tomasz Kryjak

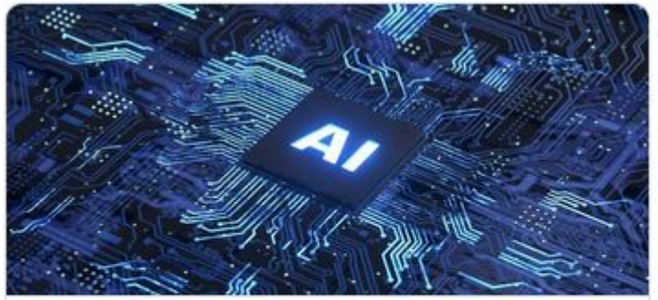
C2, 107 - AGH University of Krakow
6 October, 2023

agh.edu.pl

Before we start…

Prosze zapisać się na kurs UPEL

Name of the course: **Wbudowane systemy AI**
Password: **WsAI_2023**



Proszę przeczytać BHP i regulamin oraz podpisać listę!

While you’re listening…

You don’t understand? **Tell me!**

I’m going too fast? **Tell me!**

You have a question? **Just ask it!**

Tell me about yourselves!

How many of you have **ever** used Python language?

Do you have **any** experience with AI? ML? DNN? CNN?

Do you have **any** experience with embedded platforms? FPGAs? eGPUs?

How many of you have **ever** used Pytorch?

I guess it’s now my turn…

I’m a PhD Student here at AGH University

I specialise in Embedded Vision and Event Cameras

I love teaching!

I play bridge (and it’s awesome!!)

How will the classes look like?

1. Short introduction
2. Working independently with the instructions
3. Presentation of the solution to the teacher
4. Uploading solution to the UPEL platform

FAQ

Yes, you can use your own PC!
You should not send solutions to UPEL until you show it!
The presence is mandatory! (maximum 2 absences)
You should still complete the task if you're absent!
You have 2 weeks to complete the task!

Cheating and fairness

You CAN!

- **ask me for help**
- consult the problems encountered with other students
- work together on harder exercises
- use tools for developers (stackoverflow, GitHub Copilot, ChatGPT)

You MAY NOT!

- present solutions other than your own,
- present solutions without a basic understanding of how they work.

Don’t cheat, just ask! I will help you!

Metody zaliczenia przedmiotu

Laboratory

Na laboratorium będzie obowiązywał system punktowy. Na łączną sumę 100 punktów składają się punkty za wykonanie ćwiczeń (70 pkt.) oraz mini-projektu (30 pkt.).Na podstawie łącznej sumy punktów zostanie wystawiona ocena zgodnie z obowiązującym Regulaminem Studiów na AGH. Na końcu semestru odbędzie się kolokwium zaliczeniowe w formie ustnej uwzględniające materiał z wykładu oraz ćwiczeń. Aby otrzymać zaliczenie z przedmiotu, należy uzyskać pozytywną ocenę zarówno z kolokwium, jak i z laboratorium

Ocena końcowa = 0.7 x ocena z laboratorium + 0.3 ocena z kolokwium

1 unexcused absence - maximum grade of 4.0
2 unexcused absences - maximum grade of 3.0
3 unexcused absences - you will **NOT** receive a credit.

Ustalenie terminów zajęć

Mamy do zrealizowania 14 zajęć. Chcemy to zrobić przed świętami…

Opcja SPEEDRUN
7x środa 8:00, 7x czwartek 9:45
Skończymy do 30 listopada

Opcja PEACEFUL
10x czwartek 9:45
4x środa 8:00 (np cztery środy w listopadzie)
Skończymy do 13 grudnia

Opcja CUSTOM
Jakieś pomysły?

Consultations

Plan A

One mutually agreed consultation date.

Plan B (prefered by me)

Contact by e-mail for questions, problems. Make an appointment privately, either remotely or in person.

Piotr Wzorek
pwzorek@agh.edu.pl
309, C2

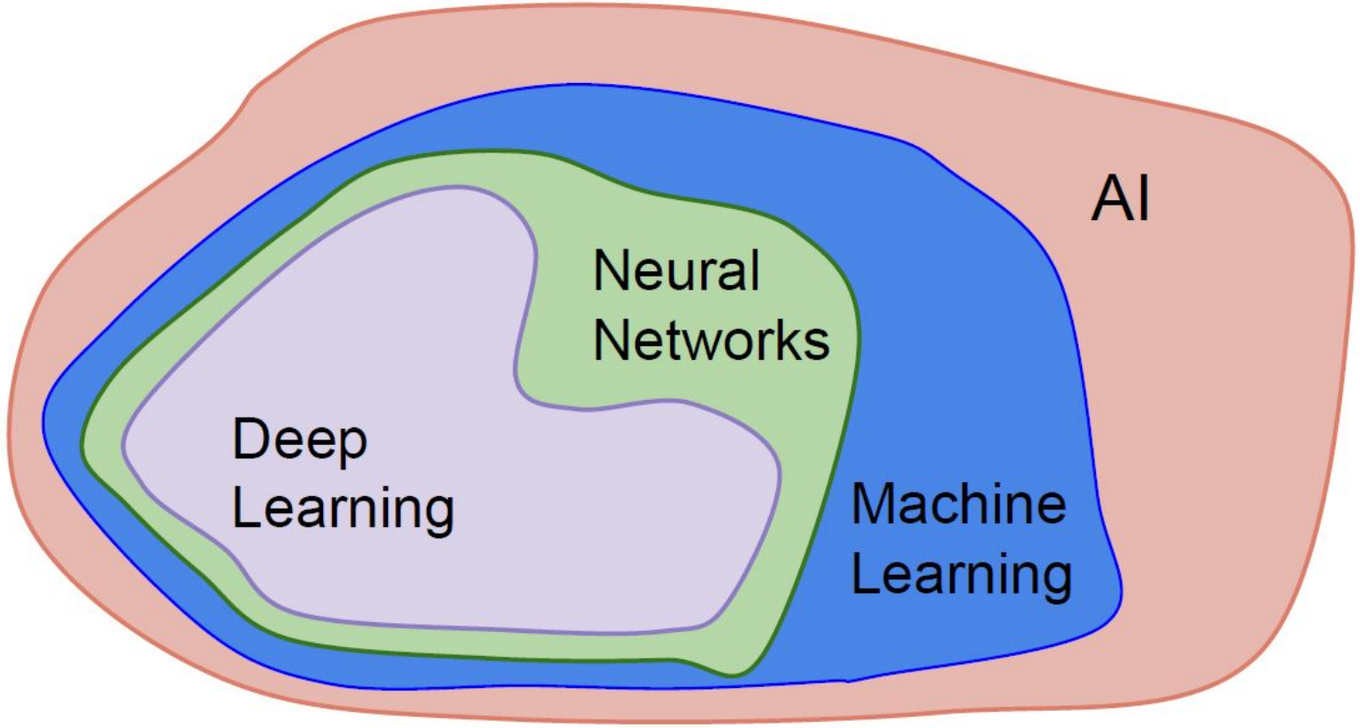
Krzysztof Błachut
kblachut@agh.edu.pl
308, C2

dr inż. Tomasz Kryjak
kryjak@agh.edu.pl
113, C3

Embedded AI - What is it?

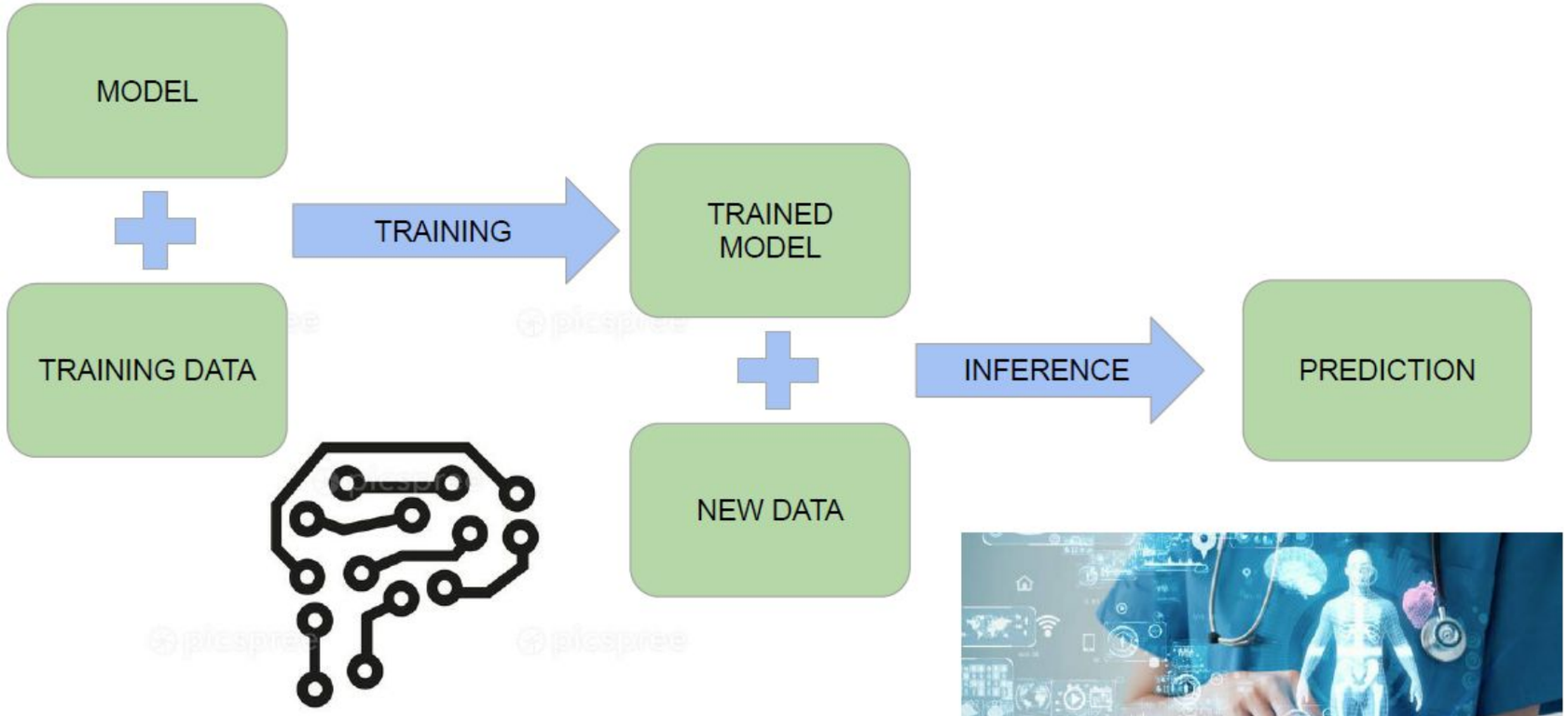
What is AI?

AI is the simulation of human intelligence by machines



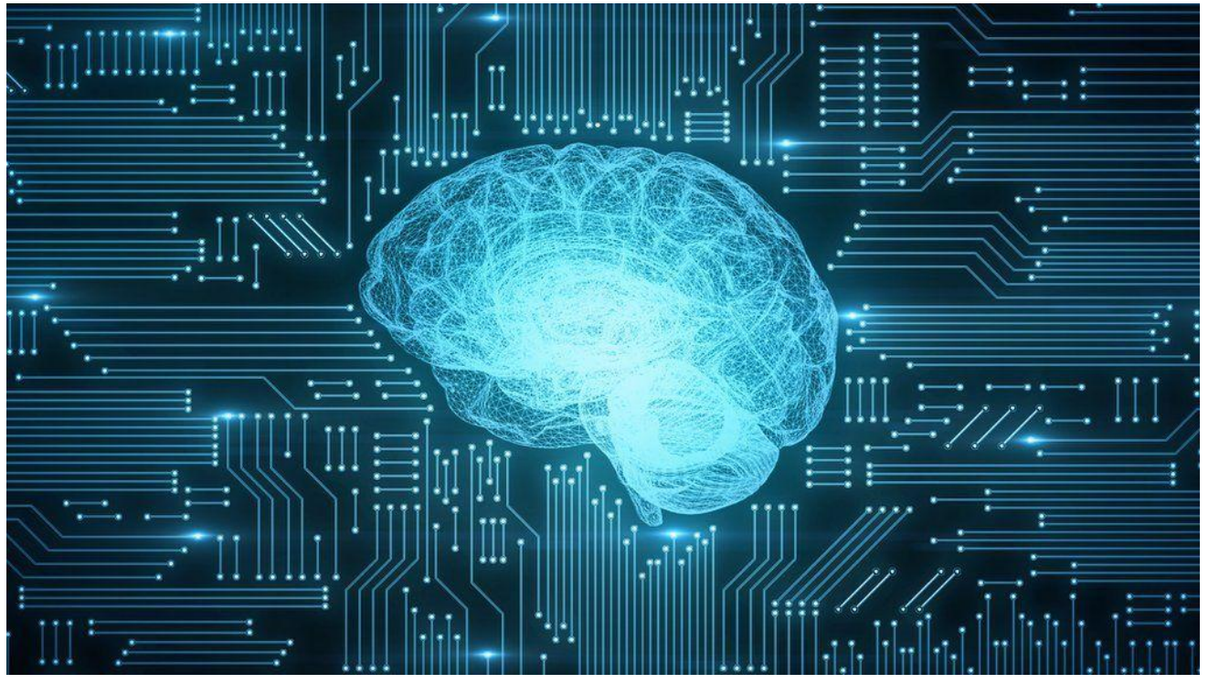
Embedded AI - What is it?

What is AI?



Embedded AI - What is it?

AI is accurate!
AI is flexible!
AI is awesome!



Embedded AI - What is it?

AI is accurate!
AI is flexible!
AI is awesome!

It is, however…
Computationally complex
Memory hungry

Embedded AI - What is it?

AI is accurate!
AI is flexible!
AI is awesome!

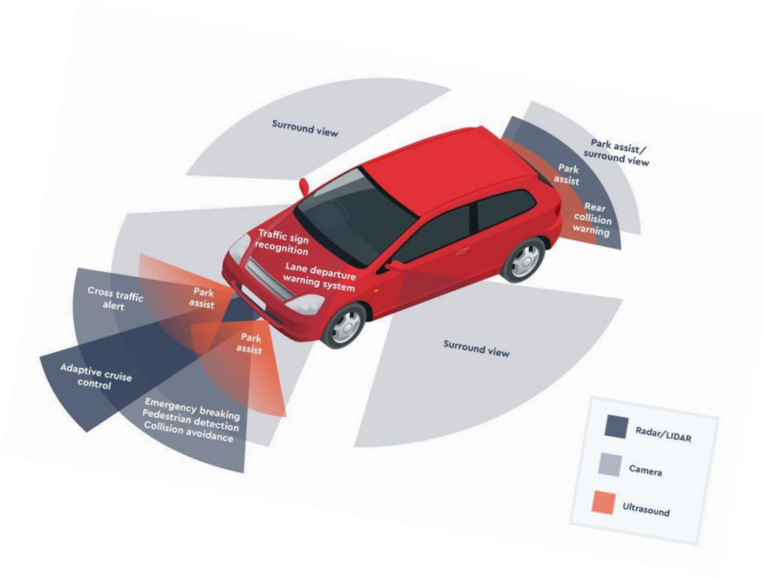
It is, however…
Computationally complex
Memory hungry

We can use Cloud Computing…
We can buy powerful GPUs…

Embedded AI - What is it?

AI is accurate!
AI is flexible!
AI is awesome!

It is, however…
Computationally complex
Memory hungry



We can use Cloud Computing…
We can buy powerful GPUs…
But not always!

Embedded AI - What is it?

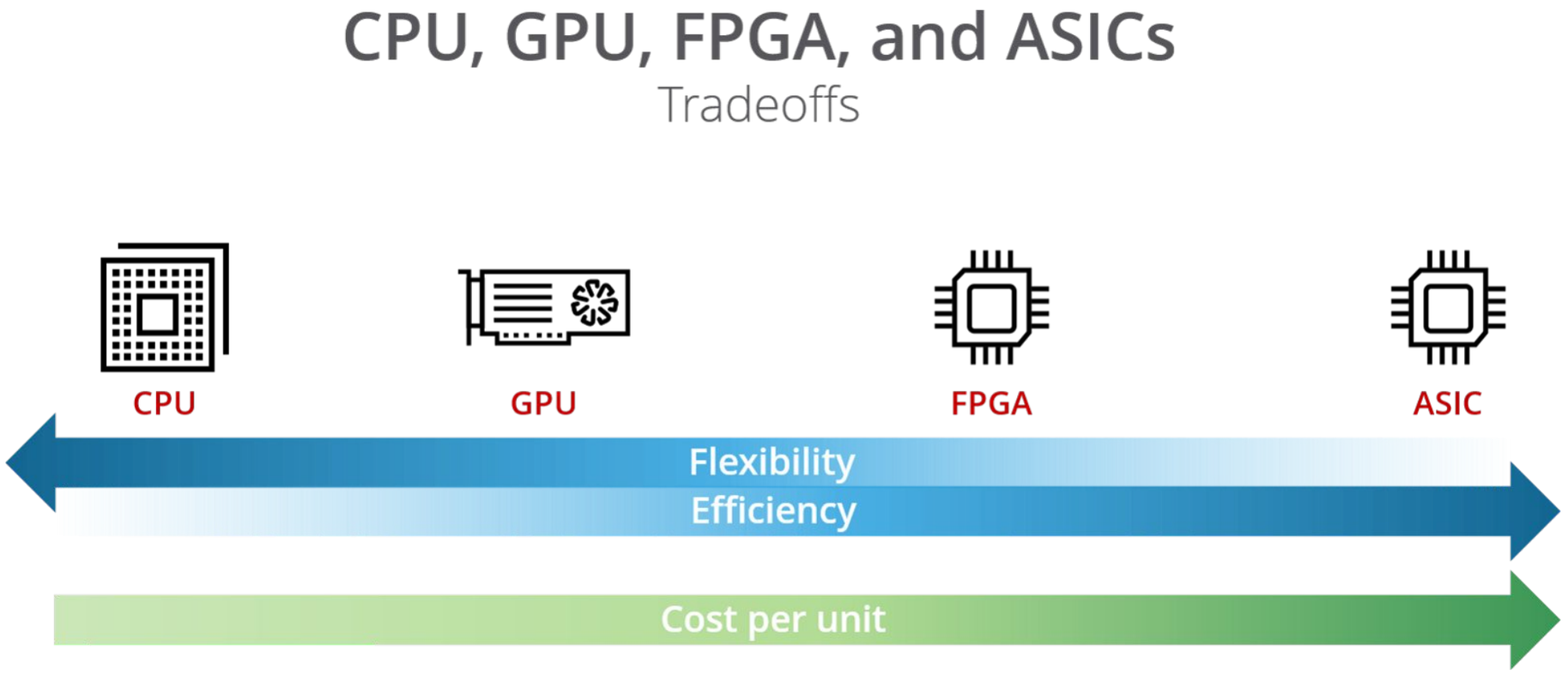
AI is accurate!
AI is flexible!
AI is awesome!

It is, however…
Computationally complex
Memory hungry

We can use specialised Hardware to deal with complex operations!
We can optimize models to reduce the weights!

We can accelerate AI to apply it in real time!

Embedded AI - What is it?



Embedded AI - Topics…

1. Neural Network Quantization
2. Neural Network Pruning
3. Unusual, hardware-focused Neural Networks
4. Hardware implementation
5. AI accelerators

Short project

Embedded AI - Topics…

1. **Neural Network Quantization**

Quantization is the process of reducing the precision of the weights, biases, and activations such that they consume less memory.

2. Neural Network Pruning
3. Unusual, hardware-focused Neural Networks
4. Hardware implementation
5. AI accelerators

Short project

Embedded AI - Topics…

1. Neural Network Quantization
2. **Neural Network Pruning**

Neural network pruning is the process of methodically eliminating parameters from an existing network in order to minimize the resource requirements.

3. Unusual, hardware-focused Neural Networks
4. Hardware implementation
5. AI accelerators

Short project

Embedded AI - Topics…

1. Neural Network Quantization
2. Neural Network Pruning
3. **Unusual, hardware-focused Neural Networks**

Spiking neural networks (SNNs) are neural networks that more closely mimics human neurons.
Binary neural networks are an extreme case of Network Quantization (only one bit!)

4. Hardware implementation
5. AI accelerators

Short project

Embedded AI - Topics…

1. Neural Network Quantization
2. Neural Network Pruning
3. Unusual, hardware-focused Neural Networks
4. **Hardware implementation**

We will perform Hardware Implementation of our Neural Networks.
We are going to run neural networks inference on eGPUs and/or FPGAs

5. AI accelerators

Short project

Embedded AI - Topics…

1. Neural Network Quantization
2. Neural Network Pruning
3. Unusual, hardware-focused Neural Networks
4. Hardware implementation
5. **AI accelerators**

We will use some AI accelerators / computation modules like Coral or Intel Neural Compute Stick

Short project

Embedded AI - Topics…

1. Neural Network Quantization
2. Neural Network Pruning
3. Unusual, hardware-focused Neural Networks
4. Hardware implementation
5. AI accelerators

Short project

AGH

RESEARCH UNIVERSITY
EXCELLENCE INITIATIVE

AGH University
of Krakow

Any questions?