Math Foundations of Data Science
Course Introduction

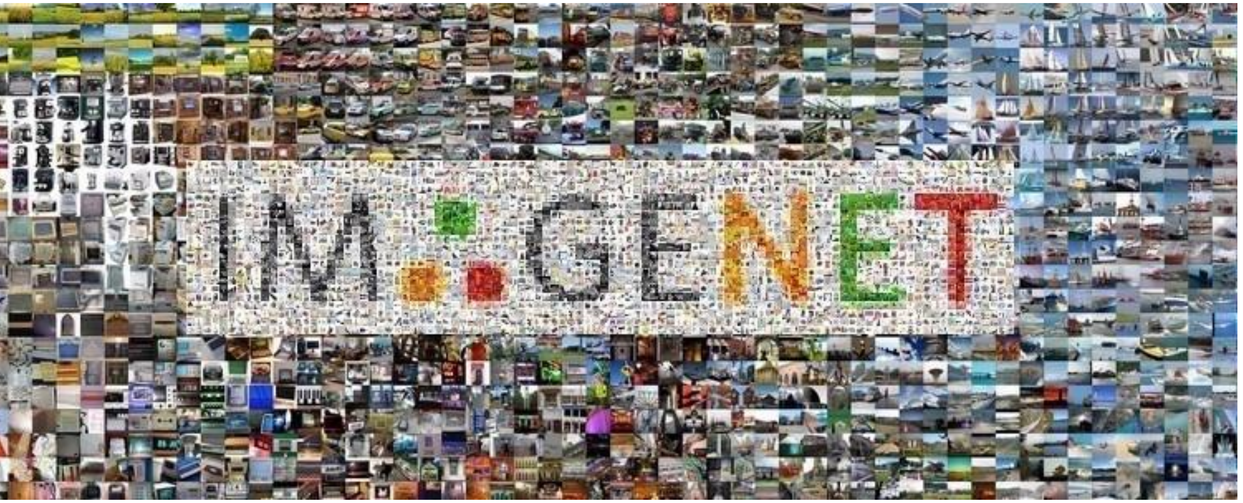
Outline

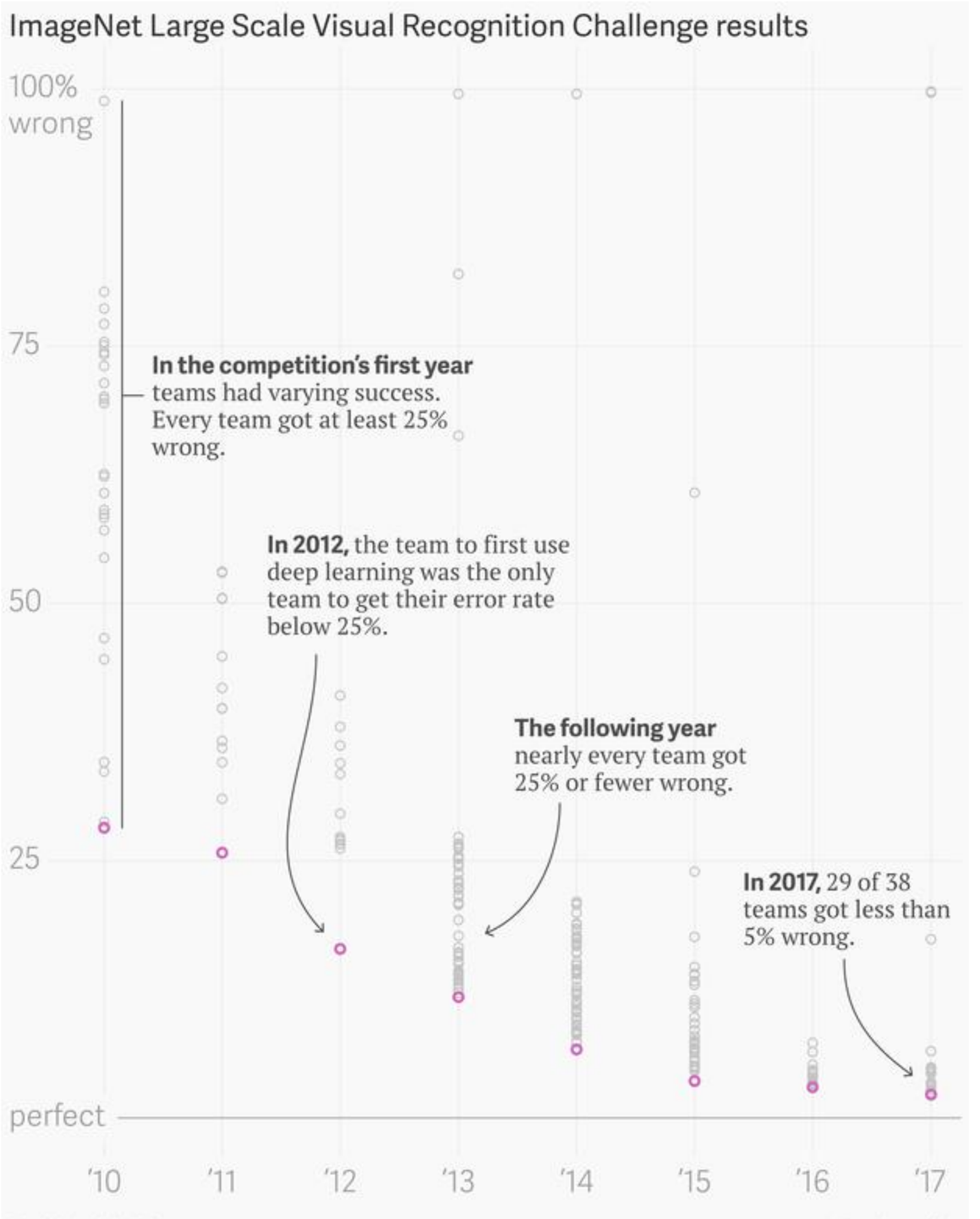
- Why this course
- Course logistics

Part I: Why this course

in the era of AI

Classify Images





https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

Detect and Segment Objects

https://github.com/matterport/Mask_RCNN

Style Transfer

https://github.com/StacyYang/MXNet-Gluon-Style-Transfer

Generative Modeling

4.5 years of face generation

2014

2015

2016

2017

2018

http://www.whichfaceisreal.com/methods.html

Generative Modeling (2024)

Generated by OpenAI Sora

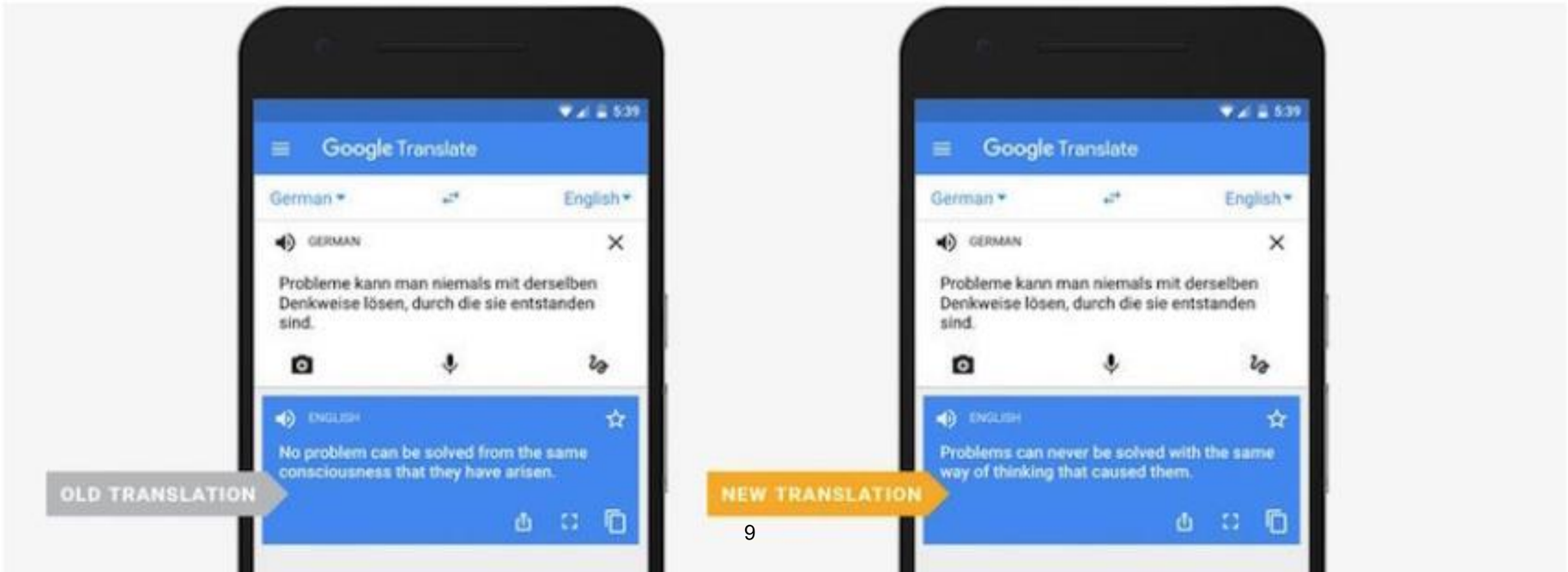
Machine Translation

https://www.pcmag.com/news/google-expands-neural-networks-for-language-translation

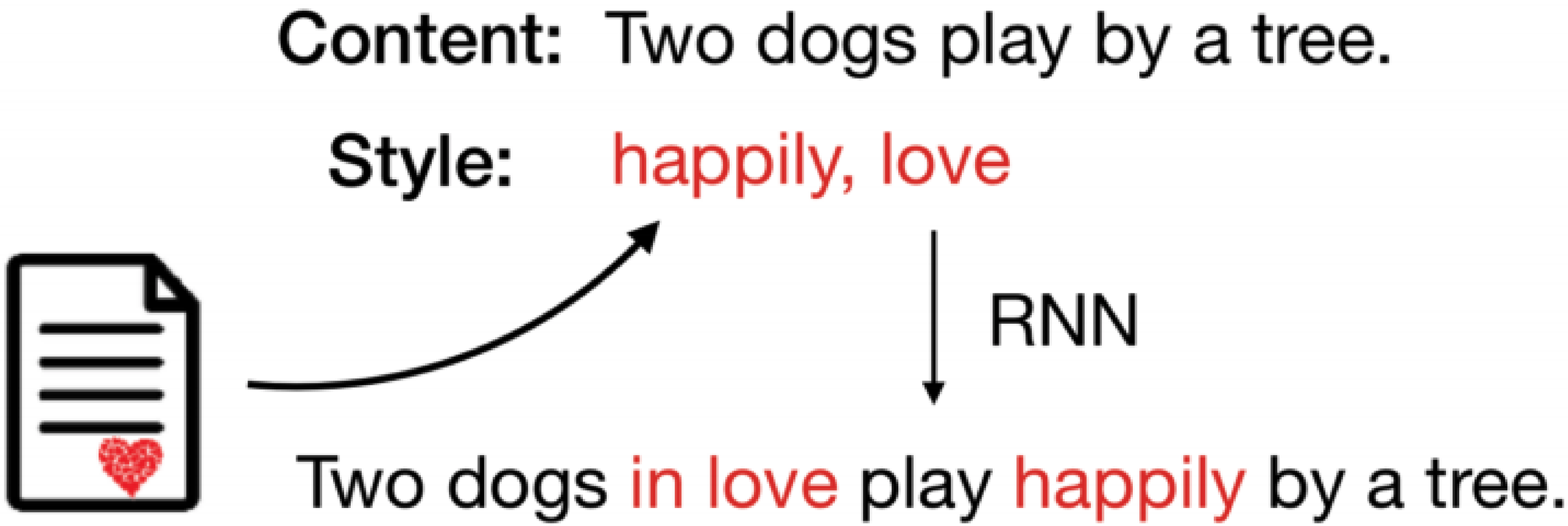
Google Expands Neural Networks for Language Translation

The new system can translate whole sentences at a time, rather than just phrases.

By Tom Brant November 15, 2016



Text Synthesis



Li et al, NACCL, 2018

OpenAI ChatGPT (in 2022):

me: why is carbon dioxide highly sought after?

ChatGPT: In recent years, CO2 has also become highly sought after as a potential source of renewable energy. When burned, CO2 releases a large amount of heat, which can be harnessed to generate electricity. It can also be used to power fuel cells, which are a type of clean energy technology that converts chemical energy into electricity.

(Accessed in Dec 2022 by Jerry Zhu)

OpenAI ChatGPT (in 2022):

me: which molecules are produce when burning co2?

ChatGPT: When carbon dioxide (CO2) is burned, it reacts with oxygen in the air to produce carbon monoxide (CO) and water (H2O). The chemical reaction for this process can be written as:
CO2 + O2 -> CO + H2O

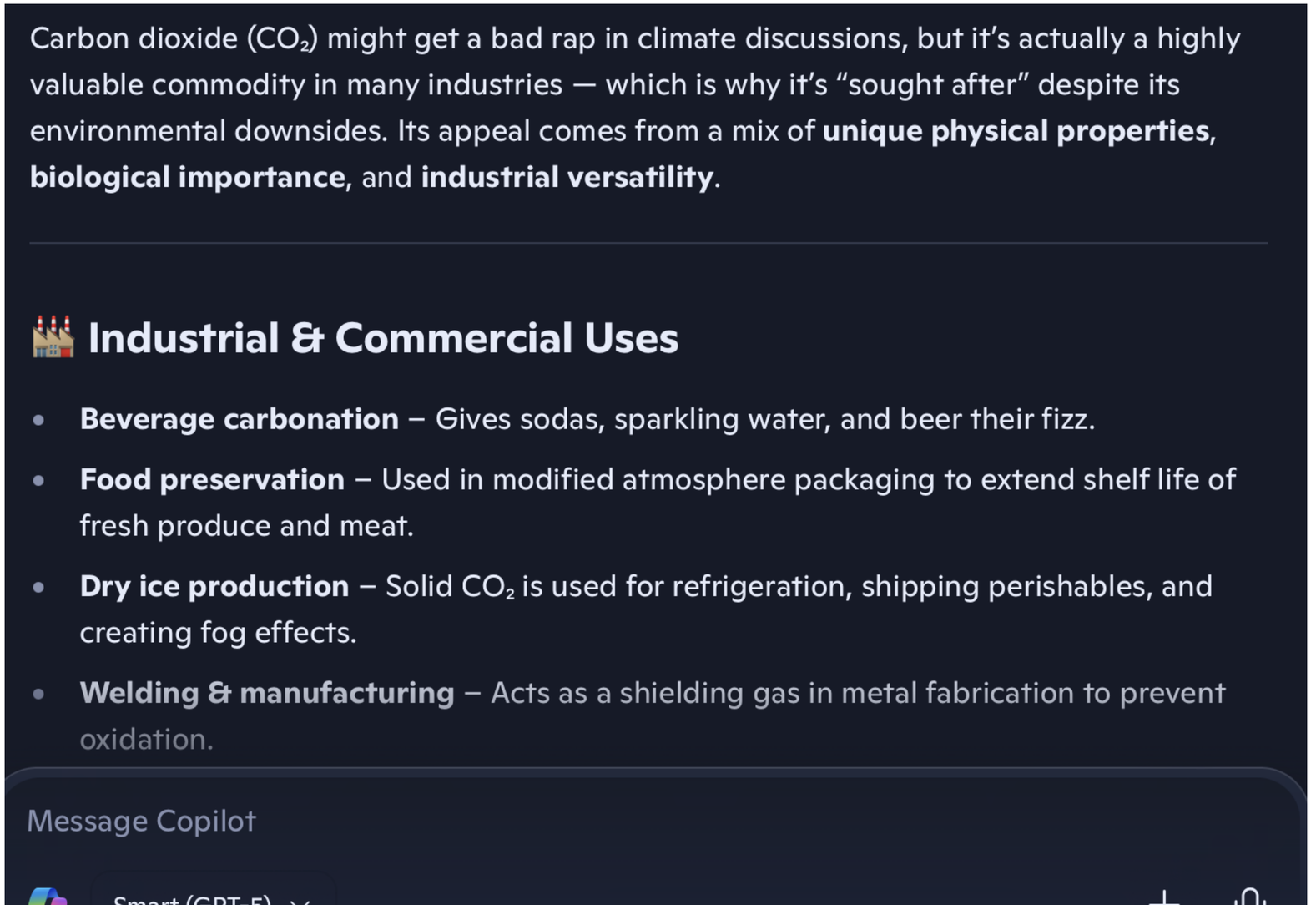
The exact products of the combustion of CO2 depend on the conditions under which it is burned, such as the temperature, oxygen concentration, and presence of other chemicals. However, in general, CO and H2O are the primary products of the burning of CO2.

(Accessed in Dec 2022 by Jerry Zhu)

OpenAI GPT5 (Sept 2, 2025)

me: why is carbon dioxide highly sought after?

GPT5:



Carbon dioxide (CO_2) might get a bad rap in climate discussions, but it's actually a highly valuable commodity in many industries — which is why it's "sought after" despite its environmental downsides. Its appeal comes from a mix of **unique physical properties**, **biological importance**, and **industrial versatility**.

🏭 Industrial & Commercial Uses

- **Beverage carbonation** – Gives sodas, sparkling water, and beer their fizz.
- **Food preservation** – Used in modified atmosphere packaging to extend shelf life of fresh produce and meat.
- **Dry ice production** – Solid CO_2 is used for refrigeration, shipping perishables, and creating fog effects.
- **Welding & manufacturing** – Acts as a shielding gas in metal fabrication to prevent oxidation.

(Accessed by Yingyu Liang)

OpenAI GPT5 (Sept 2, 2025)

me: What are some important research problems in the field of AI?

GPT5:

AI research is full of fascinating, high-impact challenges — some are deeply technical, others are ethical, social, or even philosophical in nature. Here's a curated snapshot of some of the most important open problems being tackled today:

🧠 Core Technical Challenges

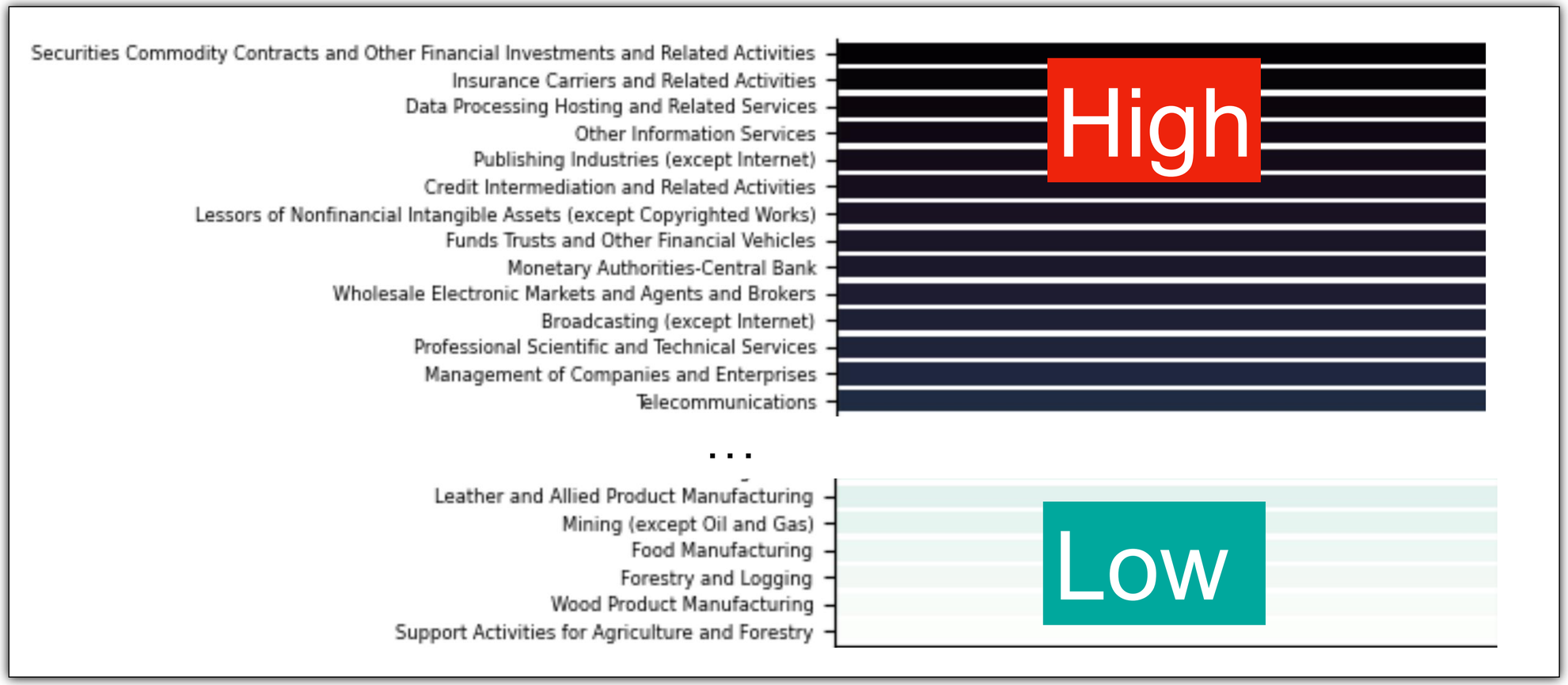
- **Robustness & Reliability** – Making AI systems less prone to catastrophic errors in recognition, reasoning, or decision-making [1].
- **Generalization & Transfer Learning** – Enabling models to learn from fewer examples, adapt to new tasks, and avoid over- or under-generalizing [1].
- **One-Shot / Few-Shot Learning** – Achieving human-like ability to learn concepts from a single or very small number of ex[illegible]les [1].
- **Catastrophic Forgetting** – Preventing models from losing previously learned knowledge when trained on new data [1].

(Accessed by Yingyu Liang)

Will AI take my job?

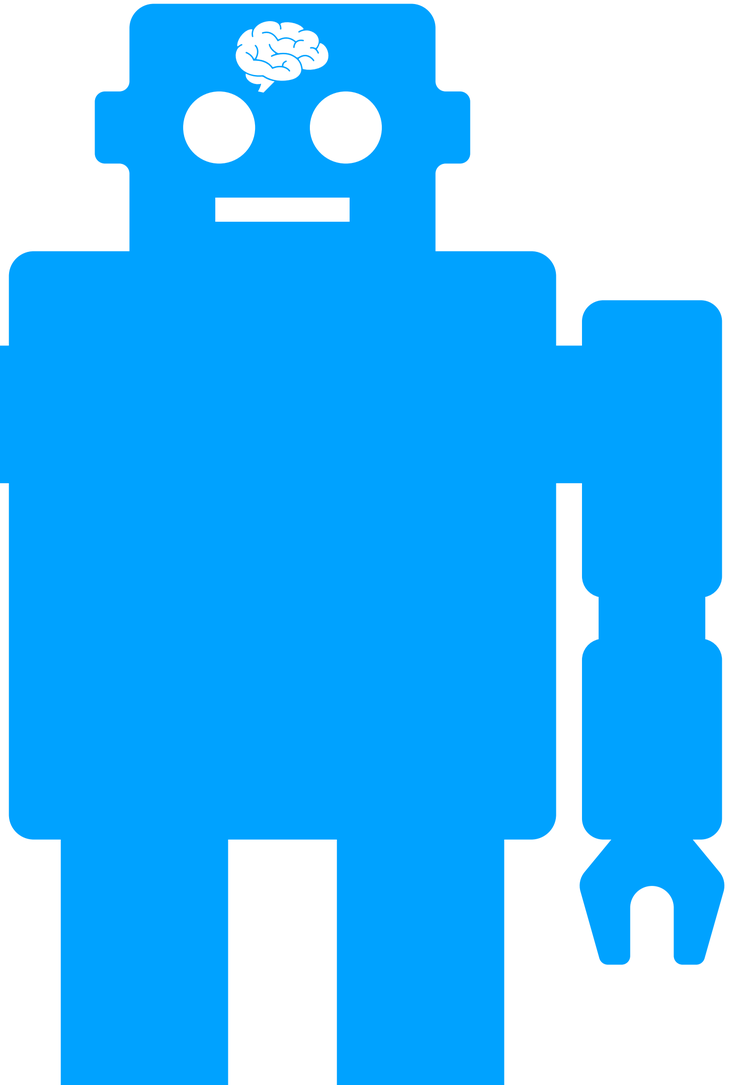
Maybe

- The more AI helps your job, the higher the replacement risk



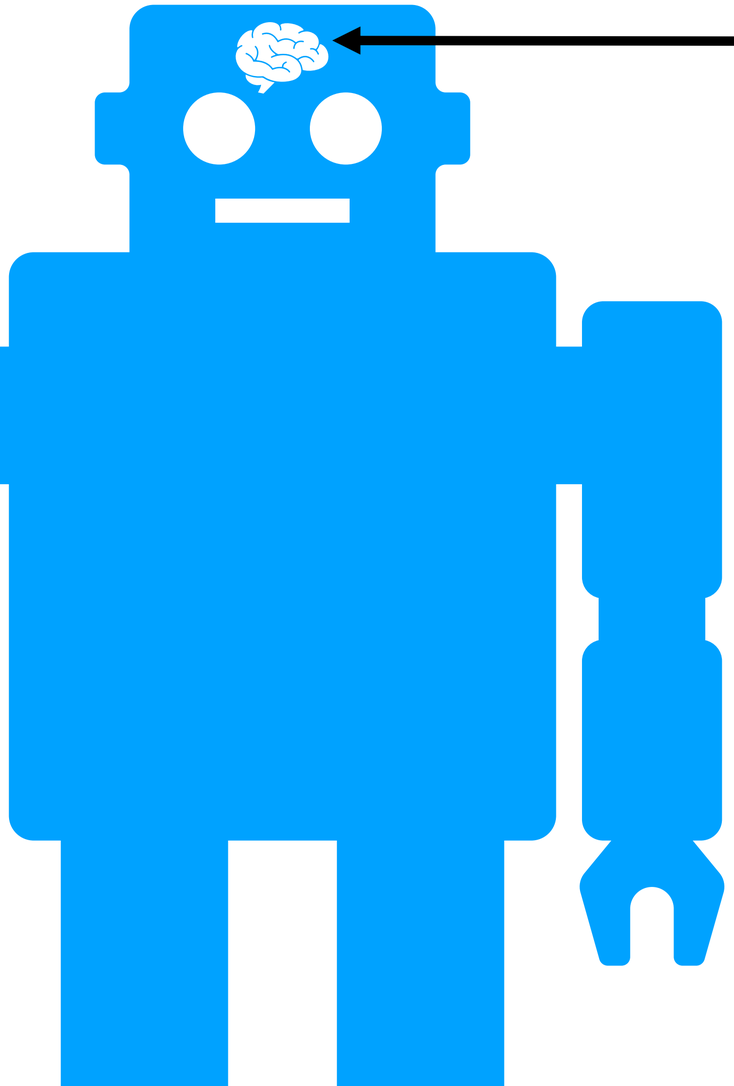
[GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. Eloundou et al. 2023]

AI Venn diagram



Artificial Intelligence

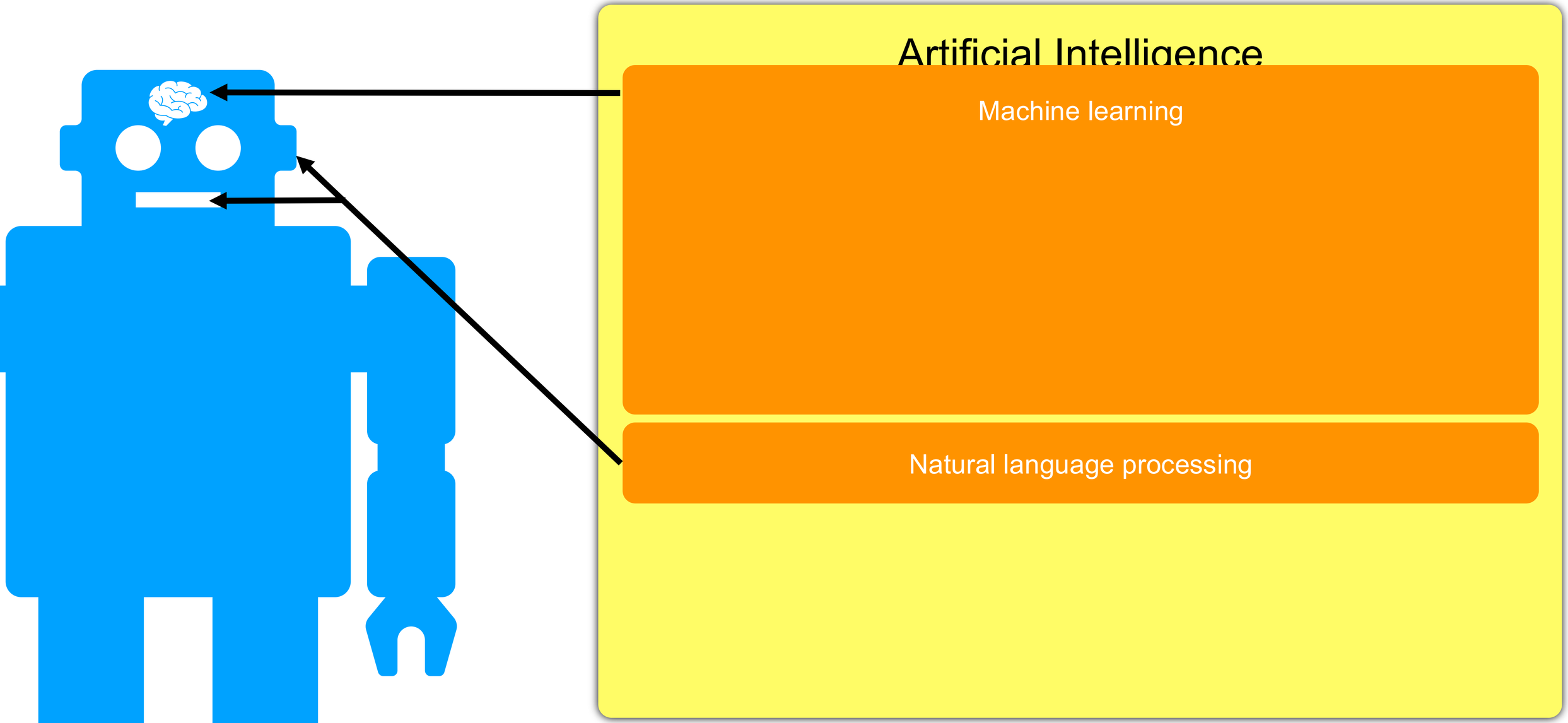
AI Venn diagram



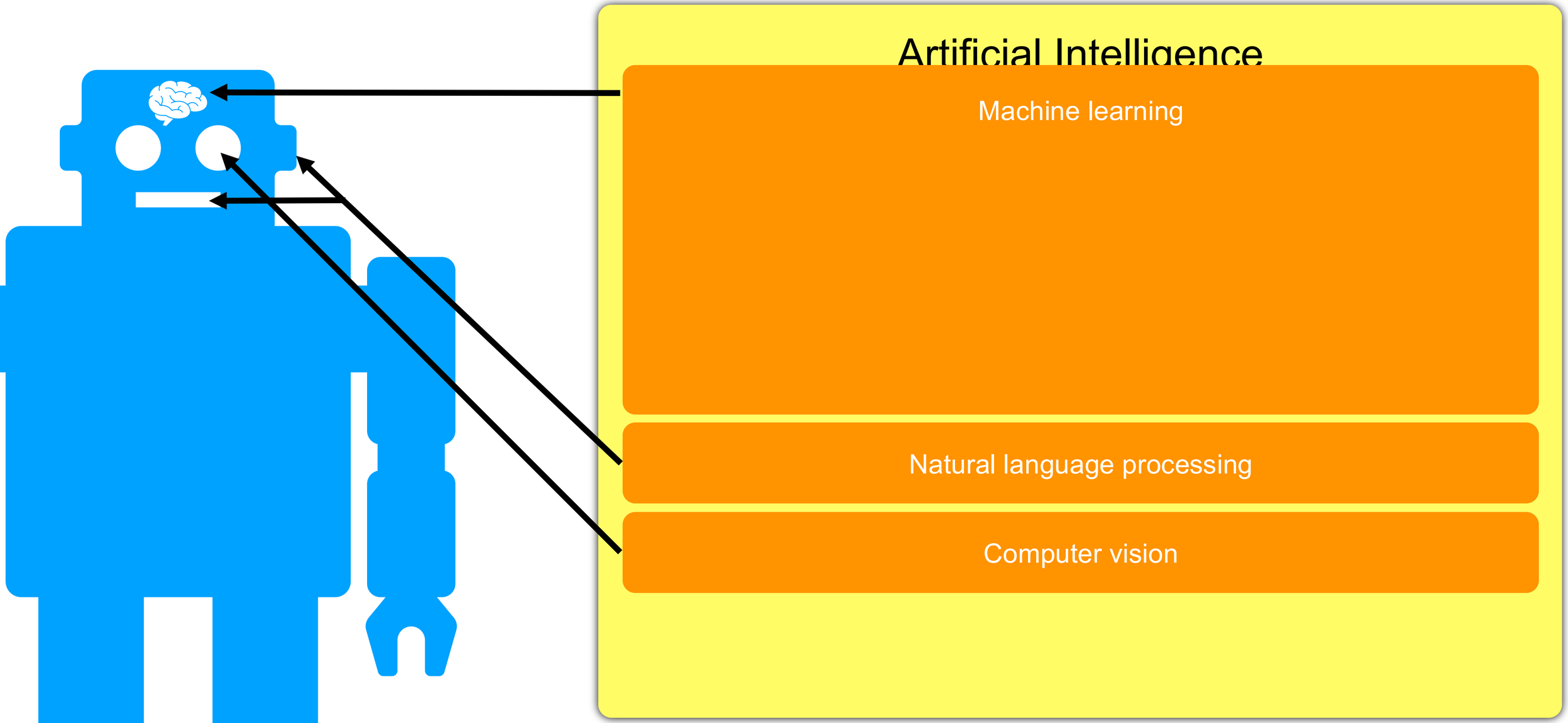
Artificial Intelligence

Machine learning

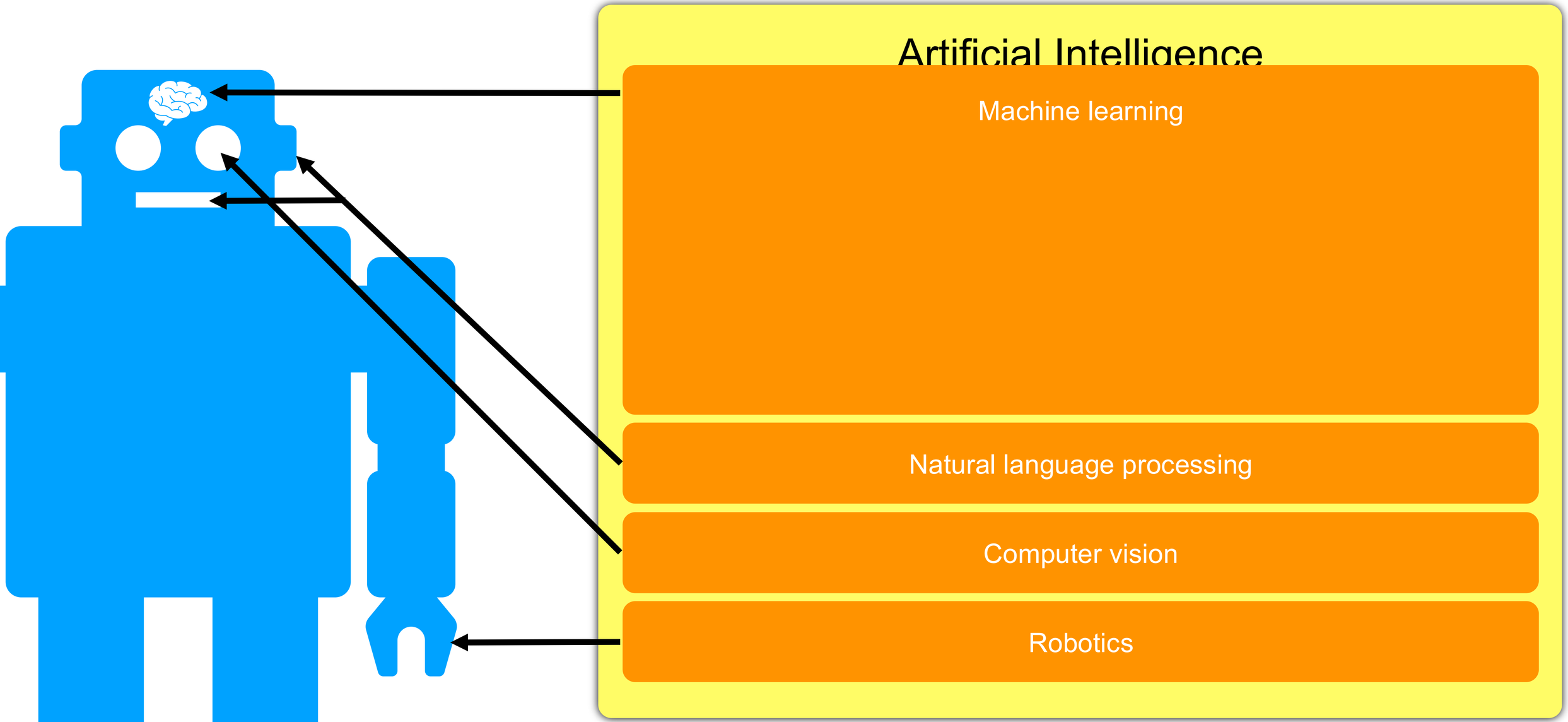
AI Venn diagram



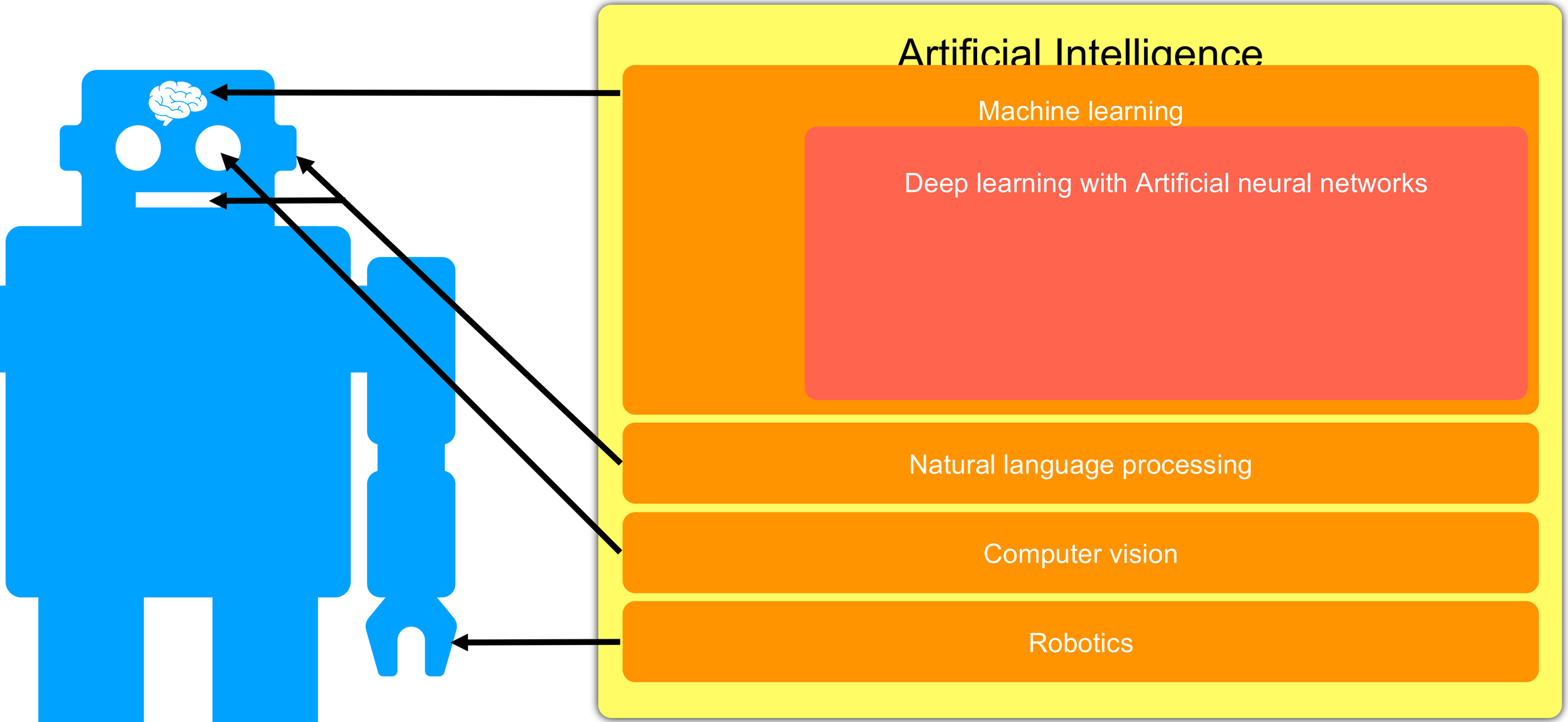
AI Venn diagram



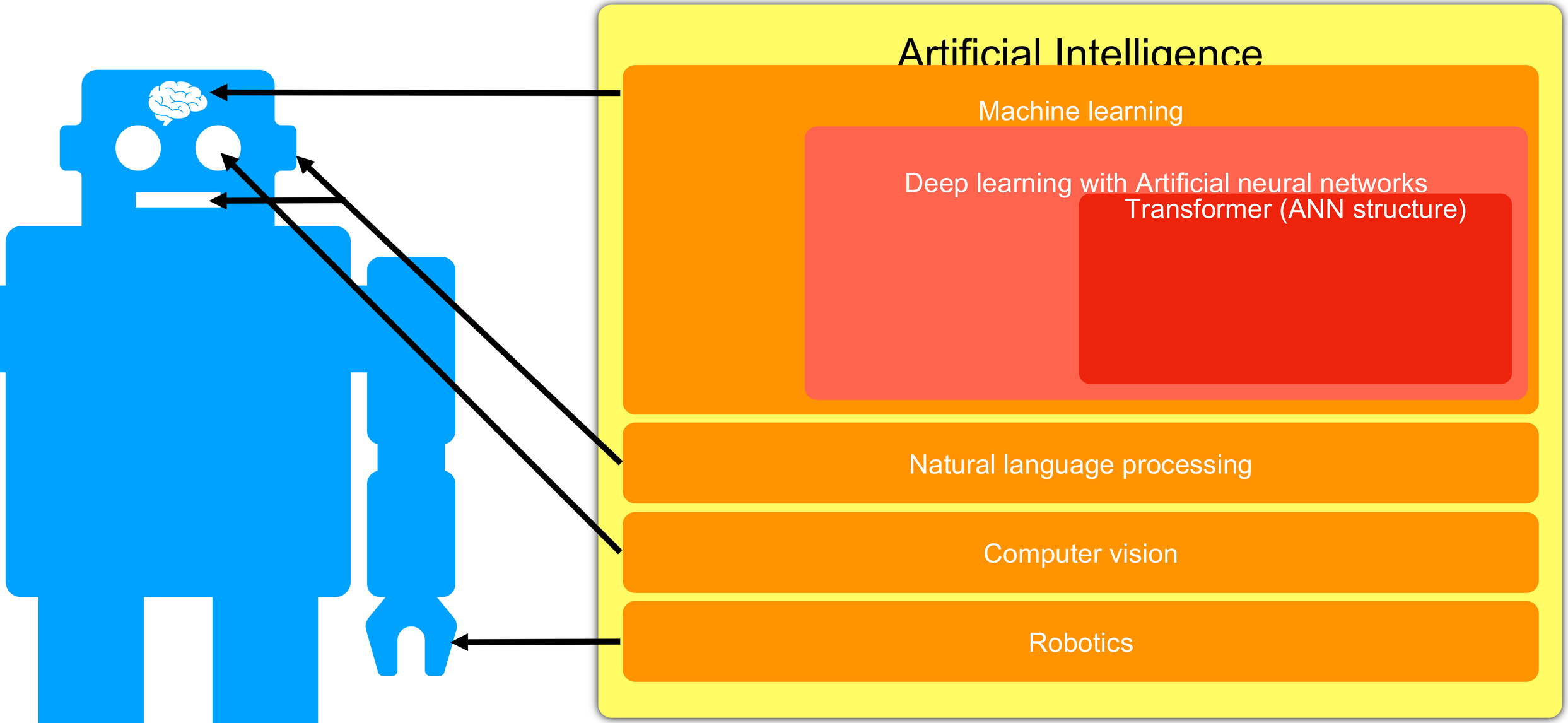
AI Venn diagram



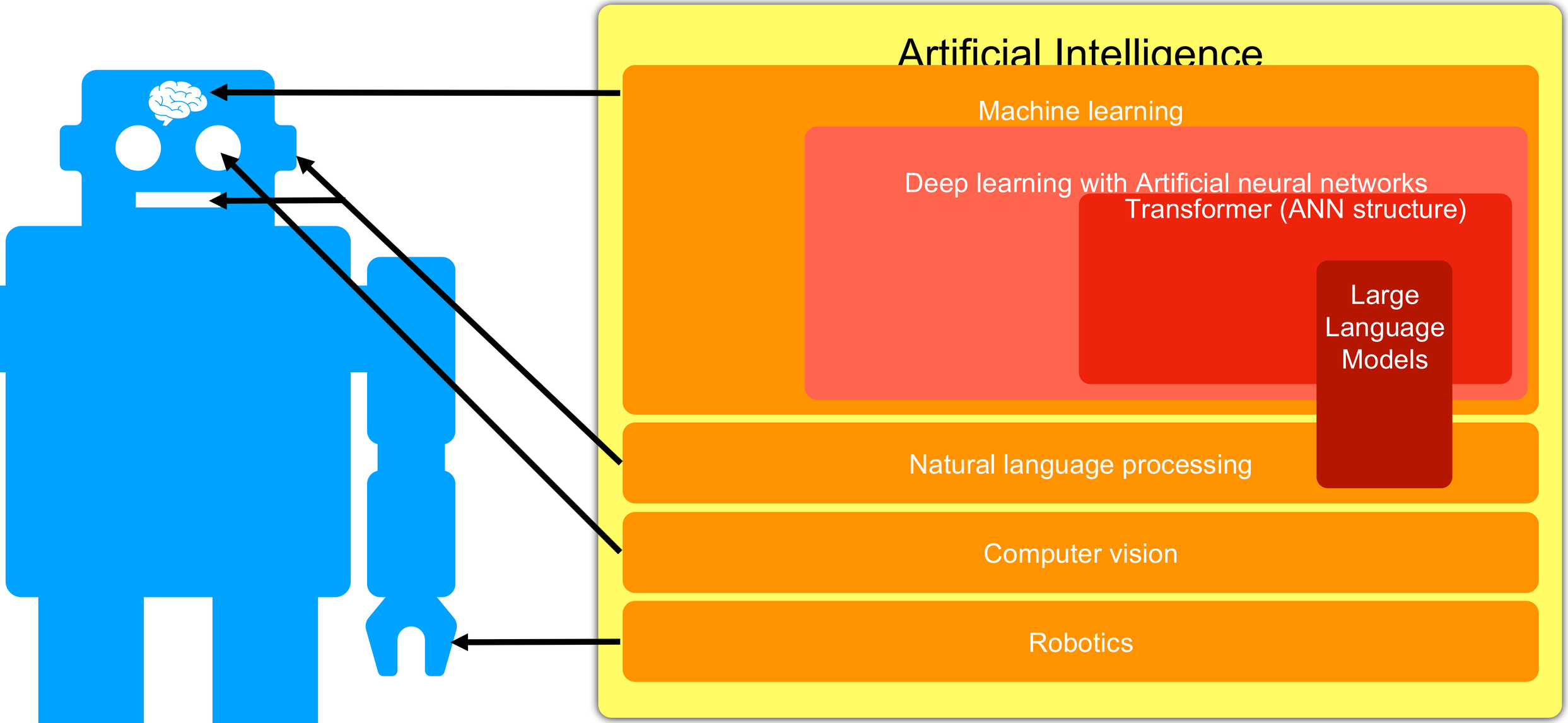
AI Venn diagram



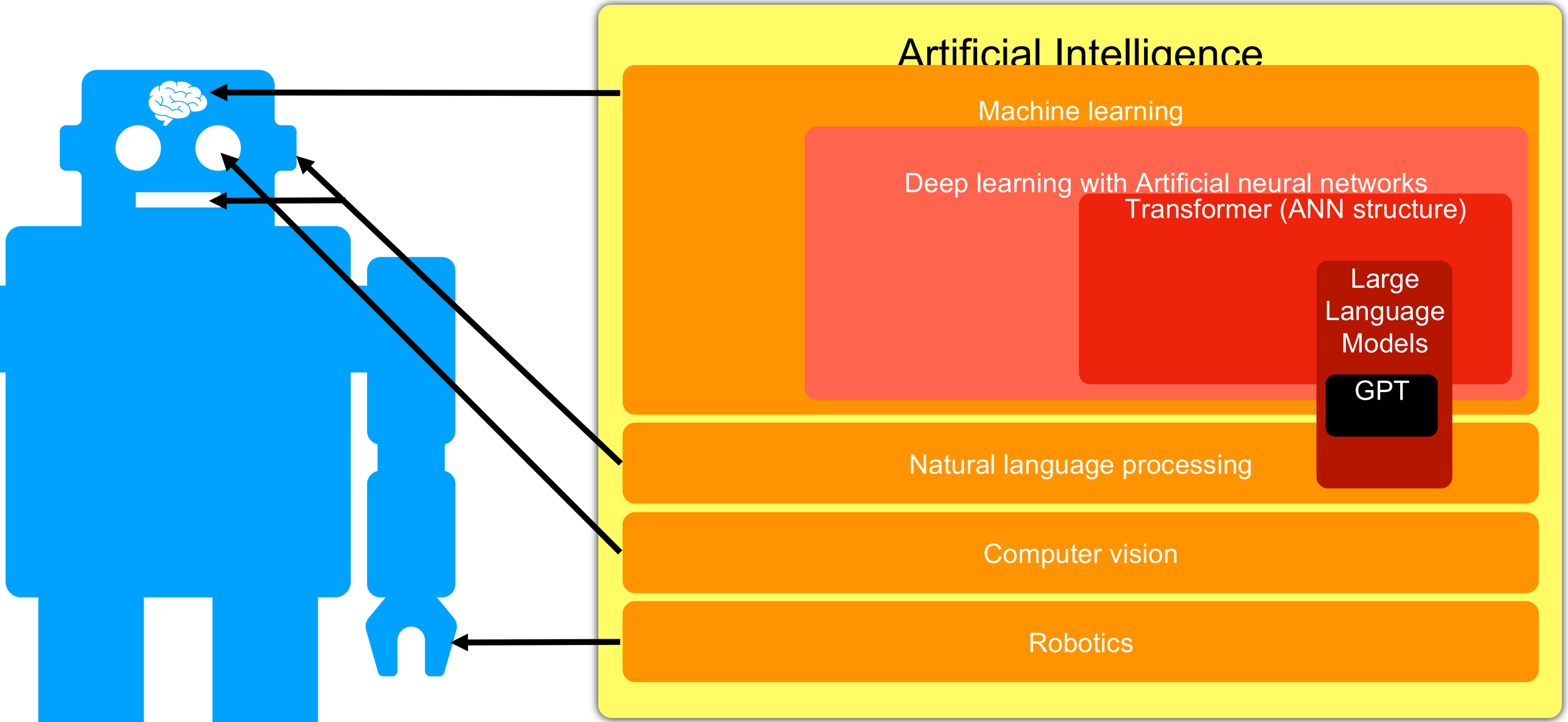
AI Venn diagram



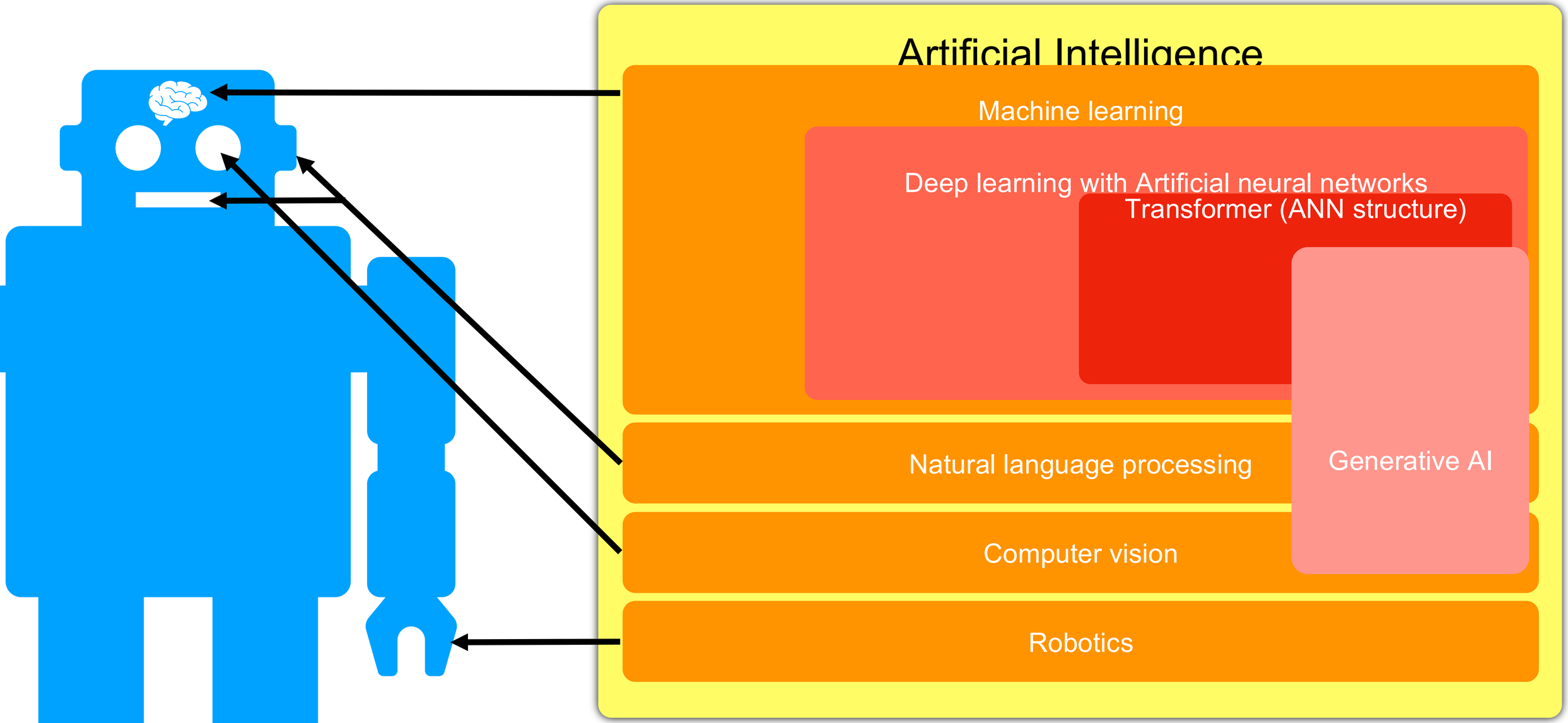
AI Venn diagram



AI Venn diagram



AI Venn diagram



Artificial Intelligence is not Magic

They rely on **fundamental** techniques in:

- Probability and Statistics
- Optimization
- Algorithms
- Logic
- Systems
- …

Part II: Course Logistics

Where to find everything?

- **Moodle**
  - grades, and other private materials *that should not be shared*
  - Slides, syllabus, policies
  - Discussion, questions, announcements

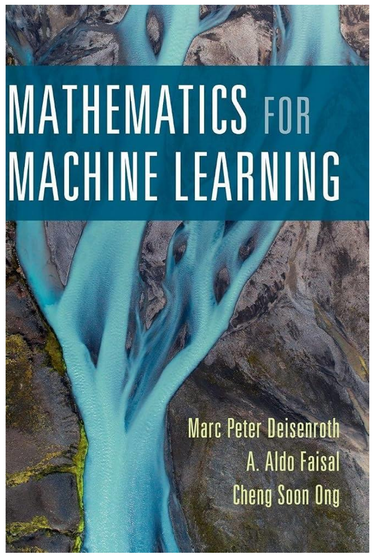
Instruction Team

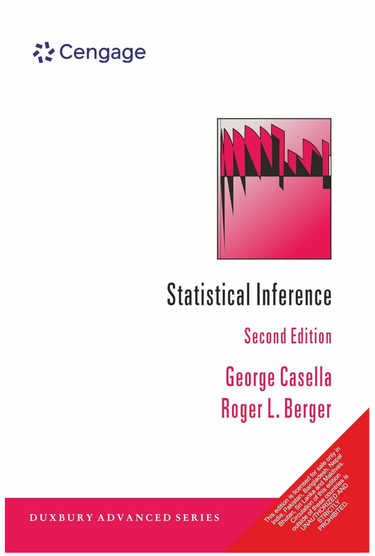
Instructor: Yingyu Liang

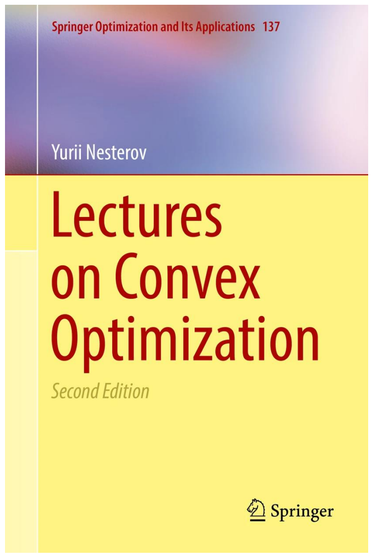
- Email: yingyul@hku.hk
- Office: Room P307, Graduate House

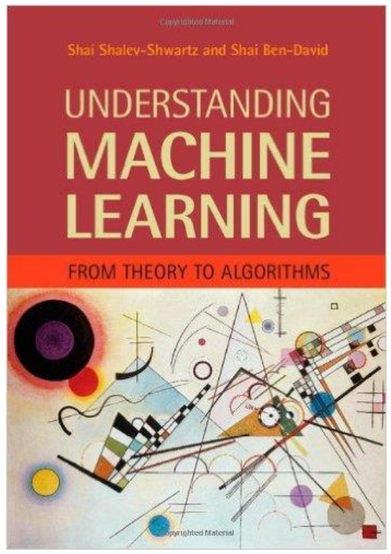
No Textbook

We will have lecture notes.
Useful References **(reference is optional, but may be useful):**









Grading scheme

- Continuous Assessment: 50% (homework + attendance)
  - *Missing two lectures are allowed by default; missing more needs justification*
- Written Examination: 50% (final exam)

- Homework assignments:
  - about 3 assignments
  - Written questions

Thanks!