
# Overview of Deep Learning

Rowel Atienza, PhD
University of the Philippines
github.com/roatienza
2023

# Artificial Intelligence (AI)

## Machines exhibiting animal or human intelligence

Ants bridging algo

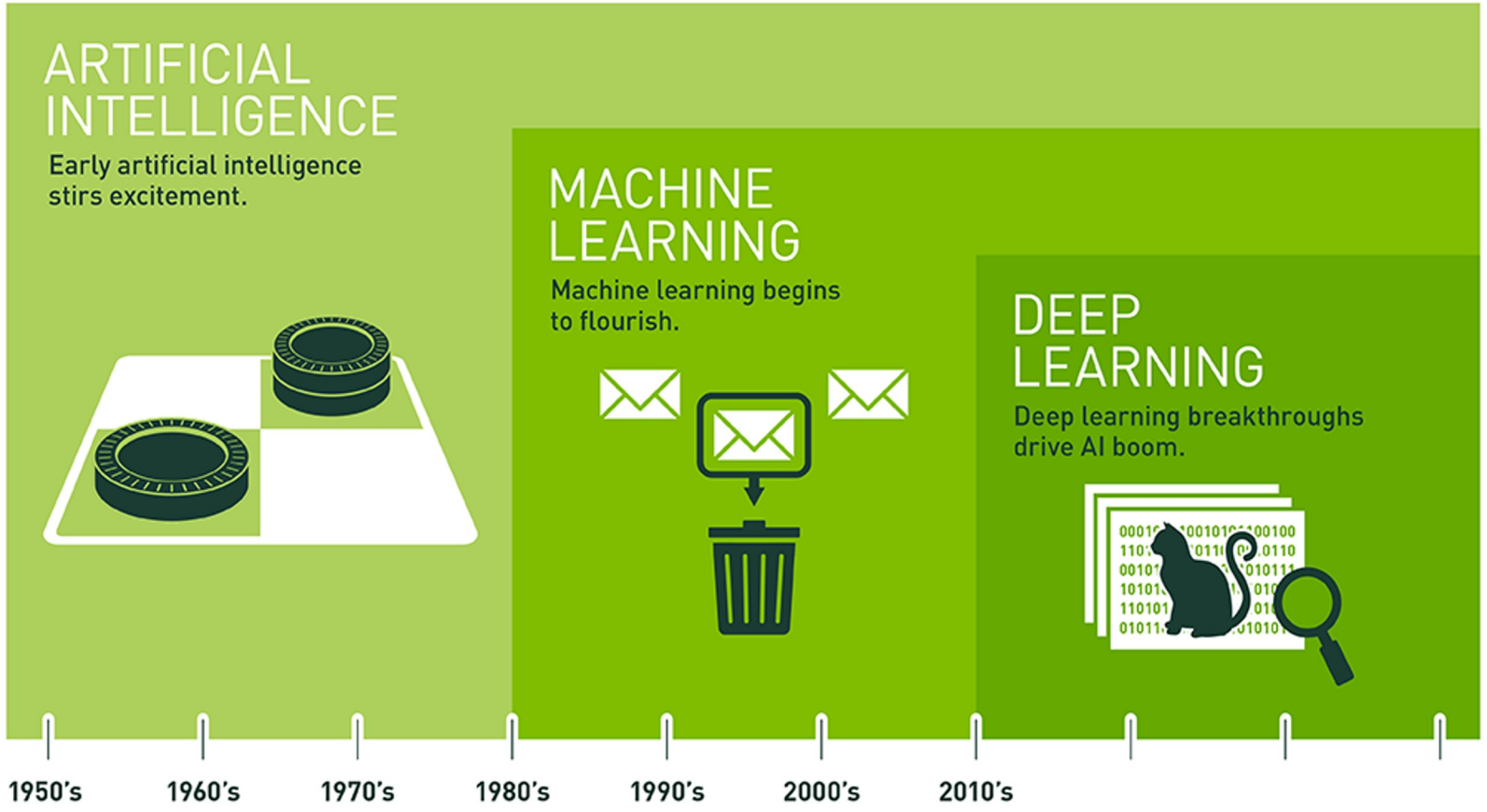
https://www.quantamagazine.org/the-simple-algorithm-that-ants-use-to-build-bridges-20180226

# Intelligence

- A very general mental capability that among other things involves the ability to:
  - Reason
  - Plan
  - Solve problems
  - Think abstractly
  - Comprehend complex ideas
  - Learning quickly and learning from experience

Journal of Intelligence 1997 Vol 24 No 1

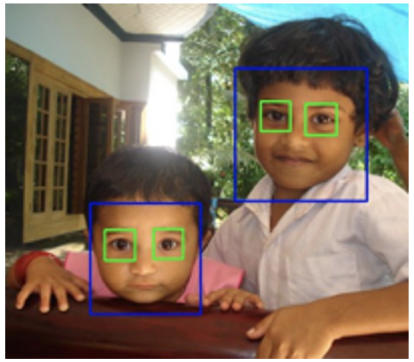
# AI, Machine Learning and Deep Learning




NVIDIA

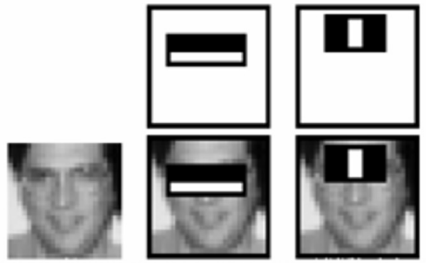
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# Artificial Intelligence, Machine Learning and Deep Learning on Face Detection



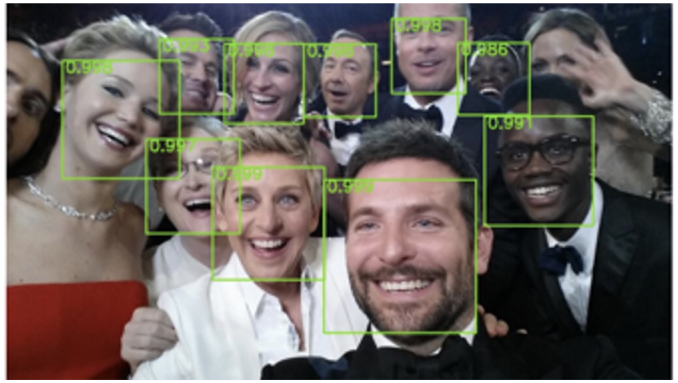
**Rule-based (AI):**
Detect facial features based on color/template
**Apply if-else-if-else**

[Viola & Jones 2001]



**Machine Learning:**
Use Haar Cascade Classifier
**Hand-crafted feature detection**

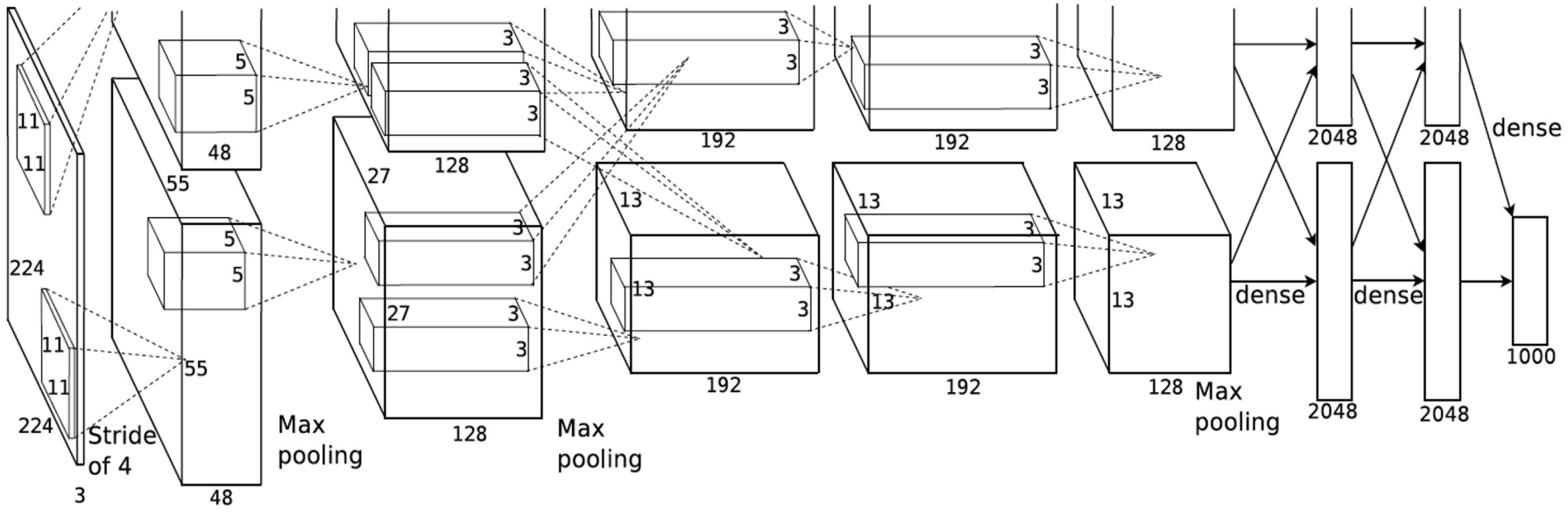
YOLO, SSD, RCNN [>2012]



**Deep Learning:**
Train a network by showing thousands of labelled region of faces
**Automatic feature detection**

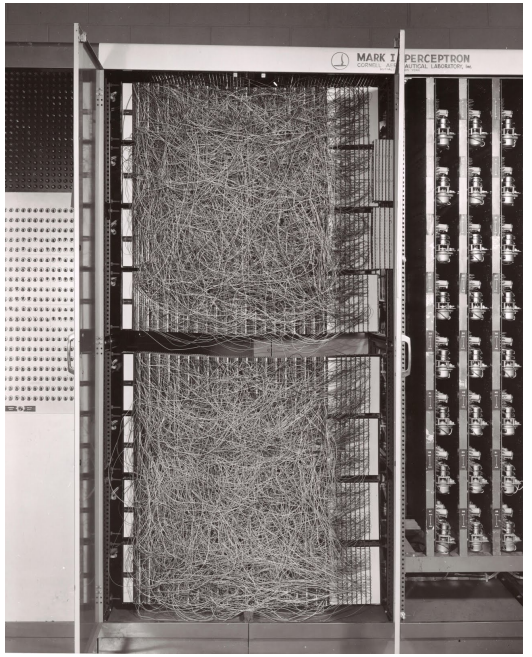
# Deep Learning Moment - 2012

- AlexNet – 650,000-neuron deep neural network won 2012 ImageNet1k competition with 15% Top-5 error rate compared to 2nd place with 26%. It's top 1 accuracy is 63.3%.



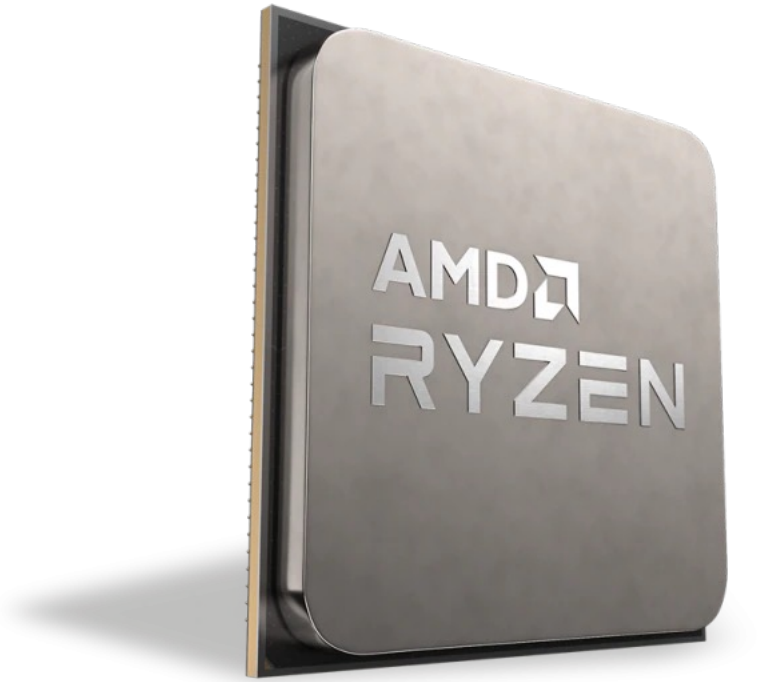
# Deep Learning - what made it work?

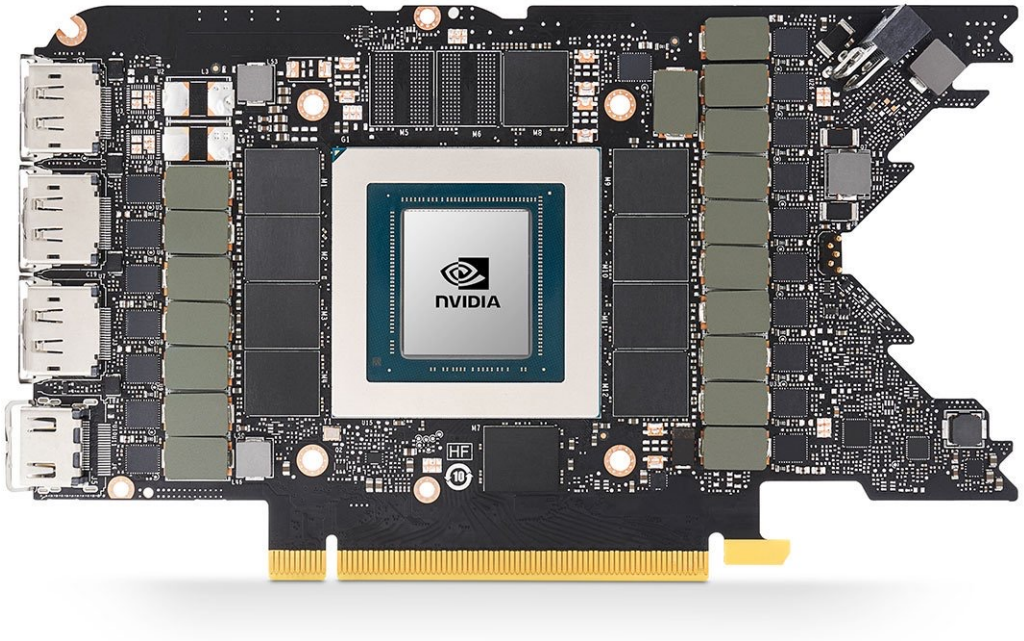
- Concepts of artificial neural network (ANN) and convolutional neural network (CNN) are old
  - Neurons in perceptron (1-layer NN) – 1958
  - Neocognitron (1980) and CNN (1989)
  - Backpropagation (1986)
- What's new?
  - Computing power – Massive number of GPU CUDA cores
  - Data – from the Internet




Perceptron is a binary classifier

# CPU (AMD RyZen) vs GPU (RTX 3090)



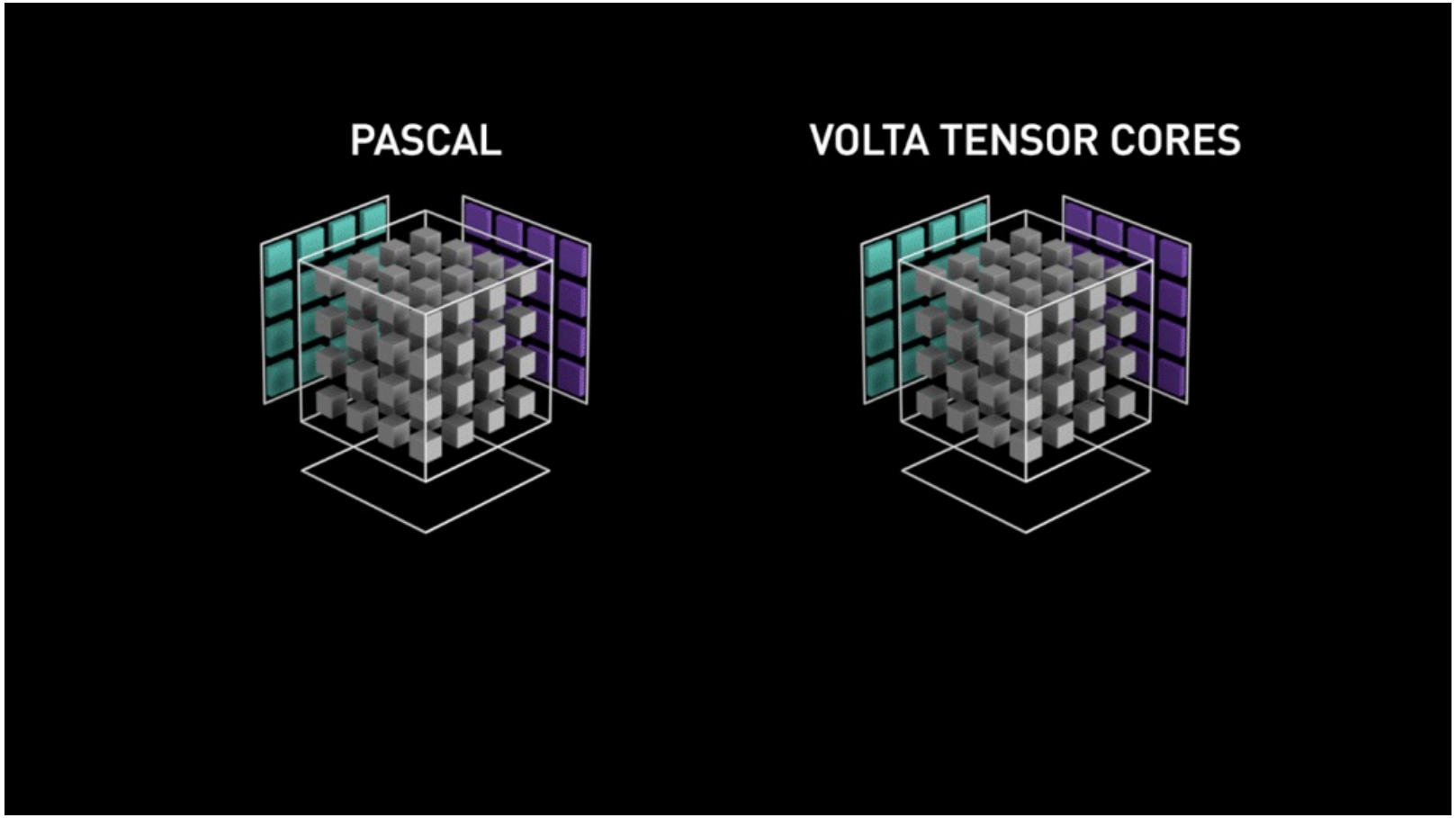

64 3.7GHz super fast cores

6.9 TFLOPS

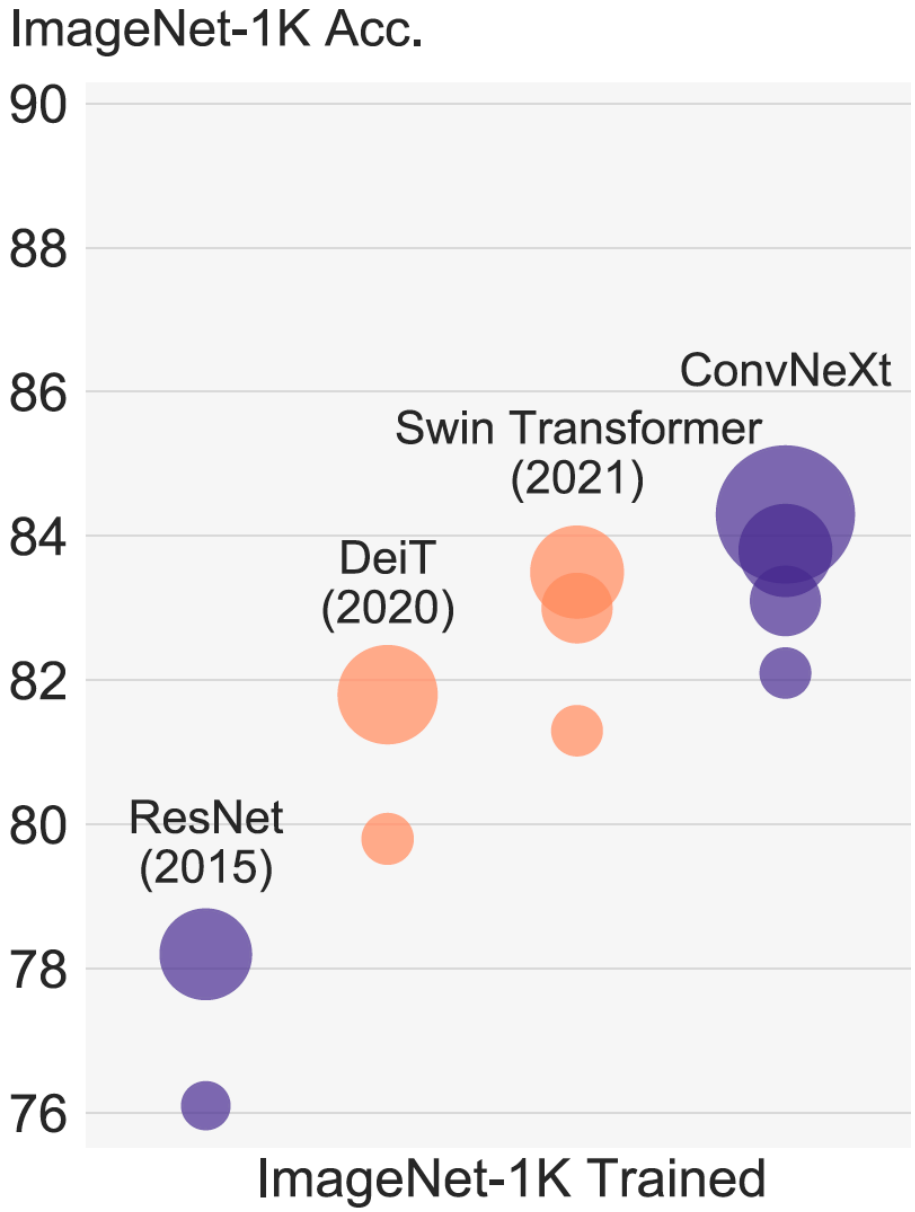
10,496 1.4GHz fast cores

35.6 TFLOPS

# 1D vs 2D Tensor Operation



# The Rest is History



# Barely scratching the surface of Artificial General Intelligence (AGI)

$Imagining$

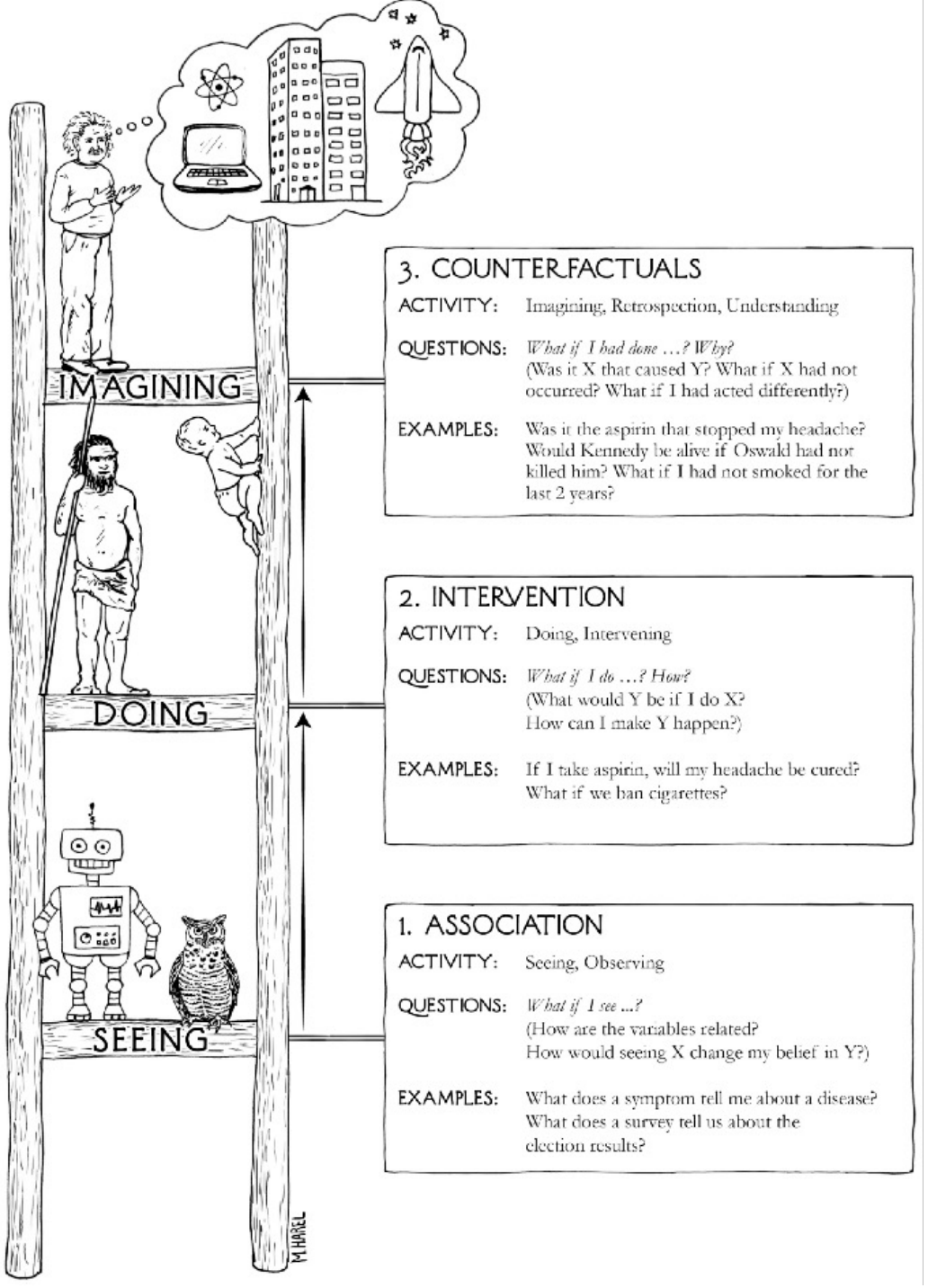
Starting to move here

$Doing: p(\boldsymbol{y}|do(\boldsymbol{x}))$

We are here

$Seeing: p(\boldsymbol{y}|\boldsymbol{x})$

Pearl, Book of Why



# Does ChatGPT (or LLMs in general) exhibit AGI?

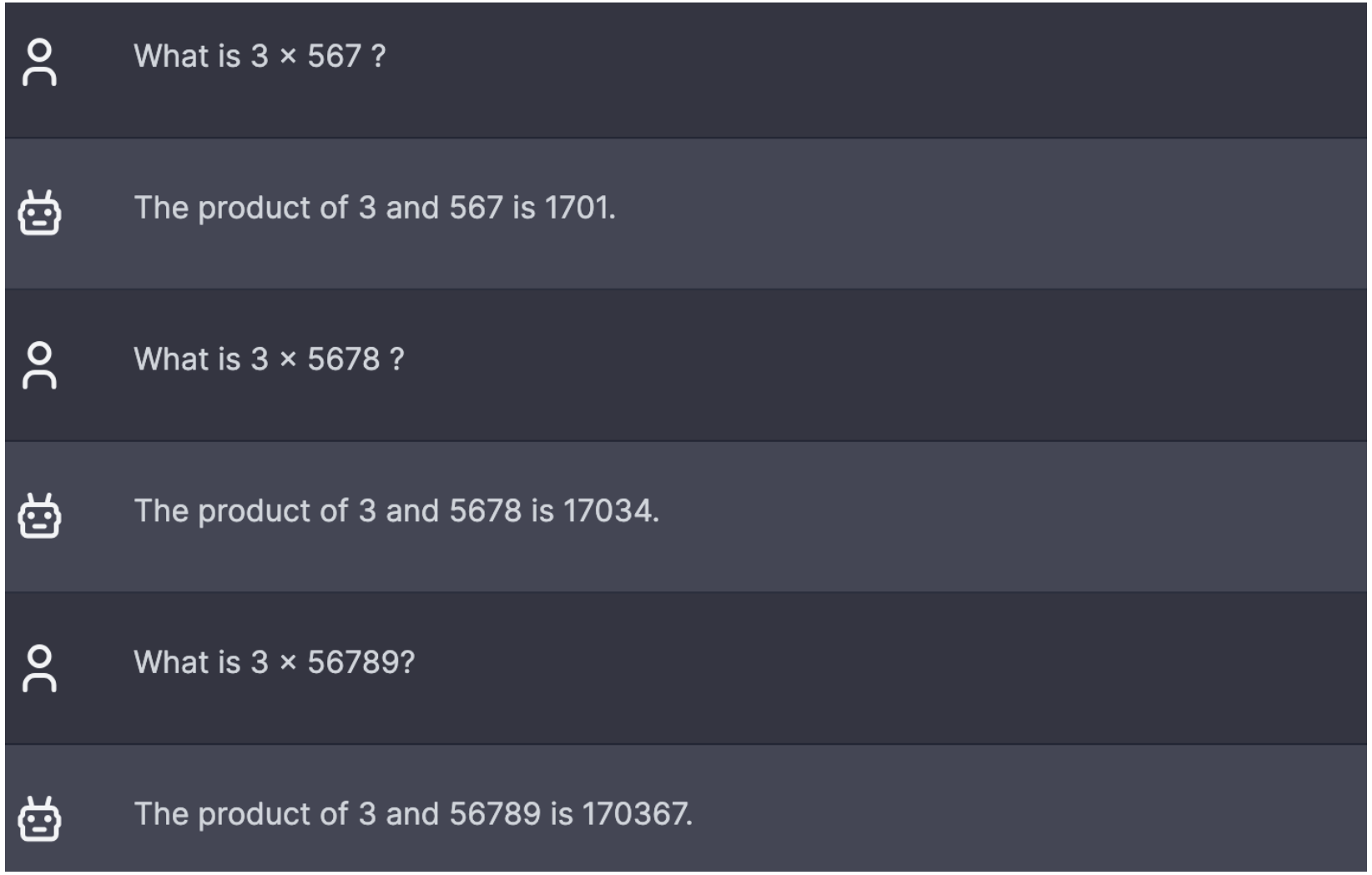
# Do Large Language Models Perform Reasoning in Problem Solving Tasks?

# Comprehensive Tests

## Faith and Fate: Limits of Transformers on Compositionality

**Nouha Dziri[1*], Ximing Lu[1,2*], Melanie Sclar[2*], Xiang Lorraine Li[1†], Liwei Jiang[1,2 †], Bill Yuchen Lin[1], Peter West[1,2], Chandra Bhagavatula[1], Ronan Le Bras[1], Jena D. Hwang[1], Soumya Sanyal[3], Sean Welleck[1,2], Xiang Ren[1,3], Allyson Ettinger[1,4], Zaid Harchaoui[1,2], Yejin Choi[1,2]**

# GPT4 Reasoning Test



What is 345 × 56789?

The product of 345 and 56789 is 19,577,205.

What is 345 / 56789?

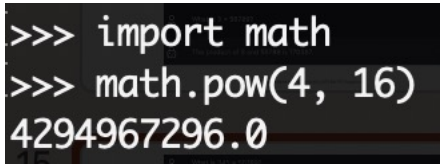
The result of 345 divided by 56789 is approximately 0.006072.

What is 4 raised to its square?

Raising a number to the power of its own square is a sizable operation. In this case, you're asking what is 4^(4^2), or 4^16.

4^16 equals 18,446,744,073,709,551,616.

Failed!



# How Humans use Reasoning to Solve Tasks

- Break a problem into sub-tasks
- A sub-task is a node
- Solving a sub-task transitions it into a new sub-task
- Repeat until all sub-tasks are completely solved

**Computation Graph!**

*Sub-task - Vertex*

*Solution – Edge or Operator*

# Computation Graph for Multiplication Algo

```
function multiply (x[1..p], y[1..q]):
  // multiply x for each y[i]
  for i = q to 1
    carry = 0
    for j = p to 1
      t = x[j] * y[i]
      t += carry
      carry = t // 10
      digits[j] = t mod 10
    summands[i] = digits

  // add partial results (computation not shown)
  product = ∑_{i=1}^{q} summands[q+1-i] · 10^{i-1}
  return product
```

$A(\mathbf{x})$

$A(\mathbf{x})$ for $\mathbf{x} = [7,49]$

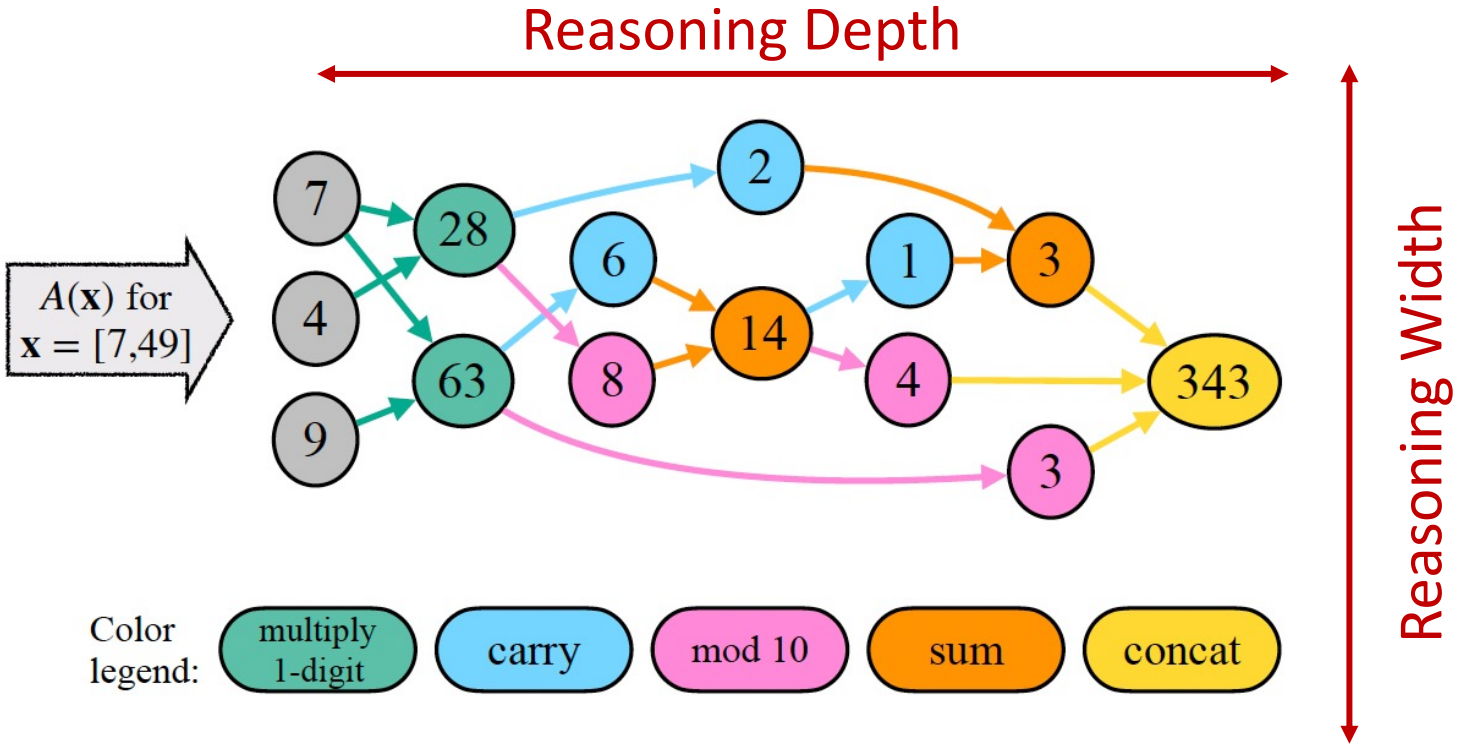
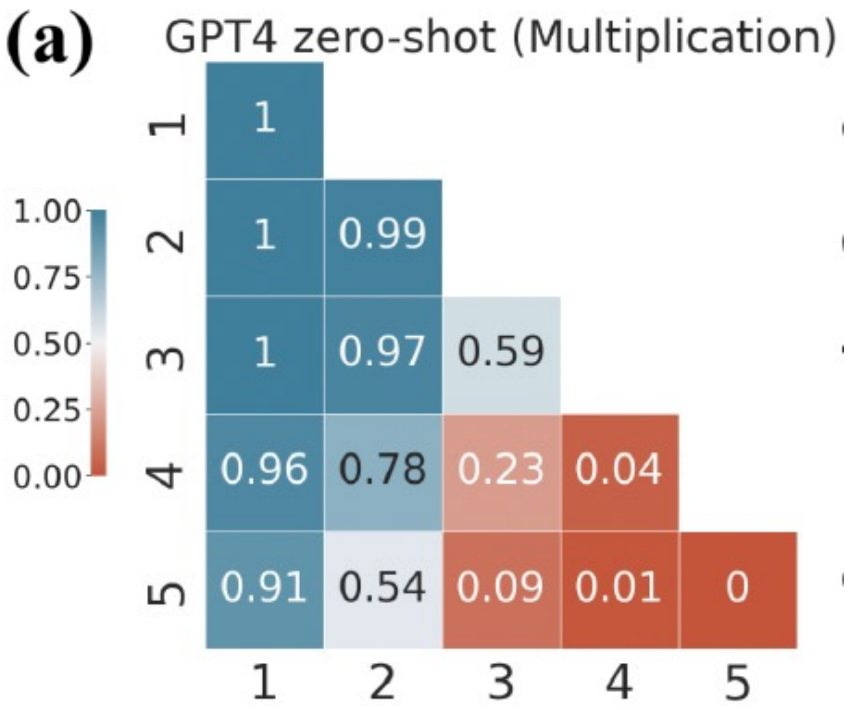



Figure 1: Transformation of an algorithm $A$ to its computational graph $G_{A(\mathbf{x})}$. The depicted example is of long-form multiplication algorithm $A$, for inputs $\mathbf{x} = [7, 49]$ (i.e. computing $7 \times 49$).

# GPT4 Zero-shot Multiplication




*No longer applies as demonstrated in the previous slides!*

# Relative Information Gain (RIG)

$$\text{RelativeIG}(Y_j, X) = \frac{H(Y_j) - H(Y_j|X)}{H(Y_j)} \in [0, 1]$$

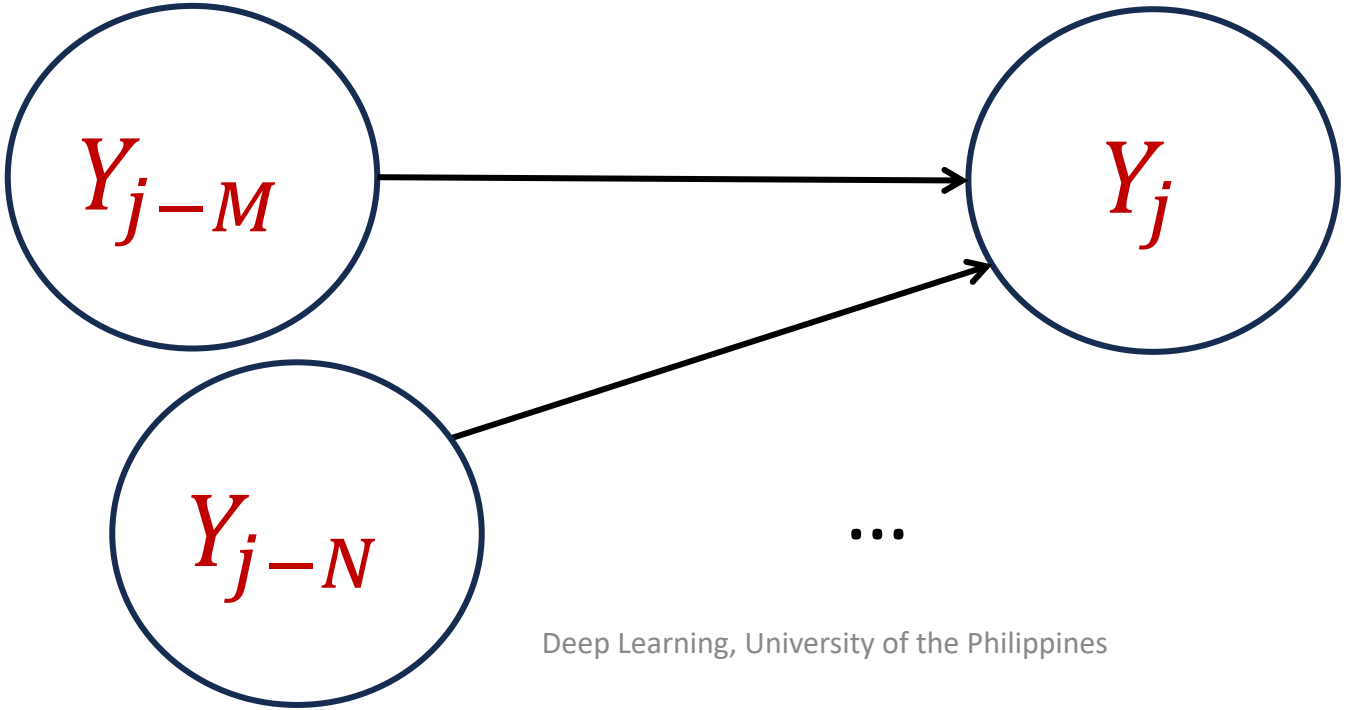
$Y_j$ : output

$X$ : input random variables

$H(Y) = -\mathbb{E}[\log p(Y)]$ : entropy

# Observation

- LLMs break down a problem into a computational graph
- LLMs can solve problems where the RIG is high between sub-tasks
- When the RIG is low, the LLM hallucinates



# 2 Critical Points

- Ability to break down a problem into a correct computation graph
- High RIG between sub-tasks

# When do LLMs fail?

- In-correct computation graph
- A presence of low RIG between sub-tasks
  - Deep reasoning graphs can amplify errors due to error propagation

# Incorrect computation graphs

- Instruction-based tuning
- Prompt engineering
- Etc

**InstructBLIP: Towards General-purpose Vision-Language Models with Instruction Tuning**

LARGE LANGUAGE MODELS AS OPTIMIZERS

# Low RIG between sub-tasks

- Higher data quality

**TinyStories**: How Small Can Language Models Be and Still Speak Coherent English?

---

**LIMA: Less Is More for Alignment**

---

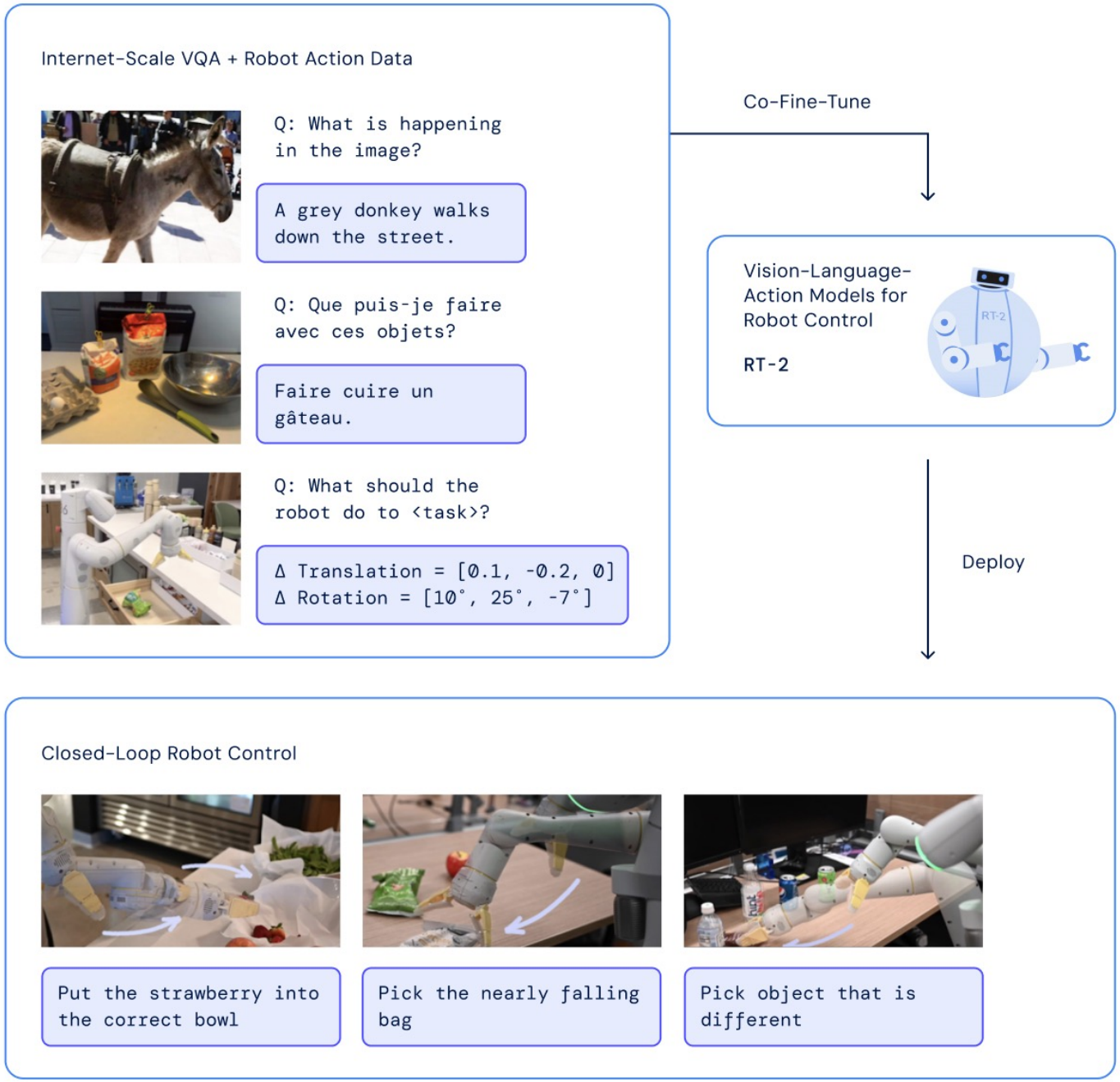
# Improving Reasoning by Grounding the Language Modality

# RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control

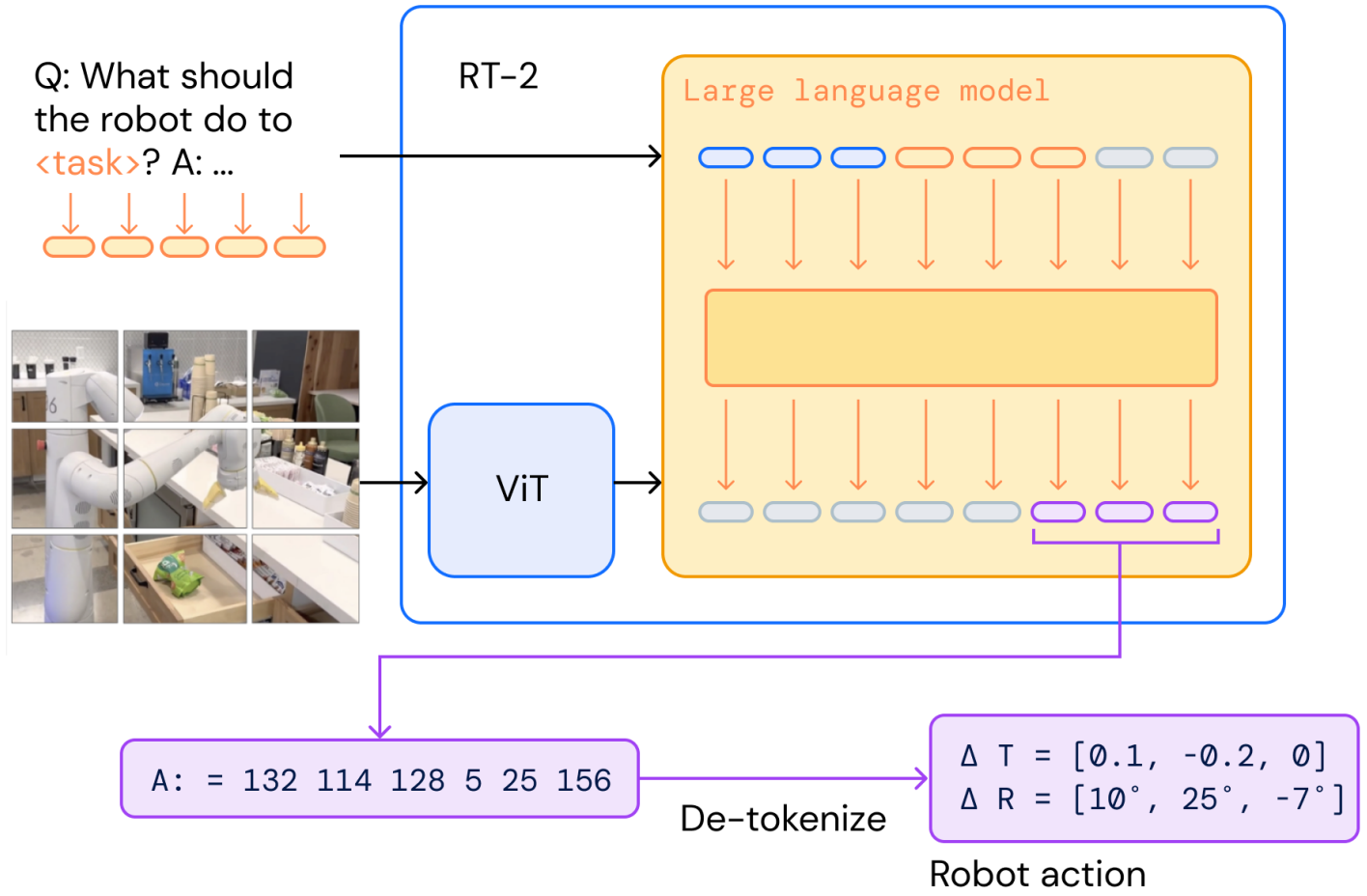
https://robotics-transformer2.github.io/

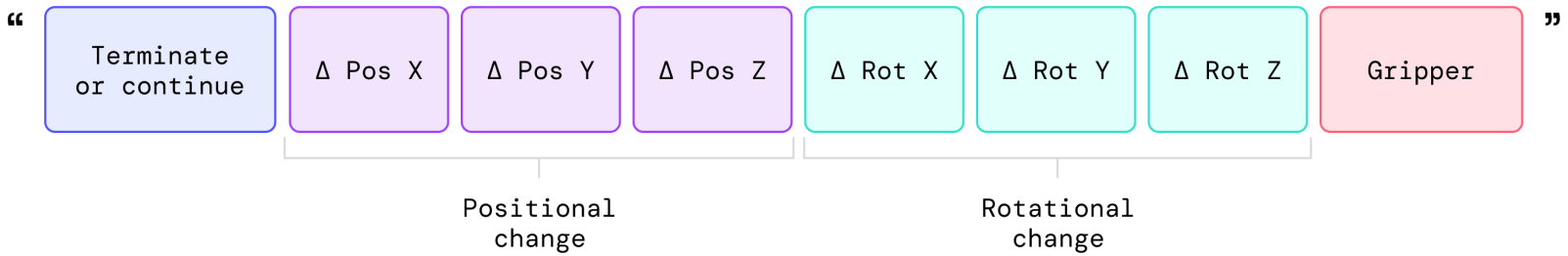

VLM

# Pre-train on Internet scale vision-language data



Co-fine-tune with real robot data.
The robot language is used during co-fine-tuning.




Coca-Cola

# LINGO-1: Exploring Natural Language for Autonomous Driving

https://wayve.ai/thinking/lingo-natural-language-autonomous-driving/

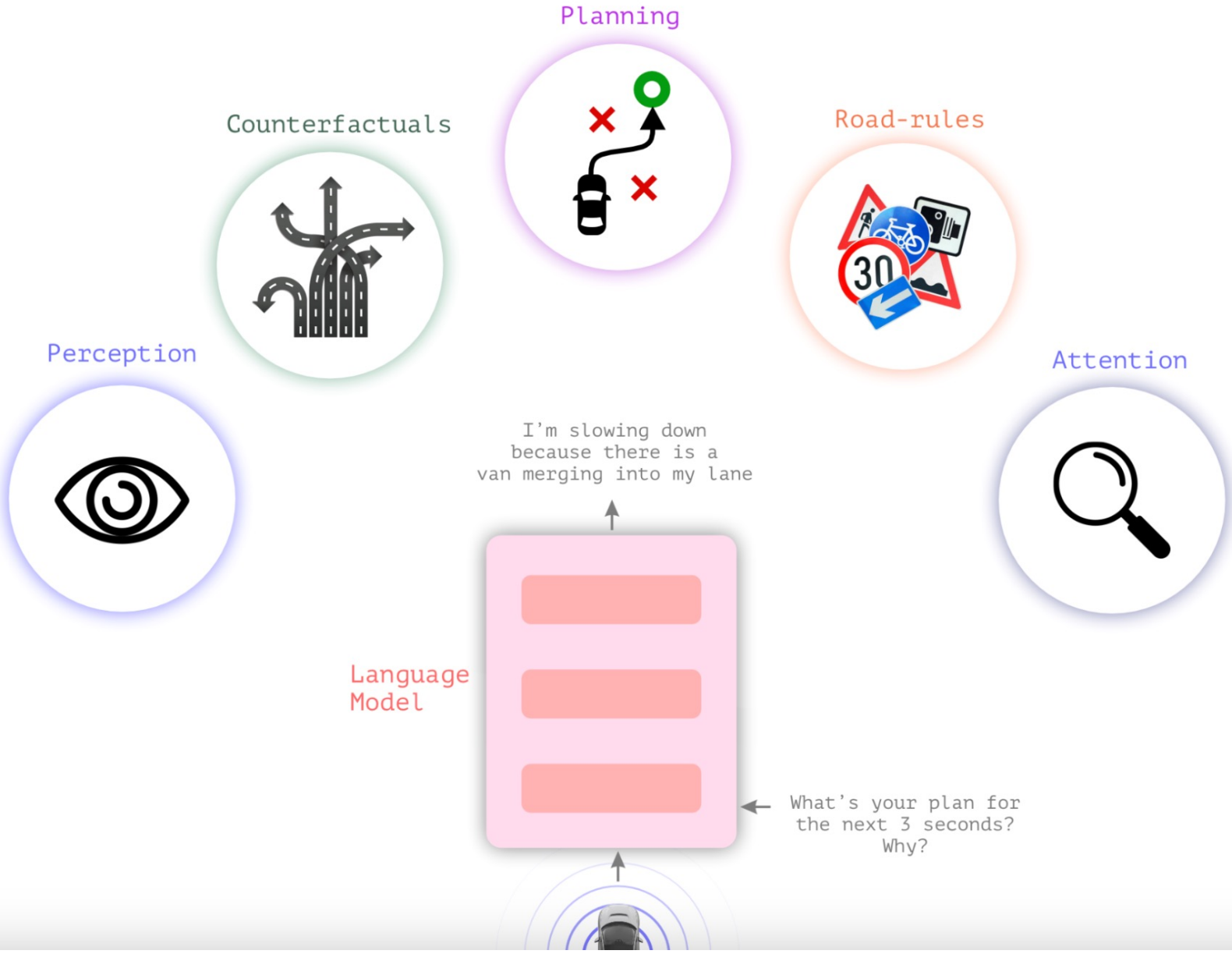
WAYVE

Technology Partners Safety Careers Company

## AV2.0

Bringing the next wave of self-driving technology to market.

# LINGO-1 Architecture



# Lingo-1: Driver Commentator for Autonomous Driving

- slowing down for a lead vehicle or a change in traffic lights,
- changing lanes to follow a route,
- accelerating to the speed limit,
- noticing other cars coming onto the road or stopped at an intersection
- approaching hazards such as roundabouts and Give Way signs,
- parked cars, traffic lights or schools,
- actions other road users are taking, such as changing lanes or overtaking parked vehicles,
- cyclists and pedestrians waiting at zebra crossings or coming up from behind the car in a cycle lane.

WAYVE

Select another segment


**London NW Route**

Ready

Close

What do you see?

...

Ask a new question

# LINGO-1 Contributions

- Improved Reasoning due to instructions from driver commentator
- Querying next actions improve model explainability

# References


Strickland, E. The Turbulent Past and Uncertain Future of Artificial Intelligence Is there a way out of AI's boom-and-bust cycle? https://spectrum-ieee-org.cdn.ampproject.org/c/s/spectrum.ieee.org/amp/history-of-ai-2655064200

Liu, Zhuang, et al. "A ConvNet for the 2020s." arXiv preprint arXiv:2201.03545 (2022).

NVIDIA.com

AMD.com

Dziri, Nouha, et al. "Faith and Fate: Limits of Transformers on Compositionality." *arXiv preprint arXiv:2305.18654* (2023).

# End