Large Language Models, ChatGPT and University Education

D. Psarras, Z. Sotireli, I. Valsamara, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 4.2

Large Language Models, ChatGPT and University Education

Large Language Models, ChatGPT and University Education

- Introduction
- Large Language Models
- GPT
- Capabilities
- Reasoning
- Limitations
- ChatGPT and University Education

Introduction

- ChatGPT is a ***Large Language Model*** (***LLM***) that is fine-tuned from a Generative Pre-Trained Transformer-3.5 (GPT-3.5) series LLM, produced by OpenAI.
- An LLM is a ***Deep Neural Network*** (***DNN***) trained to generate smooth text similar to the human-generated one.
- The fine-tuning of the GPT-3.5 is performed using supervised and reinforcement learning with human feedback [OPE2023] .

Large Language Models

The building blocks of LLMs are [AJI2023] :

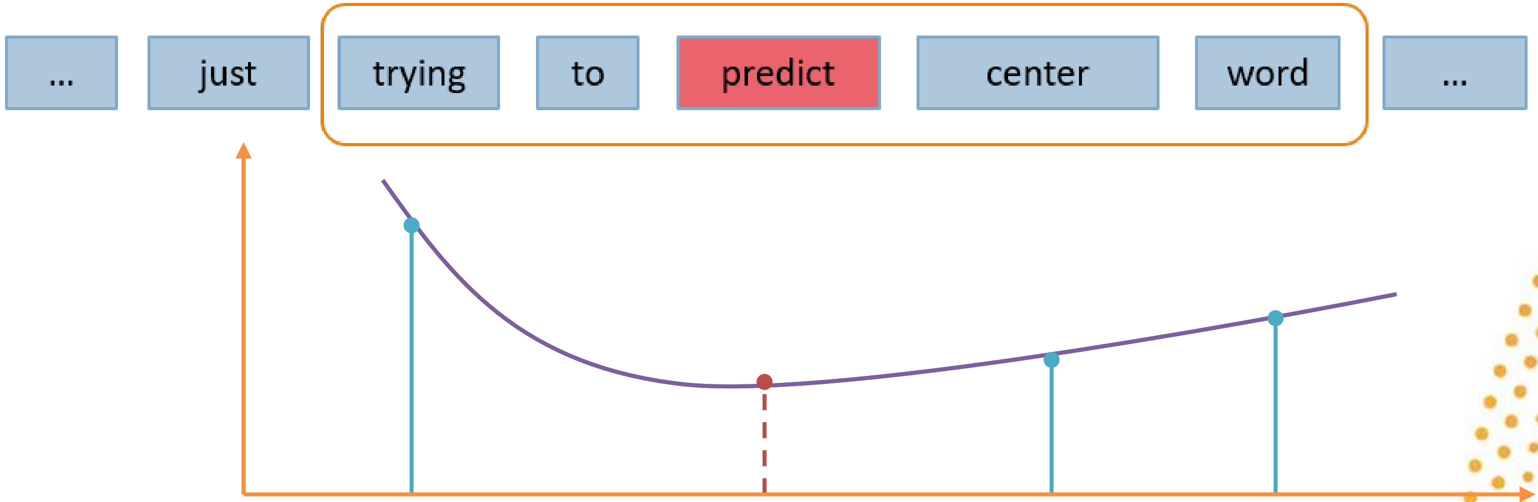
- ***Tokenization***: transforming a text in a series of tokens, e.g.,:
  - ***sub-words, words***.
- Text compression, in order to minimize the size of the encoded token, while retaining the ability to represent well text sequences.
- ***Vector embedding***: Token representation by vectors capturing their semantic meaning in a high-dimensional space.
- Vector embeddings are processed by the NN and are learned during the training.

Large Language Models

Word embeddings: ***Word2Vec***

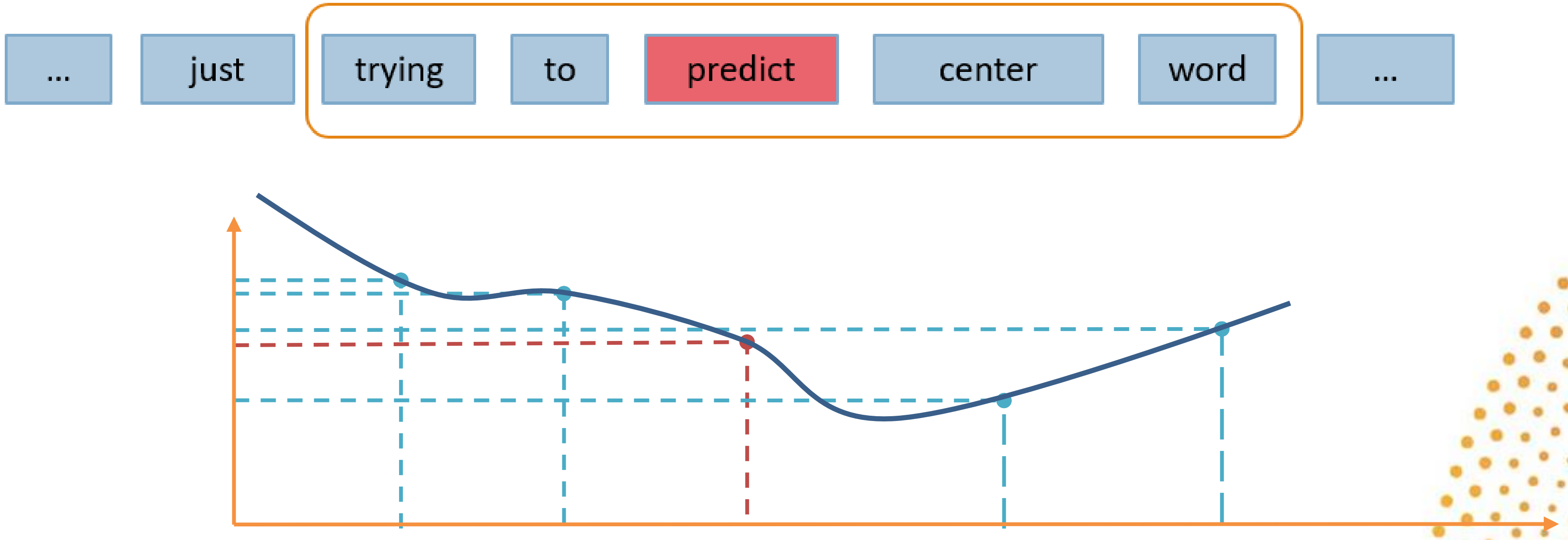
Two-layer NN trained to reconstruct linguistic context of words.

- Training is performed with pairs of context-target words.
- 2 training variations.



Word embeddings

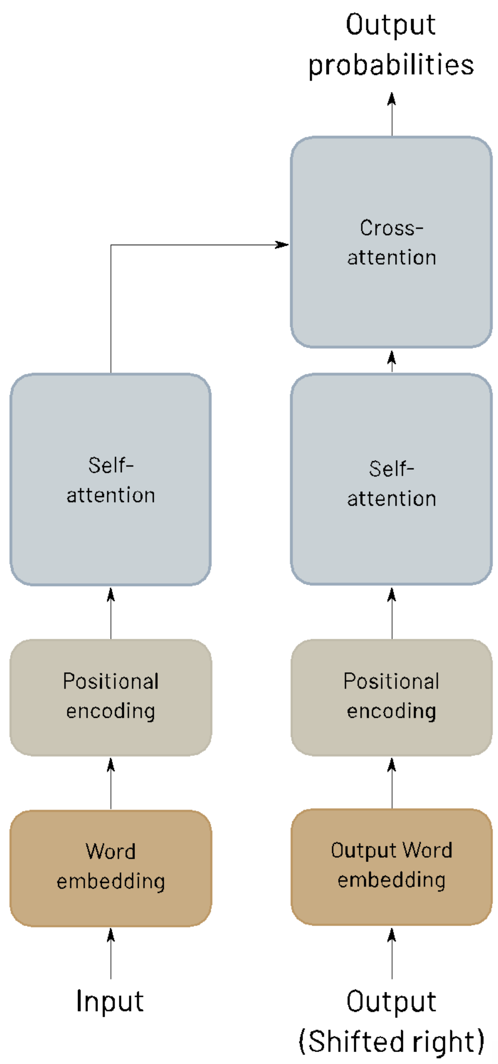
- ***Visualization of word prediction in 2D space***
- A sentence can be visualized as a curve in the vectorial space over time, connecting all its word embeddings.



Large Language Models

Transformers provide data representations based on statistical correlations of input elements (NLP tokens).

- They comprise of the ***encoder*** and ***decoder***.
- ***Self-attention*** weighs the importance of input or output tokens.
- ***Cross-attention*** cross-correlates input and output tokens.

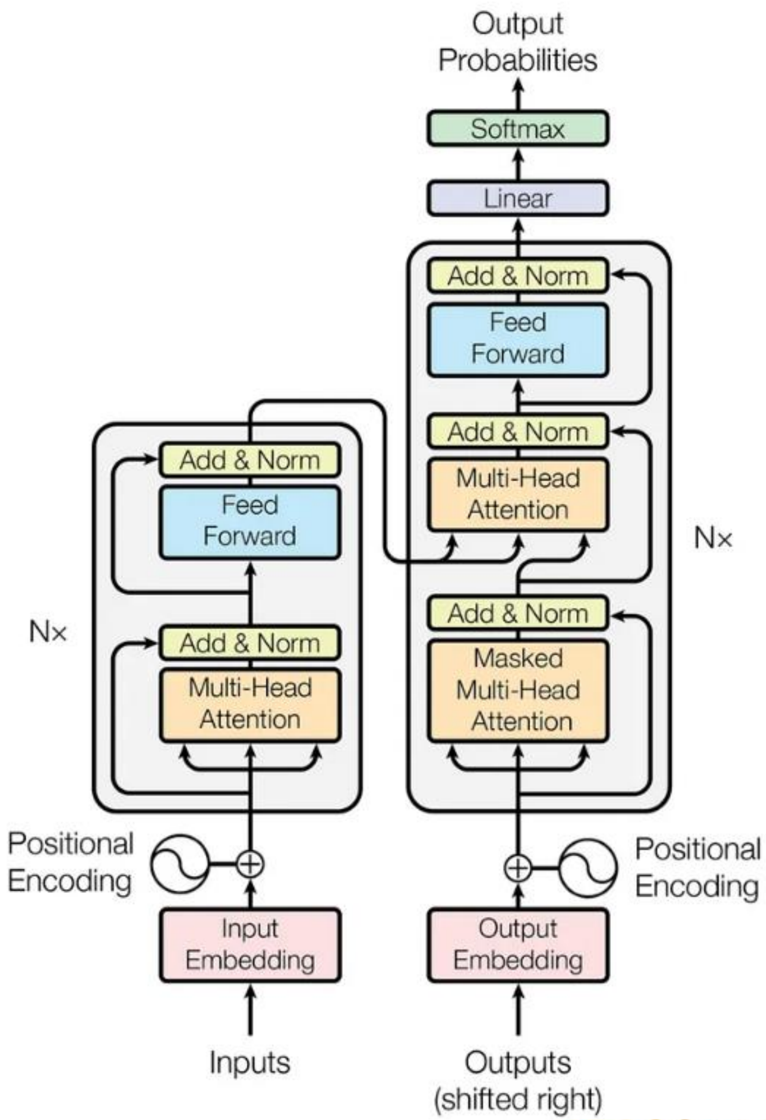


Transformer architecture.

Large Language Models

Transformers

- ***Transformers*** comprise of the encoder and decoder and use the self-attention mechanism to weigh the importance of input elements [VAS2017] .
- GPT-3.5 is a fine-tuned model of the GPT-3, which is a Transformer DNN.



Transformer architecture [VAS2017].

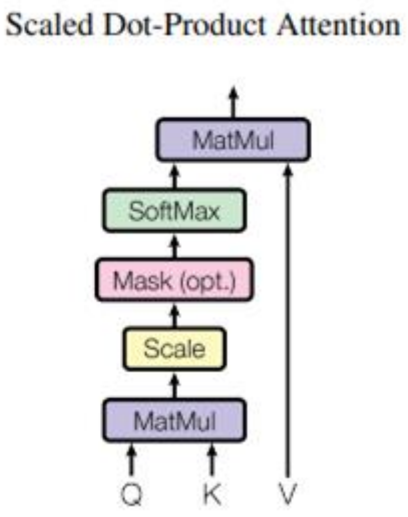
Large Language Models

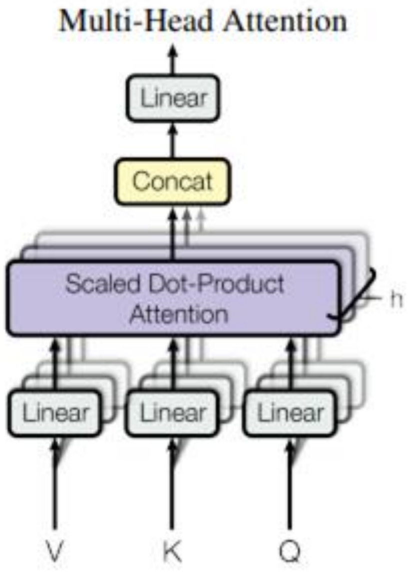
Attention module has three inputs:

- Keys $\mathbf{K}$, Values $\mathbf{V}$ and Queries $\mathbf{Q}$.
- The scaled dot-Product attention the output matrix:

$$\mathbf{A} = \sigma\left(\frac{\mathbf{QK}^T}{\sqrt{n}}\right)\mathbf{V}.$$

- Multi-head attention is applied multiple times in parallel on linearly projected





Top: Scaled Dot-Product attention. Bottom: Multi-Head attention [VAS2017].

Large Language Models

Multi-head Attention:

- Counteracts the reduced effective resolution due to averaging attention-weighted positions in single attention.
- Multi-head attention provides multiple low-scale featured map compared to a single map obtained by single attention.
- Multiple attention head are analogous to multiple kernels in a single layer in a CNN.

Large Language Models

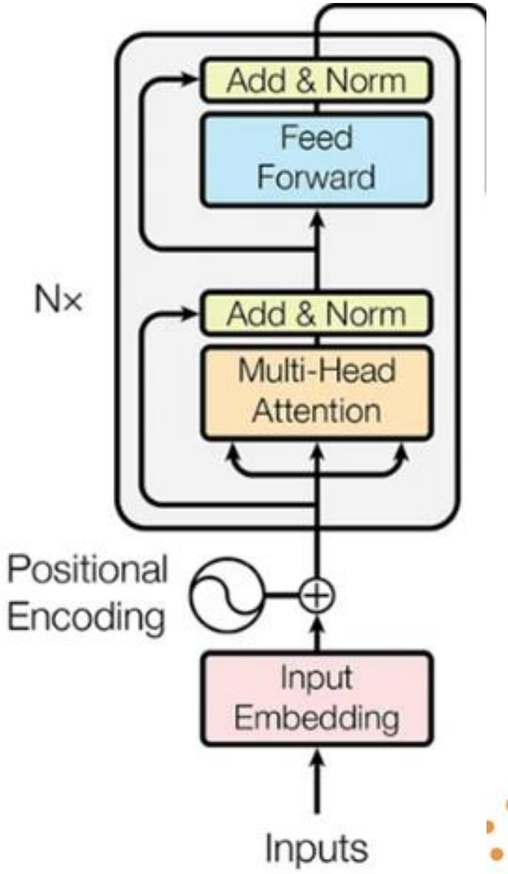
Multi-head Attention:

- Jointly attend information from different representation subspaces at different positions capturing richer interpretations (various patterns and dependencies).
- Redundancy is introduced making the model more resilient to noise or errors in individual heads (robustness).

Large Language Models

The ***transformer encoder*** processes the input sequence using two sub-layers [VAS2017]:

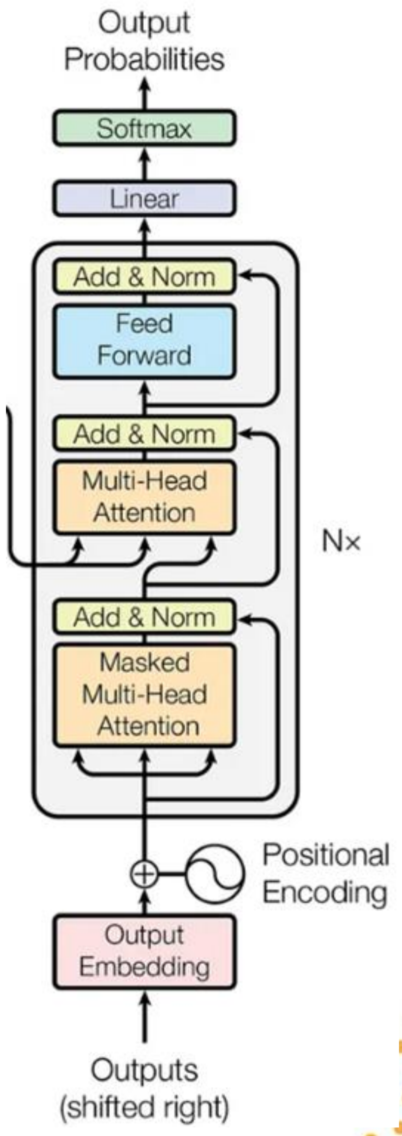
- ***Multi-head self-attention mechanism***: It attends to different parts of the sequence in parallel, inferring meaning and context.
- ***Position-wise fully connected feed-forward network***: Two linear transformations with a RELU activation in between applied to each position independently [VAS2017].



Transformer encoder [VAS2017].

Large Language Models

- The ***transformer decoder*** has an extra multi-head ***cross-attention*** sub-layer of attention, between the two sub-layers of the encoder layer.
- It outputs the probability of each vocabulary token.
- In the multi-head ***cross-attention*** sub-layer the key-value pairs $\mathbf{K}, \mathbf{V}$, are



Transformer decoder.

Large Language Models

Encoder-Decoder stacks:

- Multiple stacked encoder and decoder blocks build hierarchical representations to capture high level features and dependencies.
- Stacked blocks are analogous to increasing the depth of a CNN.

Large Language Models

LLM training and text production:

- LLMs search for text patterns and correlations in huge amounts of training data and produce statistically probable output (text).
- Essentially, they become increasingly better in learning word predictions and relations.
- This is an essential feature in outputting smooth ‘human-like’ text.

Large Language Models

LLM training and text production example:

- LLMs' reply to the query 'What is the capital of Spain?' would be 'Madrid' rather than 'death penalty', since:
- a) they encountered this semantic association (Spain, Madrid, capital) too many times in their training corpora.
- b) the learned association (Spain, country) helps them disambiguate the meaning of the query word 'capital'.
- Such statistical associations may occasionally be out of context, or semantically wrong or completely fabricated.

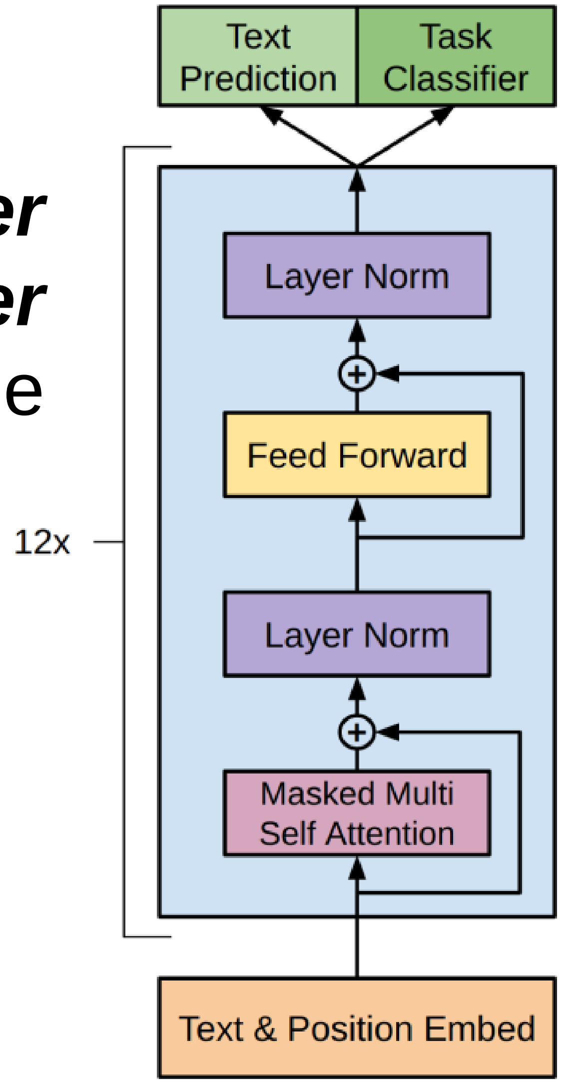
GPT

- The ***Generative Pre-Trained Transformer*** (***GPT***) is a ***decoder-only Transformer*** model that generates one token at a time [RAD2018].

- Semi-supervised training:

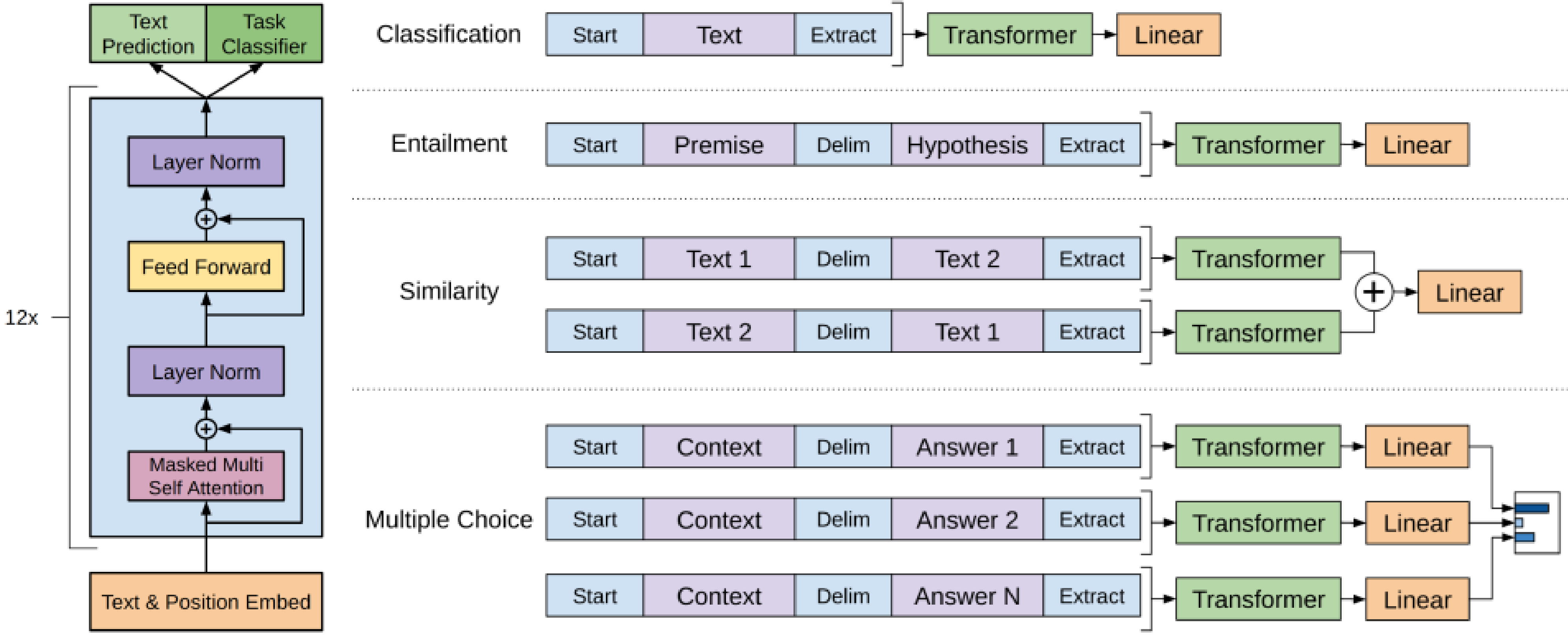
a) Unsupervised pre-training.

b) Supervised fine-tuning.



GPT architecture [RAD2018].

GPT



GPT architecture of GPT model (left). Input for GPT fine-tuning to perform various tasks (right) [RAD2018].

GPT Training stages

Unsupervised Pre-training stage:

- *Training dataset*: BooksCorpus [ZHU2015].
- *Objective*: Standard language modelling [RAD2018].

Fine-tuning stage:

- *Training dataset*: a labelled dataset corresponding to the fine-tuning task
- *Objective*: GPT model parameters adaptation to the supervised target task and language modelling [RAD2018].

GPT-2

- GPT-2 is larger than the first GPT model [RAD2019]:
  - GPT: 117 million parameters, GPT-2: 1.5 billion parameters.

- GPT-2 employs ***zero-shot learning***.
- Special tasks (text translation, question answering, etc.) can be framed in the same way as language modelling.
- WebText training dataset (internal OpenAI corpus) was used with emphasis on document quality.

GPT-2

- ***Zero-shot learning***: GPT model input is: a) a task description b) prompt.
- Example: *Translate English to French* (task description), *cheese* (prompt).

- ***One-shot learning***: GPT model input is: a) task description and b) a single task example (from the training dataset).
- ***Few-shot learning***: GPT model input is: a) task description and b) few task examples (from the training dataset).

GPT-2 Results

Zero-shot learning accuracy (ACC) on different datasets without further training or fine-tuning [RAD2019].

	LAMBADA (PPL)	LAMBADA (ACC)	CBT-CN (ACC)	CBT-NE (ACC)	WikiText2 (PPL)	PTB (PPL)	enwik8 (BPB)	text8 (BPC)	WikiText103 (PPL)	1BW (PPL)
SOTA	99.8	59.23	85.7	82.3	39.14	46.54	0.99	1.08	18.3	**21.8**
117M	**35.13**	45.99	**87.65**	**83.4**	**29.41**	65.85	1.16	1.17	37.50	75.20
345M	**15.60**	55.48	**92.35**	**87.1**	**22.76**	47.33	1.01	**1.06**	26.37	55.72
762M	**10.87**	**60.12**	**93.45**	**88.0**	**19.93**	**40.31**	**0.97**	**1.02**	22.05	44.575
1542M	**8.63**	**63.24**	**93.30**	**89.05**	**18.34**	**35.76**	**0.93**	**0.98**	**17.48**	42.16

GPT-3

- GPT-3 is a Transformer DNN with the same design and architecture as GPT-2 [BRO2020].
- It is an autoregressive model generating a continuation of an input sequence of tokens.
- GPT-3 has 10 times bigger parameter set compared to GPT-2:
  - 175 billion parameters, 96 attention layers.
  - Each layer has 96 heads. Each head is 128-dimensional.

- GPT-3 is trained on batches of size 3.2 million token.

GPT-3

- After the pretraining stage GPT-3 (in contrast to GPT-2), applies in-context learning to address fine-tuning issues [BRO2020]:
  - E.g., too many required data, overfitting.

- Training dataset comprises:
  - Common Crawl (filtered) [COM2023], WebText2 [ALE2019].
  - Books1, Books2 [JAR2020], Wikipedia [BRO2020].

GPT-4

- GPT-4 is a large multimodal model

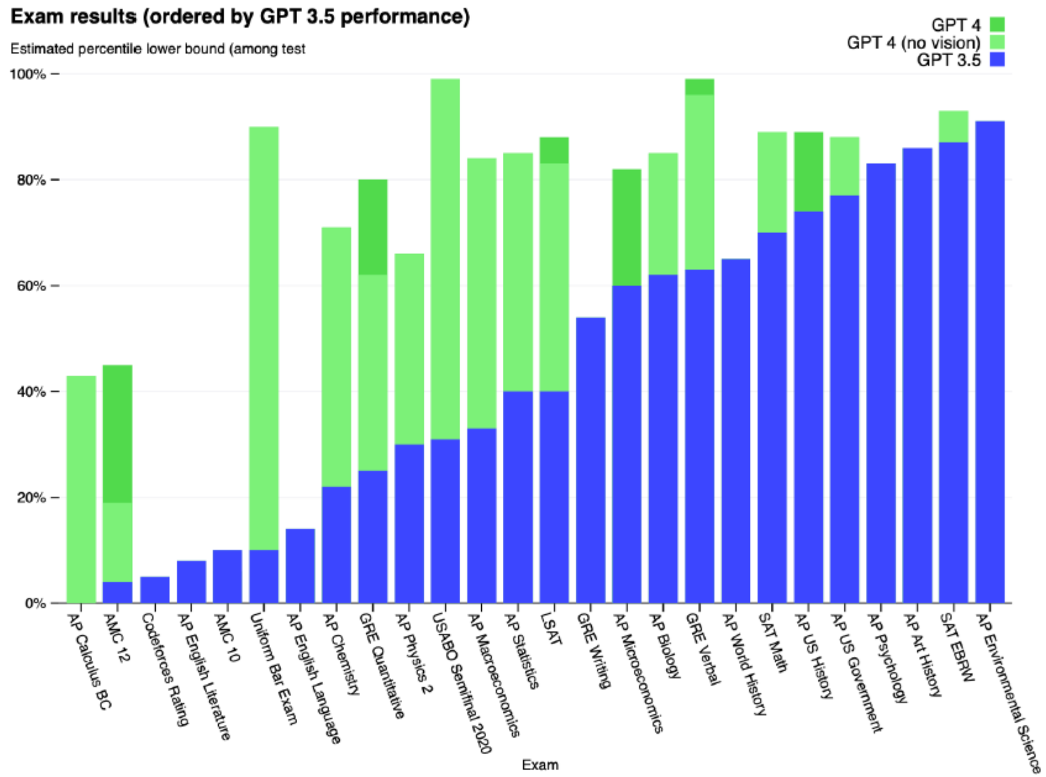
Input: Both images and text

Output: Text

- Trained on next word prediction using public and licensed data.
- Fine-tuned through ***Reinforcement Learning with Human Feedback*** (RLHF) in order to align the models output with the user's intent [OP2023].
- Models capabilities originate from the pre-training process and not the RLHF [OP2023].

GPT-4

- GPT-4 exhibits human-level performance on various professional and academic benchmarks [OP2023].



GPT performance on academic and professional exam [OP2023].

ChatGPT-4 Capabilities

- Visual inputs
- Steerability
- Significantly reduced hallucinations
- Improved safety and alignment
- Improved mathematical reasoning
- Strong performance in many languages

ChatGPT-4 Capabilities

	GPT-4 Evaluated few-shot	GPT-3.5 Evaluated few-shot	LM SOTA Best external LM evaluated few-shot	SOTA Best external model (incl. benchmark-specific tuning)
MMLU [49] Multiple-choice questions in 57 subjects (professional & academic)	**86.4%** 5-shot	70.0% 5-shot	70.7% 5-shot U-PaLM [50]	75.2% 5-shot Flan-PaLM [51]
HellaSwag [52] Commonsense reasoning around everyday events	**95.3%** 10-shot	85.5% 10-shot	84.2% LLaMA (validation set) [28]	85.6 ALUM [53]
AI2 Reasoning Challenge (ARC) [54] Grade-school multiple choice science questions. Challenge-set.	**96.3%** 25-shot	85.2% 25-shot	85.2% 8-shot PaLM [55]	86.5% ST-MOE [18]
WinoGrande [56] Commonsense reasoning around pronoun resolution	**87.5%** 5-shot	81.6% 5-shot	85.1% 5-shot PaLM [3]	85.1% 5-shot PaLM [3]
HumanEval [43] Python coding tasks	**67.0%** 0-shot	48.1% 0-shot	26.2% 0-shot PaLM [3]	65.8% CodeT + GPT-3.5 [57]
DROP [58] (F1 score) Reading comprehension & arithmetic.	80.9 3-shot	64.1 3-shot	70.8 1-shot PaLM [3]	**88.4** QDGAT [59]
GSM-8K [60] Grade-school mathematics questions	**92.0%*** 5-shot chain-of-thought	57.1% 5-shot	58.8% 8-shot Minerva [61]	87.3% Chinchilla + SFT+ORM-RL, ORM reranking [62]

Performance of GPT-4 on academic benchmarks [OP2023].

GPT-4 Limitations

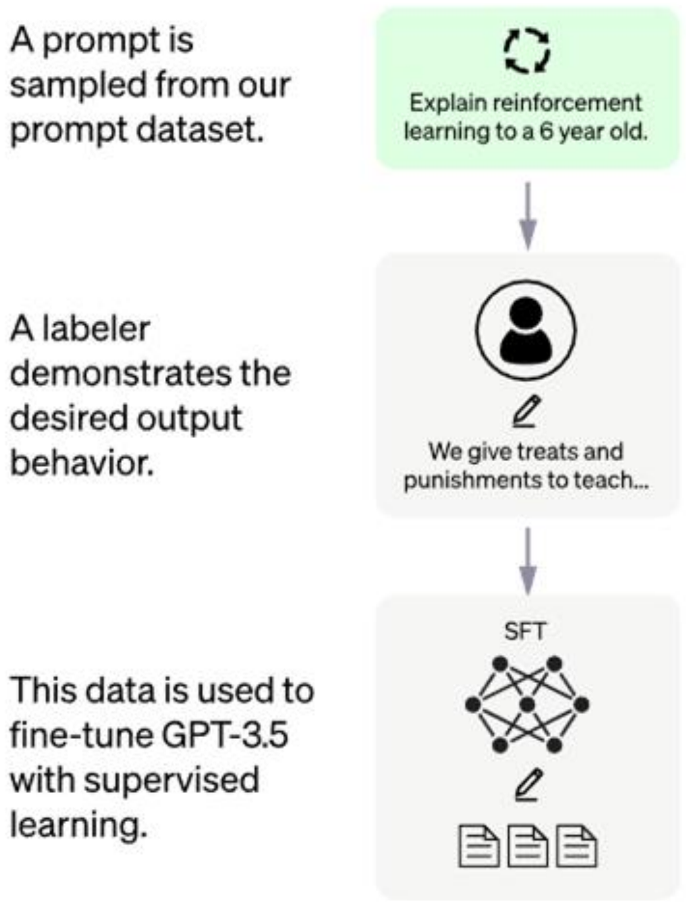
GPT-4 suffers from the same limitations as the previous GPT models [OP2023]:

- Hallucinations.
- Bias in its output text.
- Lack knowledge past 2021 and doesn't learn from its experience.
- There is still a risk of generating harmful advice, buggy code and inaccurate information. This risk has been reduced compared to older models through additional signal in the RLHF.

ChatGPT Fine-Tuning

A pre-trained 3rd generation GPT DNN for language tasks is acquired.

- Step 1: Fine-tune the pre-trained GPT DNN on a labelled dataset [OPE2023].

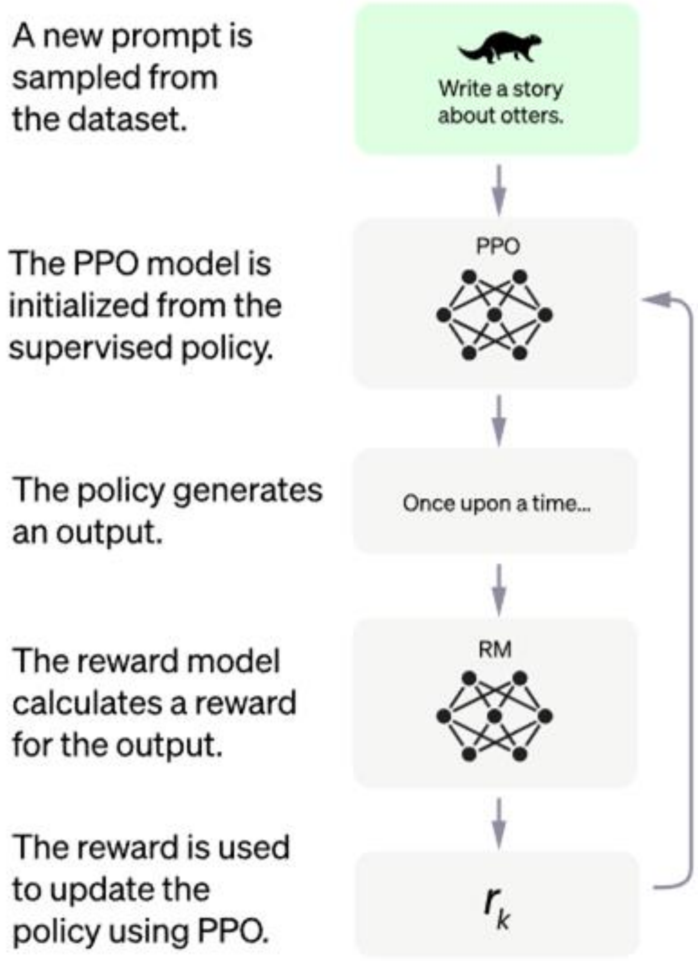


ChatGPT fine-tuning (step 1) [OPE2023].

ChatGPT Fine-Tuning

- Step 2: A reward model is trained with a scalar output.
- The output quantifies how good was the response of the fine-tuned GPT to a given prompt.

- ***Human-in-the-loop*** through ***Reinforcement Learning from Human Feedback*** (***RLHF***).



Step 3 of chatGPT fine-tuning using RLHF [OPE2023].

ChatGPT Fine-Tuning

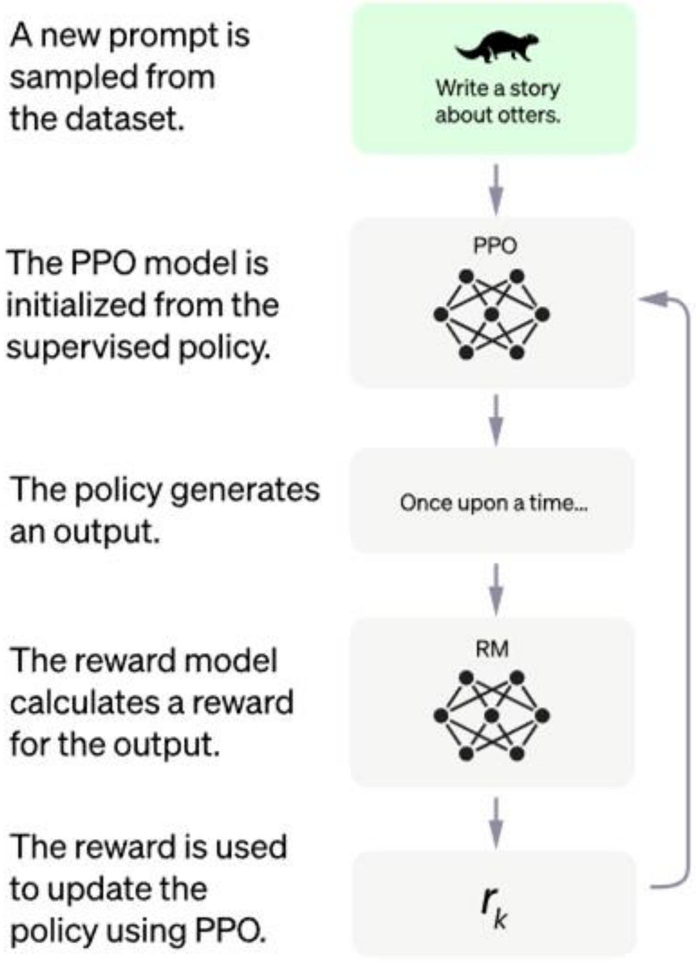
ChatGPT reward model:

- It is trained on a dataset of responses returned by the fine-tuned GPT-3 for a given prompt [OPE2023].
- For each prompt, the fine-tuned GPT outputs four responses according to a decoding strategy by sampling fresponses with the highest probability.
- The responses are labelled by a reward proportional to the quality of each output.
- Non-toxic and factual responses are given a higher reward.

ChatGPT Fine-Tuning

Reinforcement Learning with Human Feedback (***RLHF***).

- Step 3: The ***On-policy Proximal Policy Optimization*** (***PPO***) reinforcement learning algorithm is fine-tuned to optimize the scalar reward output of the reward model [OPE2023].



Step 3 of chatGPT fine-tuning using RLHF [OPE2023].

ChatGPTCapabilities

ChatGPT ***text processing*** capabilities:

- ***Translation***: chatGPT performs well translating in English [BAN2023].
- ***Summarization***: Adequate results (similar to GPT3). However, it is outperformed by SOTA works [BAN2023].
- ***Question Answering***: Near perfect scores [BAN2023].
- ***Sentiment Analysis***: It outperforms supervised SOTA works [SCA2022] and zero-shot multilingual LLM [CAH2022] (evaluation metric: F1 score) [BAN2023].

ChatGPT Capabilities

- ***Dialogue tasks***: ChatGPT generates high quality fluent human-like responses [BAN2023].
- ***Misinformation detection***: ChatGPT detected misinformation at 92% and 73.33% accuracy on covid-scientific and covid-social datasets, containing scientific and social claims related to Covid-19 accordingly [BAN2023].
- ***Code understanding and generation***: ChatGPT achieved higher score on the LinkedIn Python skills assessment than 85% of humans [CFTE].

ChatGPT Reasoning

- Despite performing well on certain reasoning tasks, ChatGPT is unreliable, as its responses are inconsistent [BAN2023].
  - Its reasoning evaluation was performed via question answering.
- ChatGPT has acceptable performance in deductive, abductive, temporal, causal and analogical reasoning [BAN2023].
- ChatGPT has weakness in inductive, spatial, mathematical, non-textual semantic and multi-hop reasoning [BAN2023].

ChatGPT Reasoning

AI4media

Categories	Testset	Results
Deductive	ENTAILMENTBANK bAbI	28/30
Inductive	CLUTRR	13/30
Abductive	αNLI	26/30
Mathematical	Math	13/30
Temporal	Timedial (formatted)	26/30
Spatial	SpartQA StepGame (hard) StepGame (diagonal) StepGame (clock-direction)	12/30 7/30 11/20 5/20
Common sense	CommonsenseQA Pep-3k (Hard)	27/30 28/30
Causal	E-Care	24/30
Multi-hop	hotpotQA	8/30
Analogical	Letter string analogy	30/30

ChatGPT results on reasoning tasks [BAN2023].

ChatGPT Limitations

- ChatGPTs responses sometimes sound plausible, while they are incorrect or nonsensical [OPE2023].
- ChatGPT responses are sensitive to tweaks in input phrasing and prompt repetition [OPE2023].
- Training data bias causes excessively verbose responses and overuse of certain phrases [OPE2023].
- In translation, it still lacks excellent ability to successfully translate English in other languages [BAN2023].

ChatGPT Limitations

- In the case of an ambiguous query, the model guesses what the user intended to say, rather than ask for clarifying questions [OPE2023].
- ChatGPT sometimes responds to harmful instructions or outputs biased answers.
  - The Moderation API is used to flag certain types of unsafe content [OPE2023].
- ChatGPT has a limited understanding of low-resource languages, due to low training data volume [BAN2023].

ChatGPT Limitations

- There are concerns and limitations due to lack of controllability and knowledge grounding in ChatGPT responses [BAN2023].

- ChatGPT fails in basic reasoning for recommendations.
  - It fails to answer correctly 66% of the time [BAN2023].

ChatGPT Limitations

ChatGPT hallucinations

- Statistical token associations may occasionally be out of context, or semantically wrong or completely fabricated.
- LLM training optimizes an ***objective function*** (or ***reward***) for a certain task.
- ***AI alignment problem***: a misaligned AI system may optimize some objective function, but not necessarily the intended one.
- Alternatively, an aligned AI system may get stuck in local minima and work sub-optimally.

ChatGPT Limitations

ChatGPT hallucinations

- Reward functions can induce ChatGPT into hallucinating facts, rather than admitting ignorance.
- Hallucinations can become even more serious when ***human-in-the-loop*** LLM retraining or fine-tuning is employed.
- Users can trigger hallucinated replies, e.g., that ‘the Pope is a pop singer’, as the LLM thinks it maximizes its reward.

ChatGPT Limitations

ChatGPT hallucinations

- Humans make such judgement errors as well:
  - Sensory illusions, wild children’s imagination.
- The human mind creates ***mental images*** of the world that map reality, yet are completely artificial, real, but different from reality.
- Arts can be considered as a form of creative expressed hallucination.

ChatGPT Limitations

ChatGPT hallucinations

- In principle, ***Generative AI fabricates imaginary outputs***.
- They may deviate from the training data and ‘common human sense’.
- Depending on their ***social use***, we can call them Art or Fake data or hallucinations.

ChatGPT: Questionmarks

- ***Does ChatGPT have access to external resources? No.***
  - Yet, if suitably trained ChatGPT can provide lots of factual information.
  - If not, what is its ***knowledge storage capacity***?
- ***Should LLMs have access to external resources? Yes.***
  - Knowledge graphs? Algebraic computations (Symbolic Algebra)?
  - This combination has great potential, e.g., in search.

ChatGPT: Questionmarks

- ***Can LLMs provide hints on how human memory works?***
  - Associative memories, Hopfield networks.
  - CNNs can store some training data information.
  - Transformer-based LLMs are based on ***statistical associations***.

- ***Relation between human imagination and ChatGPT hallucination***?
- Kids are particularly good at fabricating facts or stories.

ChatGPT: Questionmarks

- ***Does ChatGPT have explicit reasoning mechanisms***?
  - No, it has been trained as a pure language model.
  - However, its replies ***show*** some reasoning capabilities.
- ***'Text is all we need' to learn reasoning***?
  - Language/text contain many examples of reasoning.
  - Reasoning as a result of learning-by-examples?
  - ***If proven, it is a Nobel-level breakthrough.***
    - It can reconcile Machine Learning and Symbolic AI.

ChatGPT: Questionmarks

Does ChatGPT have explicit reasoning mechanisms?

- Humans learn from their mothers, relatives, and peers how to think, based on countless everyday discussions.
- An eventual LLM ‘inference by example’ capacity may hint towards ways that ***humans learn to think***.

ChatGPT: Questionmarks

AI4media

***Causal, approximate reasoning*?**

- LLM output (statistical event cross-association):

  ‘It has repeatedly been observed (or better, has been found in the literature) that plants thrive, when the sun shines’.

- Causal argumentation:

  ‘Plants thrive when the sun shines, because they use sunlight in their photosynthesis’.

ChatGPT: Questionmarks

AI4media

- ***Do LLM/ChatGPT have abstraction mechanisms?***
  - Their internal structure and functionalities are unknown.
  - Clustering and concept creation? Rule creation?

- ***Can ChatGPT provide explicit language modelling***?
  - Derivation of **grammar and syntax rules**.

- ***ChatGPT explainability?***

ChatGPT: Questionmarks

- ***Do LLMs/ChatGPT have affect?***
  - Absolutely not in the human sense.
  - Yet, it is a disgrace that they can create such an impression to unsuspecting public, when texting like ‘I love you’.
  - Machines are very good in understanding certain affect signals, e.g., ***facial expressions***.

LLM criticism

- ‘*Human intelligence can work well with few data*’ (Chomsky, 2023) [CHO2023]: ***completely wrong***.
- The contrary is true: both machine and human learning require massive training, in terms of data, architecture complexity and energy needs.
- ***Is it possible that similar laws govern both machine and human learning?***

LLM criticism

Criticism:

- ‘*Current LLMs have many deficiencies*’,
- ‘*They do just massive plagiarism*’,
- ‘*They know nothing about particular domains*’,
- ‘*They are not multimodal, e.g., supporting visual perception*’ (except GPT-4).

- ***Completely wrong claims***. LLMs are only at the start. Great advances are expected.
- Such nihilistic criticism is similar to the ill-fated criticism of Rosenblatt’s perceptron by Minsky and Papert that led to the AI winter at the end of the 1960s.

ChatGPT in Education

- ChatGPT can change the way we search and retrieve information.
- It has the capacity to help students reply to scientific questions.

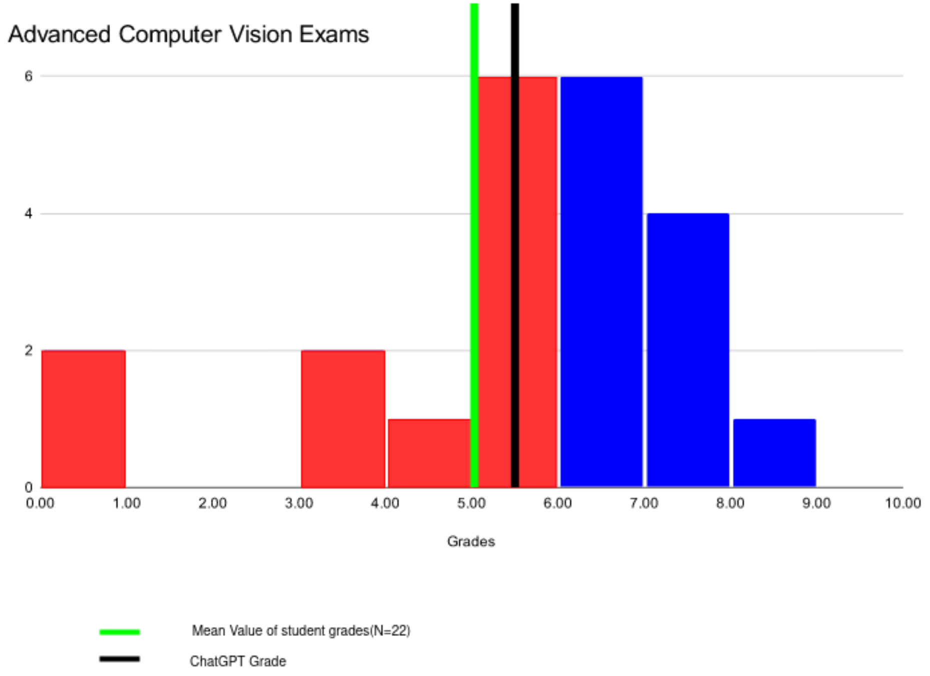
- ChatGPT changes:
  - Educational project execution and examination.
  - Educational exams.

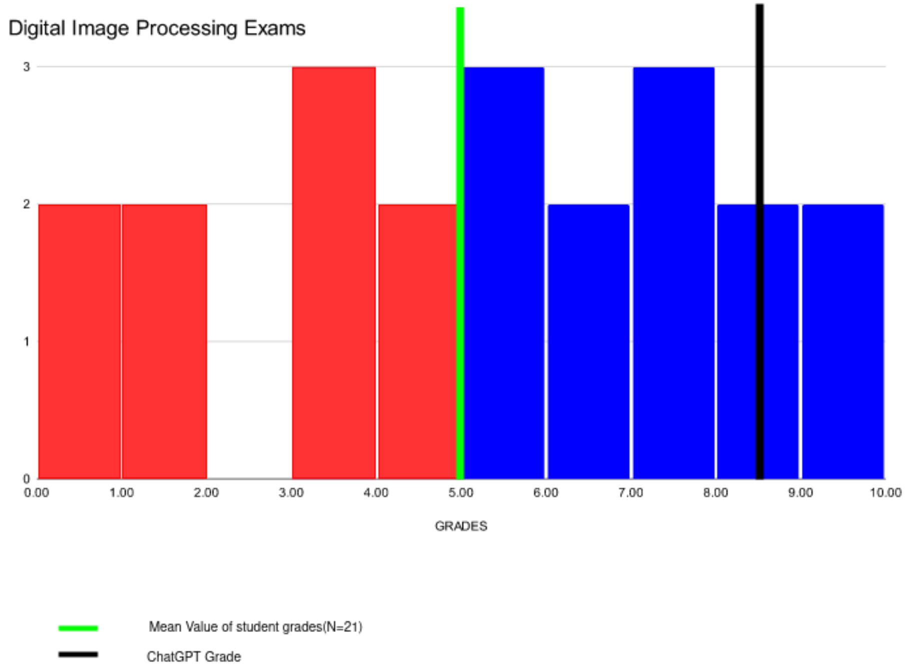
ChatGPT in Education

‘Scientific’ capacity of ChatGPT:

- Good at replying factual questions on known topics.
- It has certain capacity to reply mathematical questions.
- It can solve programming exercises very well (e.g., in Python).
- Currently, it can neither process nor output diagrams and figures.

ChatGPT in Education





ChatGPT in CS/ECE exams: very good score in mathematical questions.

ChatGPT in Education

ChatGPT and Mathematics:

- It can solve mathematical questions.
  - Complex numbers, z/Fourier transform properties.
- Replies tend to be verbose.
- It fails on some simple mathematical questions
  - It cannot always handle well reasoning with negative numbers.

    Question: He asked me to give 2 Euros. I gave him 3; how much I owe him?

    ChatGPT: You owe him 1 Euro

ChatGPT in Education

- ChatGPT in CS/ECE programming exercises: excellent scores (10/10).

Here is a Python code that implements the functionality you specified:

python   Copy code

```python
import random

def find_mean(numbers):
    """Finds the mean value of a list of numbers"""
    return sum(numbers) / len(numbers)

def moving_average(signal, window_size):
    """Applies the moving average filter of size 'window_size' on a signal"""
    filtered_signal = []
    for i in range(len(signal) - window_size + 1):
        filtered_signal.append(find_mean(signal[i:i+window_size]))
    return filtered_signal

# Test the code
numbers = [random.randint(1, 100) for i in range(3)]
print("Random numbers:", numbers)
print("Mean value of the numbers is:", find_mean(numbers))

signal = [random.randint(1, 100) for i in range(10)]
print("Random signal:", signal)
print("Filtered signal with window size 3:", moving_average(signal, 3))
```

ChatGPT programming: arithmetic mean and 1D moving average filter routines in Python.

LLMs and AI in Education

ChatGPT opportunities.

- LLMs can be used as a new education tool with massive impact in education.
- We have to research how to best use it.

- Its interaction with other teaching modes must be researched.
- Can it be used to trigger creative thinking, while speeding up tedious processes?

LLMs and AI in Education

IT and AI opportunities in education.

- What is the impact of IT and AI in teaching Mathematics?
- What is the impact of LLMs in teaching languages?
- What is the impact of Deep Arts in Arts Schools?

- What is the ***long-term impact of IT and AI*** in human memory?
- Will brain be ‘restructured’ to be, e.g., devoted more to thinking tasks than to memory?
- Can we observe such findings from historical records?

Artificial General Intelligence

Is AGI the next step after LLMs?

- A deeper understanding of LLM operation is needed.
- The exact GPT4 architecture and parameters (transformer network weights) are a well-kept corporate secret.
- A deep LLM functionality understanding would be difficult, even if LLMs were open, due to their immense complexity.
- Neuroscience did not advance enough to understand brain and human intelligence.

Artificial General Intelligence

Is AGI the next step after LLMs?

- Most probably AGI will be VERY different from human intelligence.
  - Airplanes are different then birds, yet they obey the same laws of Physics.
- The physical substrate of AI and human intelligence is very different.
  - Robots have very limited but different physical intelligence.
  - Things may change by developing biological robots.
- ***Life evolution by-design*** than through physical selection.
- Massive ***human-machine symbiosis*** at various levels.

Artificial General Intelligence

AI4media

Is AGI the next step after LLMs?

- Will AGI be any different from human intelligence from a behavioral point of view that is worth talking about?
- Today ***too many*** commoners cannot make the difference.
- The phenomenon is intensified by:
  - Lack of proper education.
  - Access of machines remotely.
  - Unwise claims and behavior of AI agents to the general public, e.g.,:
  - AI halucinations being misunderstood as imagination.
  - False claims of sentiments (internal affect states) by machines.

Artificial General Intelligence

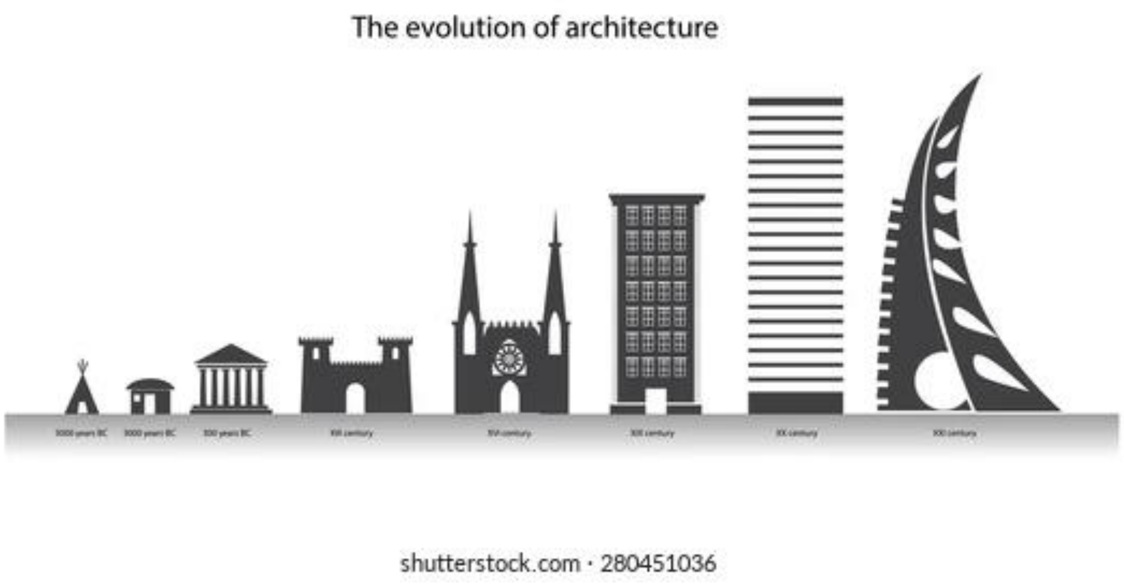
Layman's technophobia

- ***Fear of the unknown*** as commoners cannot understand AI.
- Machines appear to be intelligent and possibly better at that than the humans themselves.
- They are ***massively better*** in certain tasks, e.g., computations, memory/retrieval.
- Machines appear to be ***sentient***.
- Humans are awed by ChatGPT 'intelligence' much more than by other Generative AI methods, e.g., Deep Arts.
- ***Any technophobia can be socially destructive.***

Artificial General Intelligence

Scientific technophobia

- Very recent trend: scientists fearing the unknown.



Parable: AI and the tower of Babel.

Artificial General Intelligence

Can AI be stopped or delayed?

- ***AI is the response of humanity to a global society and physical world of ever-increasing complexity.***
- The physical and social complexity increase processes are ***very deep and seeming relentless***.
- ***AI is a blessing, but it can become a curse.***
- Political, ethical, and regulatory concerns cannot and should not stop AI research [FUT2023].
- Scientific technophobia leads nowhere [NYT2023].

Artificial General Intelligence

Can AI be stopped or delayed?

- ***AI research can and should become more open, democratic, scientific and ethical.***
- Simple AI regulatory examples:
  - AI system registry,
  - Clear indication that somebody converses with a machine.
- AI deployment should be regulated and can be temporarily delayed.
  - Geopolitical aspects must be dealt by international cooperation.

Citizen Morphosis

Information and Knowledge Society

- Information society: exponential increase of data/information, linear increase of knowledge.
- Knowledge society: exponential increase of knowledge?
- AI, IT and ***citizen morphosis*** are our only hope to have a smooth transition from the current Information Society to a Knowledge Society.
- Else, humanity may face a catastrophic social implosion, if proven unable to advance and pass knowledge to new generations (see ***start of Medieval Times***).

Citizen Morphosis

Citizen morphosis (rather than education) emphasizes the need for conscious citizens:

- with critical thinking, deduction and precise communication skills, imagination, and emotional intelligence;
- being able to understand, adapt, and ultimately harness the tremendous new technological and economic possibilities and employment prospects.
- Such a level of education is sought after today in many job positions internationally.

Citizen Morphosis

Major overhaul of education at all levels to master knowledge development and uptaking needs.

- The need for such education permeates all levels of education and all social strata.
- A 1/3-2/3 society, where 1/3 of the population understands and benefits from scientific progress, while the remaining 2/3 lags, being impoverished and technophobic, is simply not sustainable.
- Need to educate women, minorities and Global South to improve the global education level.

Citizen Morphosis

The ***basic AI and IT concepts*** are simple and can be taught at all educational levels:

- Data clustering, similarity, classification etc.
- Educational readjustment for their teaching by ***rearranging the curriculum of Mathematics and Informatics***.
- A (partial) mathematatization of education is inevitable.
- It is not certain that it is feasible, given the traditional separation of the sciences and the humanities.
- ***Classical studies*** are also an ideal tool for developing critical thinking and precision

AI and University Education

Changes will be drastic and will come very soon.

Schools of 'Information Science and Engineering' with departments of:

- Computer Science/Informatics,
- Mathematics
- Computer Engineering
- Artificial Intelligence Science and Engineering
- Internet/Web Science.

AI and University Education

AI Science and Engineering: A new scientific discipline?

- CSE spawning new disciplines ***through specialization***:
  - Web science
  - Data science
  - AI Science and Engineering.
- New scientific methodologies are not ***necessarily*** essential.
- Poor terminology?
- Past experience: ***Physics spawning Engineering disciplines***
  - Electrical Engineering, Mechanical Engineering.

AI and University Education

Creation of departments for '***Cognitive and Social Science and Engineering***' in Schools of Arts and Humanities.

- Groundbreaking proposal.
- Currently, the Humanities face the greatest pressure from LLMs and AI.
- The mathematization of classical subjects (e.g., Linguistics, Sociology) has advanced significantly.
- Alternative? Creation of departments for '***Philological/Linguistic Engineering***' or '***Social Engineering***' in Science/Engineering Schools.

AI and University Education

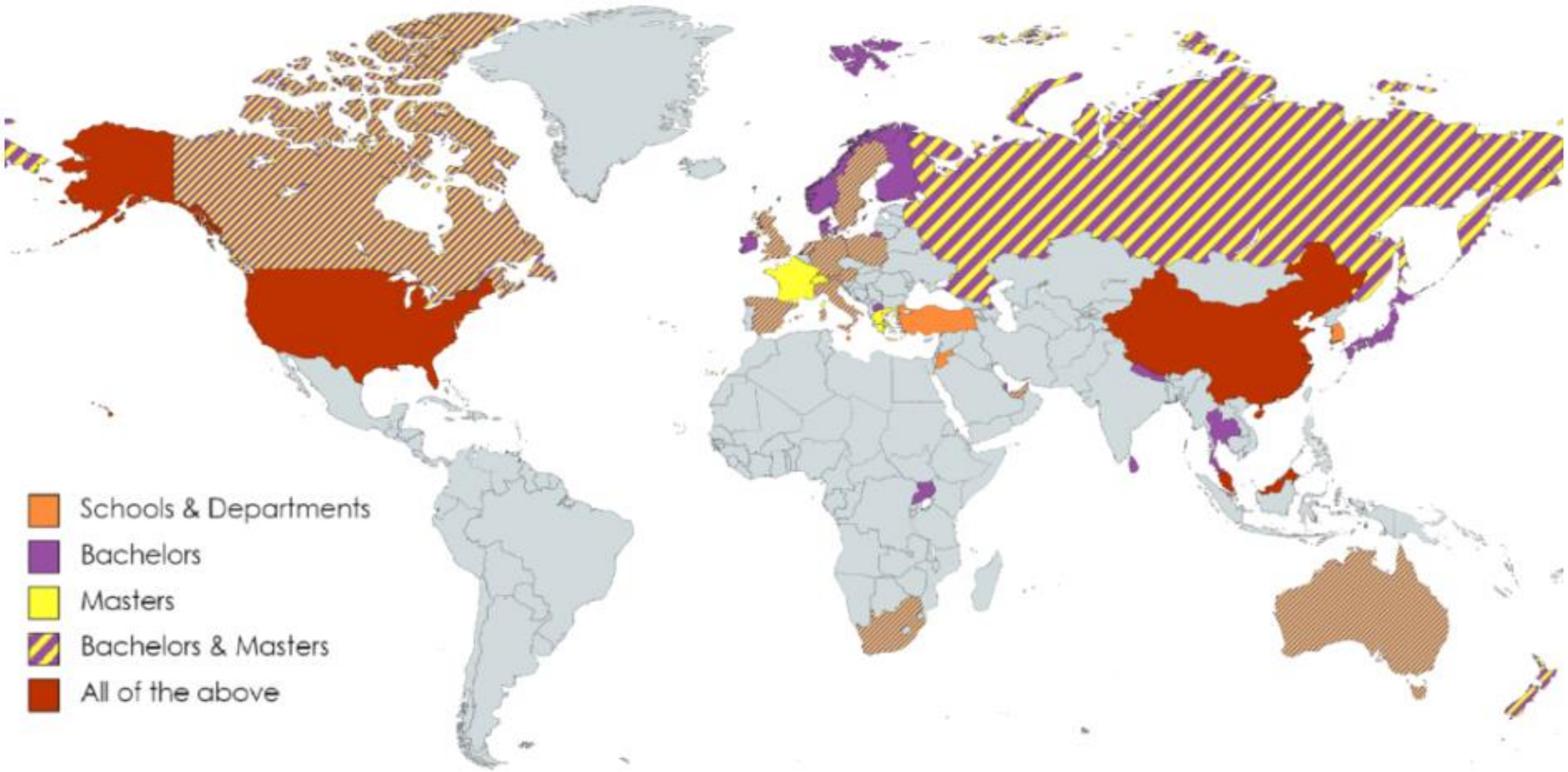
Creation of departments for '***Bio-Science and Engineering***' in Schools of Health Sciences, including:

- Biomedical Engineering, Genetic Engineering and Systems Biology.

Mandatory inclusion of Mathematics and Computer Science courses in all disciplines without exception.

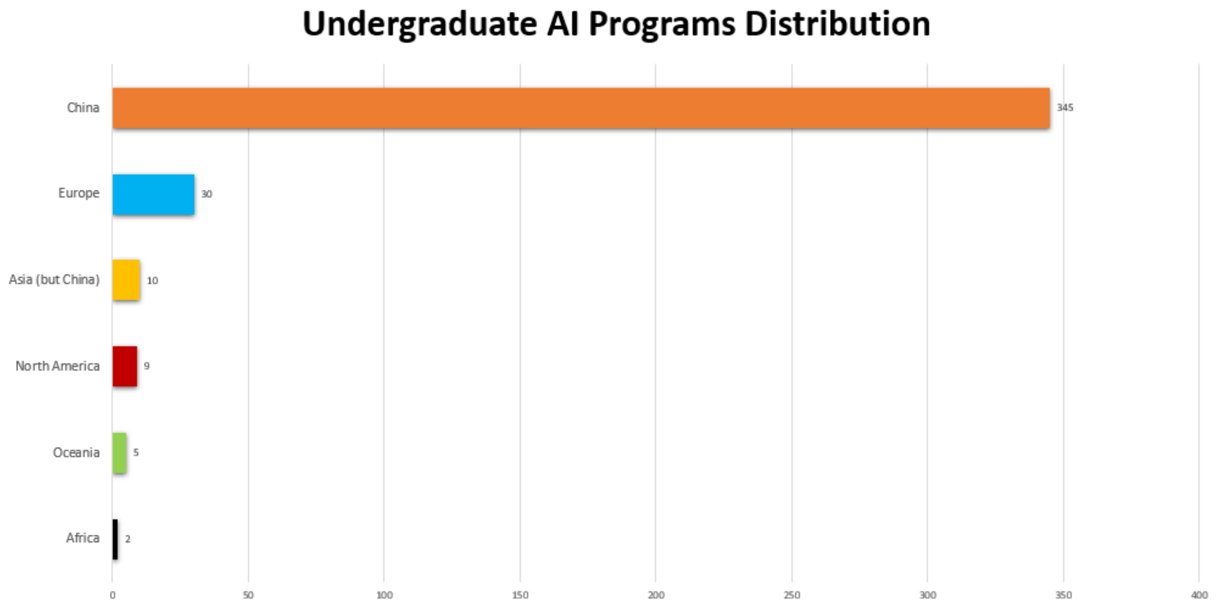
- Simply, one (poor) course in Statistics does not meet the current needs.

University Education on AI

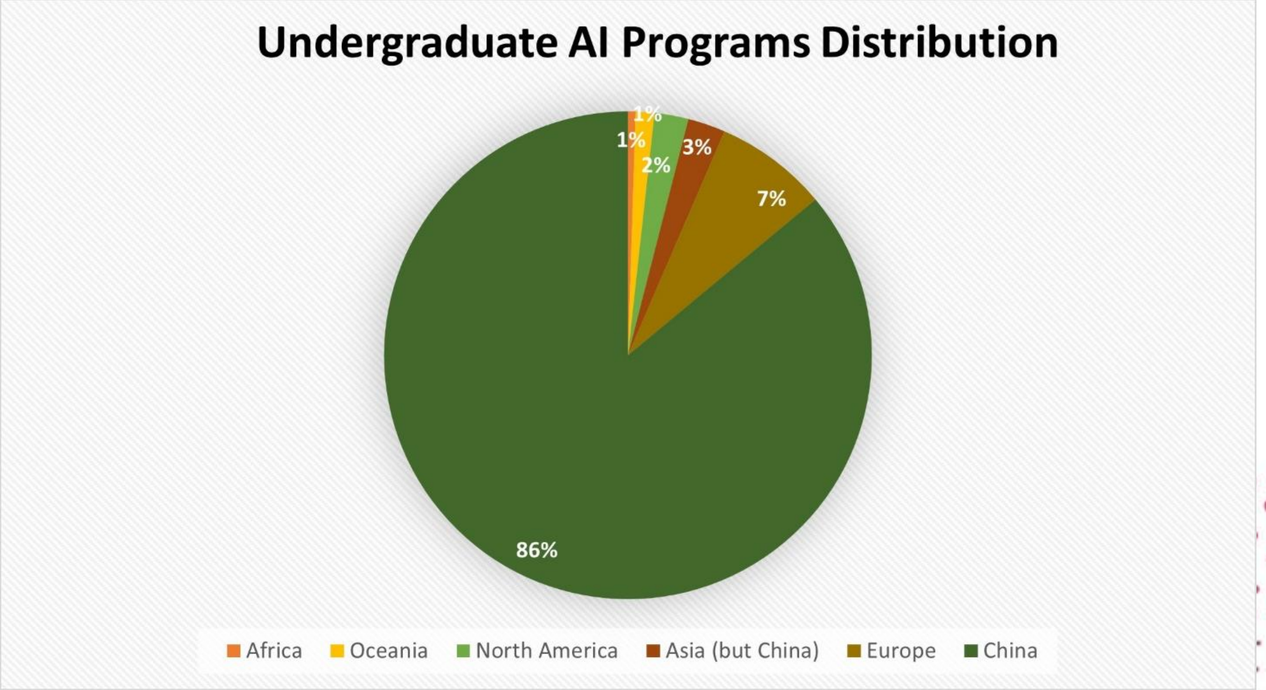


Countries that offer AI studies.

University Education on AI

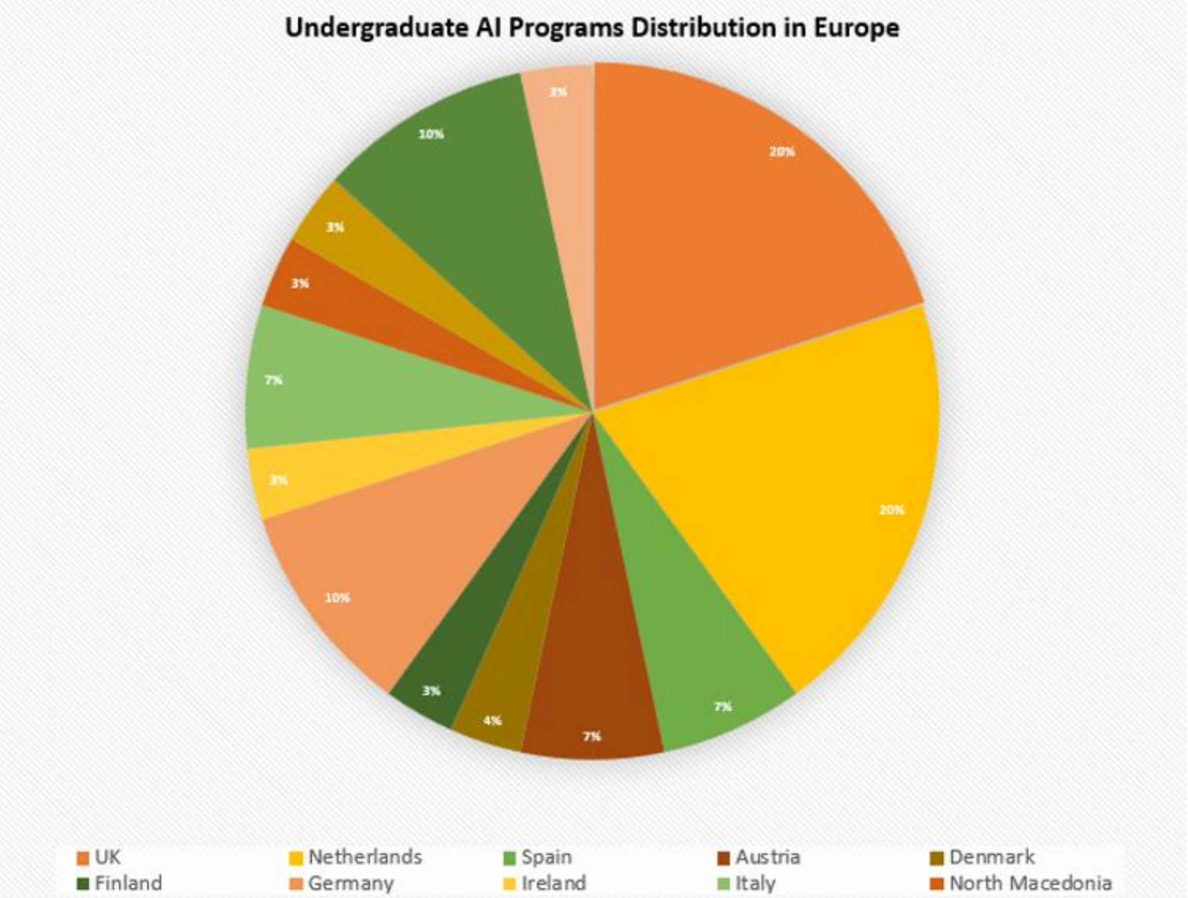


Number of undergraduate AI programs worldwide.



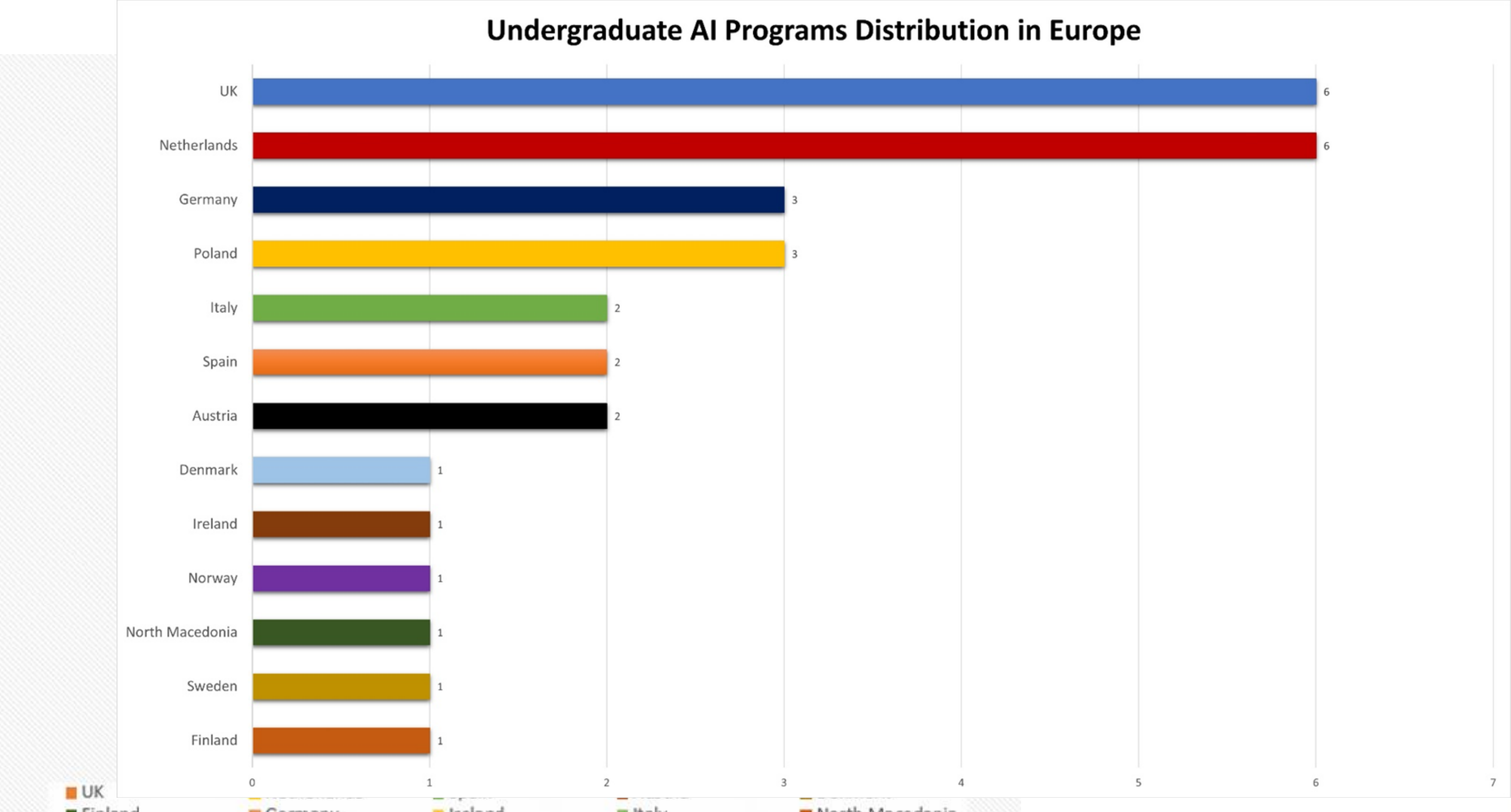
Global distribution of undergraduate AI studies.

University Education on AI



Distribution of undergraduate AI programs in Europe

University Education on AI



Geographical distribution of AI undergraduate programs in Europe.

University Education on AI

AI-centered Schools & Departments (examples):

- Machine Learning Department, Carnegie Mellon University, USA.
  - https://www.ml.cmu.edu/

- Institute for AI, Tsinghua University, China.
  - https://ml.cs.tsinghua.edu.cn/thuai/#/

- School of Intelligence Science and Technology, Peking University, China.
  - https://www.cis.pku.edu.cn/English/Home.htm

- Department of AI, College of Informatics, Korea University, S. Korea
  - http://xai.korea.ac.kr/eng/company/greeting?language=eng

University Education on AI

Undergraduate AI Studies (examples):

- BSc in Data Science and AI, Nanyang Technological University, Singapore.
  - https://www.ntu.edu.sg/education/undergraduate-programme/bachelor-of-science-in-data-science-artificial-intelligence
- BSc in AI , University of Technology Sydney, Australia.
  - https://www.uts.edu.au/study/find-a-course/bachelor-artificial-intelligence
- BSc in AI and Decision Making, Massachusetts Institute of Technology , USA.
  - http://catalog.mit.edu/degree-charts/artifical-intelligence-decision-making-course-6-4/
- BSc in AI, The University of Edinburgh , UK.
  - https://www.ed.ac.uk/studying/undergraduate/degrees/index.php?action=view&code=G700
- BSc in AI, Vrije Universiteit Amsterdam, Netherlands.
  - https://vu.nl/en/education/bachelor/artificial-intelligence
- BSc in AI, Polytechnic University of Catalonia, Spain.
  - https://www.upc.edu/en/bachelors/artificial-intelligence-barcelona-fib

University Education on AI

Curriculum of BSc on AI, ***CMU, USA*** (example)

https://www.cs.cmu.edu/bs-in-artificial-intelligence/

- Principles of Imperative Computation
- Integration and Approximation
- Mathematical Foundations for Computer Science
- Great Theoretical Ideas in Computer Science
- Matrices and Linear Transformations
- Calculus in Three Dimensions
- Concepts in AI
- AI: Representation and Problem Solving
- Parallel and Sequential Data Structures and Algorithms
- Probability Theory for Computer Science
- Introduction to Machine Learning
- Introduction to Computer Systems
- Computer Vision
- Natural Language Processing
- Modern Regression
- Neural Computation
- Autonomous Agents
- Cognitive Robotics: The Future of Robot Toys
- Planning Techniques for Robotics
- Mobile Robot Algorithms Laboratory
- Robot Kinematics and Dynamics
- Deep Reinforcement Learning & Control

AI and University Education

Curriculum of BSc on AI, CMU, USA

- Mobile Robot Algorithms Laboratory
- Robot Kinematics and Dynamics
- Deep Reinforcement Learning & Control
- Deep Learning Systems: Algorithms and Implementation
- Intermediate Deep Learning
- Machine Learning for Structured Data
- Machine Learning for Text and Graph-based Mining
- Introduction to Deep Learning
- Advanced Methods for Data Analysis
- Search Engines
- Speech Processing
- Computational Perception
- Computational Photography
- Design of Artificial Intelligence Products
- Human AI Interaction
- Designing Human Centered Software
- Human Robot Interaction

AI and University Education

Curriculum of MSc in Machine Learning, UCL, UK (example)

https://www.ucl.ac.uk/prospective-students/graduate/taught-degrees/machine-learning-msc

- Applied Machine Learning
- Advanced Topics in Machine Learning
- Approximate Inference and Learning in Probabilistic Models
- Probabilistic and Unsupervised Learning
- Statistical Natural Language Processing
- Reinforcement Learning
- Machine Vision
- Supervised Learning
- MSc Machine Learning Project
- Machine Learning Seminar
- Bayesian Deep Learning
- Statistical Learning Theory
- Applied Deep Learning
- Graphical Models

AIDA at a glance

AIDA
ARTIFICIAL INTELLIGENCE DOCTORAL ACADEMY

The International AI Doctoral Academy (AIDA) is a joint initiative of the European R&D Projects:

AI4media — www.ai4media.eu

VISION — www.vision4ai.eu

HUMANE AI NET — www.humane-ai.eu

elise — European Network of AI Excellence Centres — www.elise-ai.eu

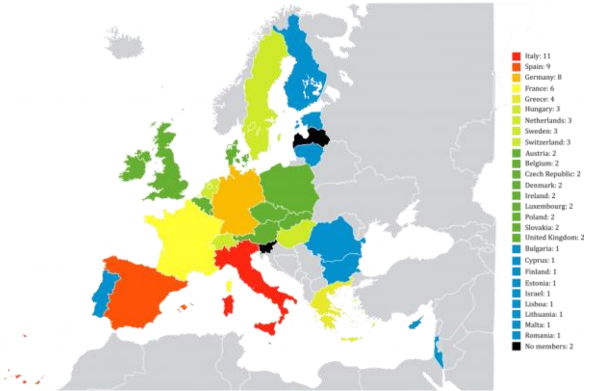
TAILOR — www.tailor-network.eu

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE) and No. 952215 (TAILOR)

AIDA at a glance

Membership:

- **77 members (AI Universities, R&D centers, Industry)**



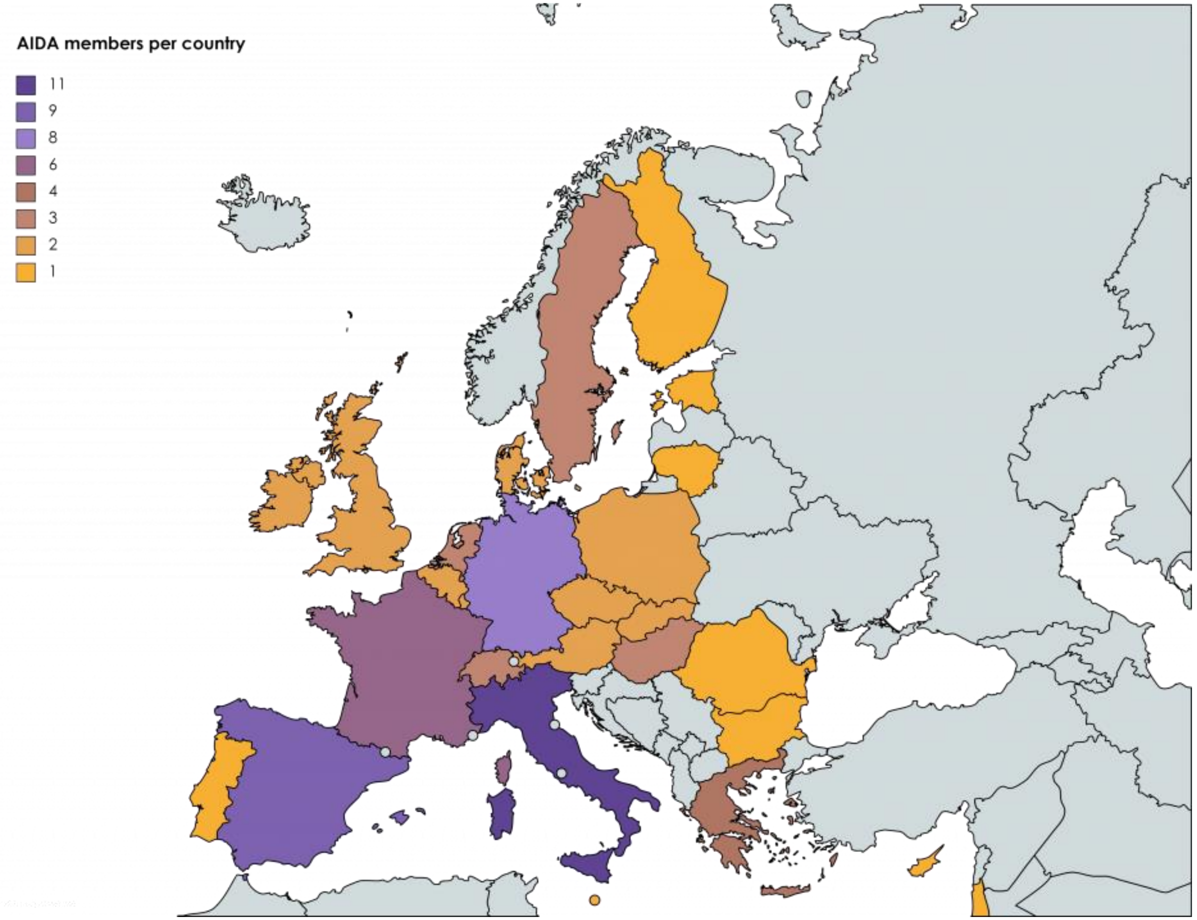
Focus on AI education (PhD/Postdoc)

Operation highlights:

- **AIDA Lecturers**: 121
- **AIDA Students**: 191 (176 PhD and 15 Post Docs)
- Junior fellows programme launched **58 applications,** 39 completed, 12 ongoing, 7 scheduled for the next period
- Courses:
  15 Spring 2023 courses (ongoing/in preparation): 6 semester courses/lecture series and 9 short courses
  14 Fall 2023 courses (8 short courses and 6 semester courses/ lecture series)
- **AIDA courses** have attracted a total of **1,760+ participants**
- **AIDA email list registrants**: 806+
- **32 Lectures** for AI Excellence Lecture Series. Attract ~135 attendees on average
- **70 educational materials** in repository: 69 AI4Media (50 of them from AUTH)
- **17 curators** (all from AI4media)

AIDA structure

AIDA Members



Geographic distribution of AIDA Membership.

AIDA has already become a well-recognized reference point in European AI education, widening its members, attracting more students, and enhancing its rich AI curriculum.

AIDA Members

- ## Universities

Universities that initially signed the AIDA Memorandum of Understanding.

- **Aristotle University of Thessaloniki (AUTH)**. Representative: Ioannis Pitas
- **Alma Mater Studiorum Università di Bologna**. Representative: Paolo Torroni
- **Bar-Ilan University**. Representative: Sarit Kraus
- **Ca' Foscari University of Venice (UNIVE)**. Representative: Marcello Pelillo
- **Charles University**. Representative: Jan Hajič
- **Eindhoven University of Technology**. Representative: Joaquin Vanschoren
- **Graz University of Technology**. Representative: Horst Bischof
- **Katholieke Universiteit Leuven**. Representative: Peggy Valcke / Bart De Moor
- **La Sapienza, DIAG**. Representative: Giuseppe De Giacomo
- **Lancaster University**. Representative: Hossein Rahmani
- **Linköping University (LiU)**. Representative: Fredrik Heintz
- **Poznan University of Technology.** Representative: Mikołaj Morzy
- **Queen Mary University of London**. Representative: Ioannis Patras
- **RWTH Aachen Center for Artificial Intelligence**. Representative: Gerhard Lakemeyer
- **Technische Universität Wien (TU Wien).** Representative: Thomas Eiter
- **TU Darmstadt**. Representative: Kristian Kersting
- **Umeå Universitet**. Representative: Juan Carlos Nieves
- **Universidad de Málaga**. Representative: Manuel López-Ibáñez
- **Università degli Studi di Firenze**. Representative: Alberto del Bimbo
- **Università degli Studi di Trento**. Representative: Nicu Sebe
- **Università di Pisa**. Representative: Dino Pedreschi
- **Università ta' Malta**. Representative: G.N Yannakakis
- **Università degli Studi di Milano (UNIMI)**. Representative: Nicolò Cesa-Bianchi
- **Universitat Politècnica de València.** Representative: Vicent Botti
- **Universitat Pompeu Fabra (UPF)**. Representative: Vicenç Gómez
- **Universitatea Politehnica din București**. Representative: Bogdan Ionescu
- **Université Côte d'Azur**. Representative: Lucile Sassatelli
- **Université d'Artois**. Representative: Bogdan Ionescu
- **Université Paris-Saclay**. Representative: Lina Ye
- **Universiteit van Amsterdam**. Representative: Cees Snoek
- **University College Cork – National University of Ireland**. Representative: Barry O'Sullivan
- **University of Basel (UNIBAS)**. Representative: Malte Helmert
- **University of Genoa (UNIGE)**. Representative: Lorenzo Rosasco
- **University of Lorraine**.Representative: Slim Ouni
- **University Of Modena and Reggio Emilia (UNIMORE)**. Representative: Prof. Rita Cucchiara
- **Univerzita Komenského v Bratislave**. Representative: Martin Homola
- **Uniwersytet Warszawski**. Representative: Andrzej Nowak
- **Vrije Universiteit Brussel**. Representative: Ann Nowé
- **University of the Basque Country UPV/EHU.** Representative: Koldo Gojenola Galletebeitia
- **Università Cattolica del Sacro Cuore. R**epresentative: Giuseppe Riva
- **Information Systems Insitute, Haute Ecole Spécialisée de Suisse occidentale. R**epresentative: Henning Müller
- **University of Alicante.** Representative: Jose Garcia-Rodriguez
- **University of Montpellier.** Deputy: Anne Laurent

AIDA Members

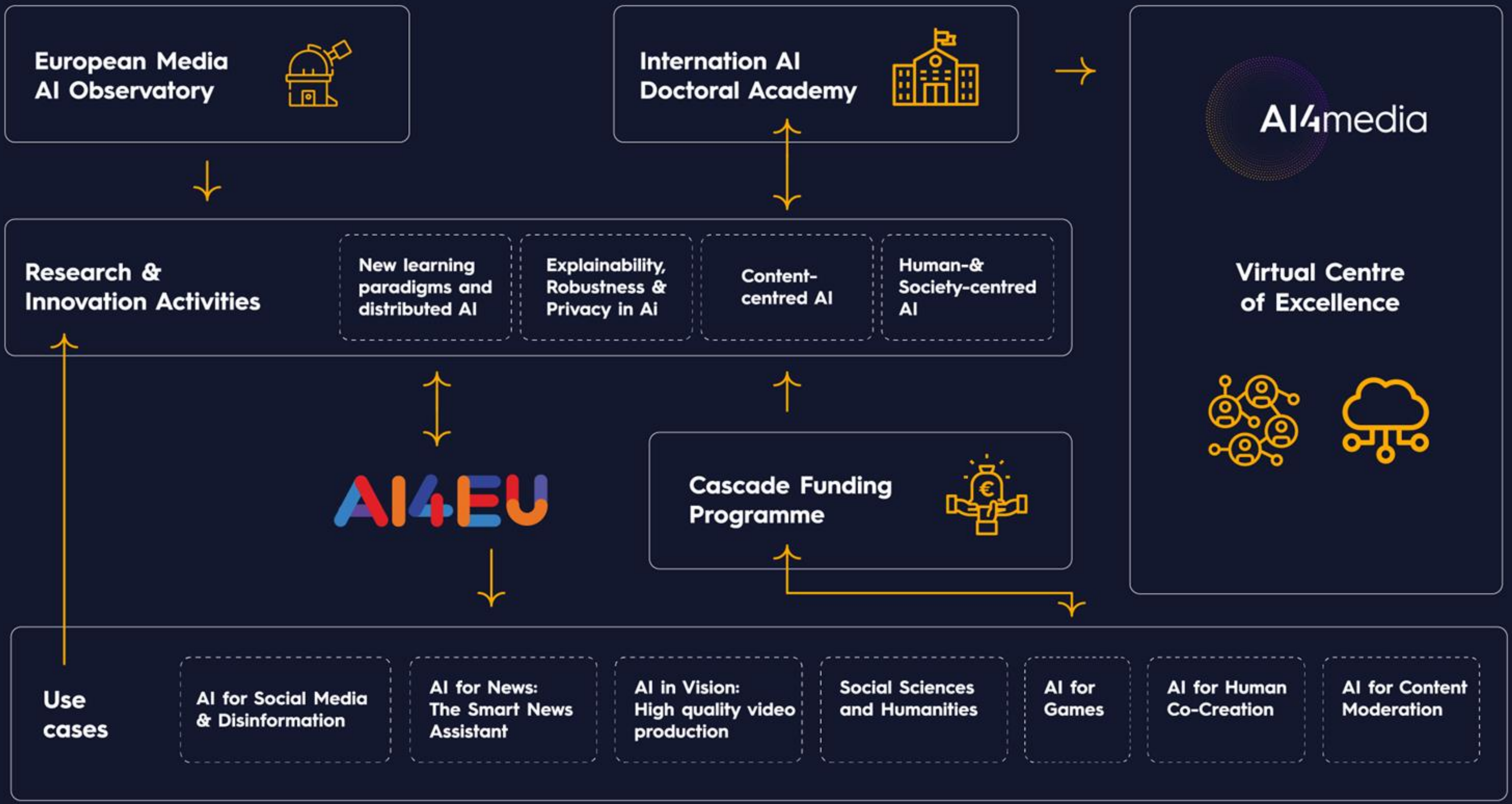
• Research/Industry Members

- **Athens Technology Center**. Representative: Danae Tsabouraki
- **Computer Vision Centre (CVC)**. Representative Dimosthenis Karatzas
- **Ethiko Kentro Erevnas Kai Techologikis Anaptyxis**. Representative: Ioannis Kompatsiaris
- **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V – Institute for Digital Media Technology IDMT**.Representative: Patrick Aichroth
- **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V – Institute for Intelligent Analysis and Information Systems IAIS**. Representative: Joachim Köhler
- **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V – Institute for Manufacturing Engineering and Automation IPA**. Representative: Marco Huber
- **Grassroots Arts and Research UG**. Representative: Carmen Mac Williams
- **IBM Ireland Ltd**.Representative: Killian Levacher
- **Idiap Research Institute**. Representative: Daniel Gatica-Perez
- **Imagga Technologies LTD**. Representative: Georgi Kostadinov
- **Kempelen Institute of Intelligent Technologies**. Representative: Maria Bielikova
- **Luxembourg Institute of Health**. Representative: Jasmin Schulz
- **Modl.ai APS**. Representative: Lars Henriksen
- **Stichting Nederlands Instituut voor Beeld en Geluid**. Representative: Johan Oomen
- **Volkswagen Group (Machine Learning Research Lab)**. Representative: Patrick van der Smagt
- **SZTAKI**. Representative: Csaba Benedek
- **Advanced Center for Aerospace Technologies (CATEC)**. Representative: Antidio Viguria Jiménez
- CNR **(Consiglio Nazionale delle Ricerche).** Representative: Fabrizio Sebastiani

AI4Media at a glance

AI4Media's Mission

- Deliver the **next generation AI Research and Training at the service of media, society and democracy**
- Ensure the embedding of **ethical and trustworthy AI** into future AI deployments
- Reimagine AI as a **human-centered, trusted and beneficial enabling technology for media and society**


European Media AI Observatory
Internation AI Doctoral Academy
AI4media
Research & Innovation Activities
New learning paradigms and distributed AI
Explainability, Robustness & Privacy in Ai
Content-centred AI
Human-& Society-centred AI
Virtual Centre of Excellence
AI4EU
Cascade Funding Programme
Use cases
AI for Social Media & Disinformation
AI for News: The Smart News Assistant
AI in Vision: High quality video production
Social Sciences and Humanities
AI for Games
AI for Human Co-Creation
AI for Content Moderation

Unique selling point

AI4media

Next Generation AI for the Media

Core Research	Real-world applications	Impact of policy & regulation	Societal concerns	Education & training

Core Research


Multi-modal content


Human & Society in the centre


New-gen of machine learning systems


Trustworthy AI solutions

Real-world applications


Fact-checking & verification


Automated game design


News production automation


Content moderation


Human-machine artistic co-creation

Impact of policy & regulation


Monitoring of EU regulatory landscape


New policy recommendations

Societal concerns

Analysis of media AI impact & societal concerns

Education & training

AIDA
ARTIFICIAL INTELLIGENCE DOCTORAL ACADEMY

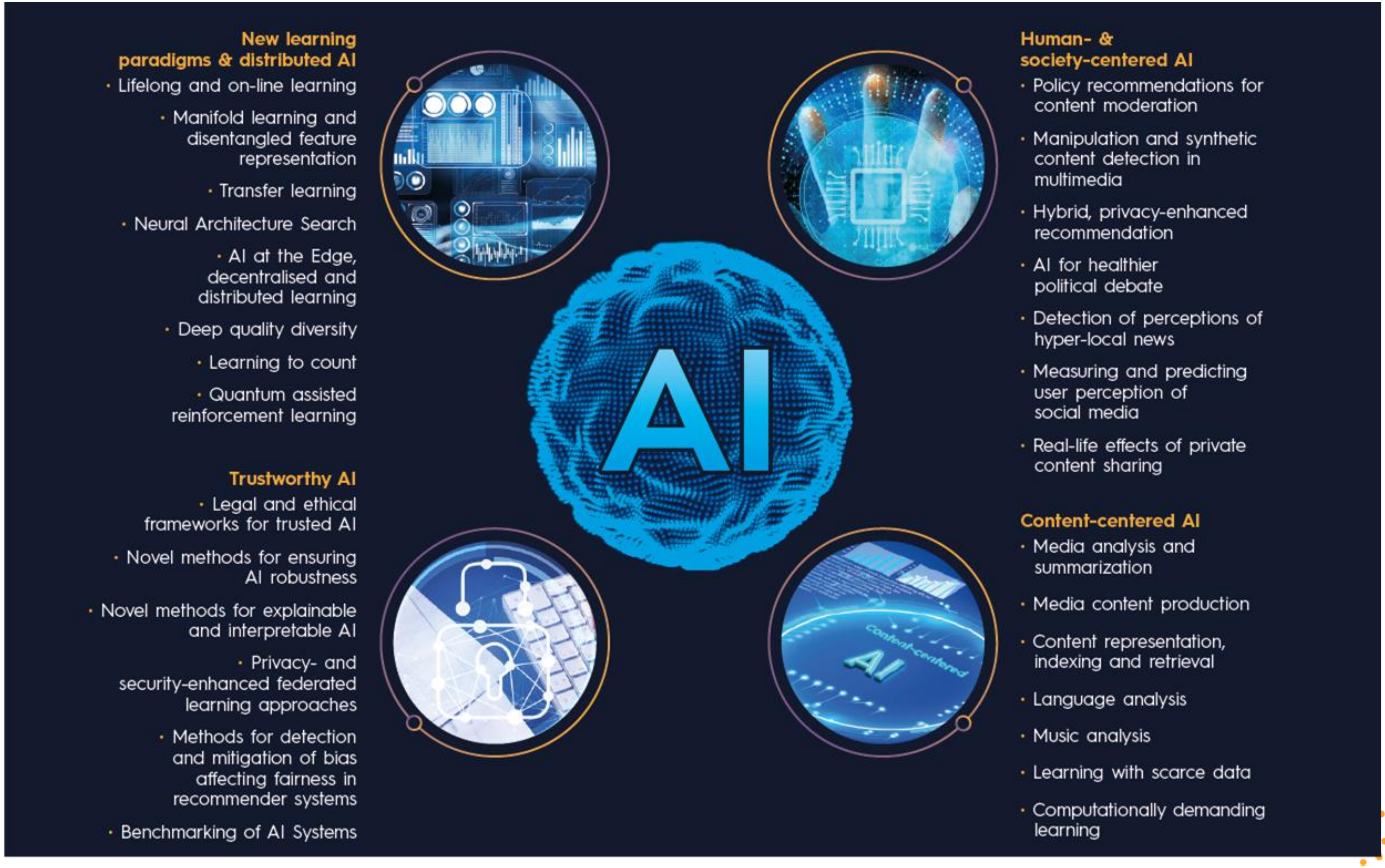
Junior Fellows Exchange Program

Strategic Research Agenda on AI for the media industry

- Lays out the strategic plan for AI4Media's R&I activities and aims to become a useful source of information for AI researchers, media practitioners and policymakers
  - main research themes to be tackled by the consortium
  - current challenges of media AI
  - research directions that need to be pursued to address the challenges
  - applications of AI in the media industry
  - potential impact of this research

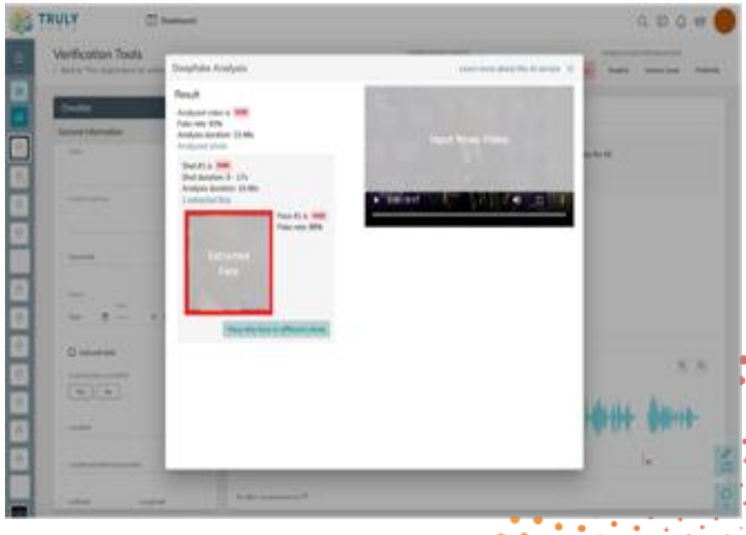
https://www.ai4media.eu/strategic-research-agenda-on-ai-for-the-media-industry/

Research & innovation in AI for Media

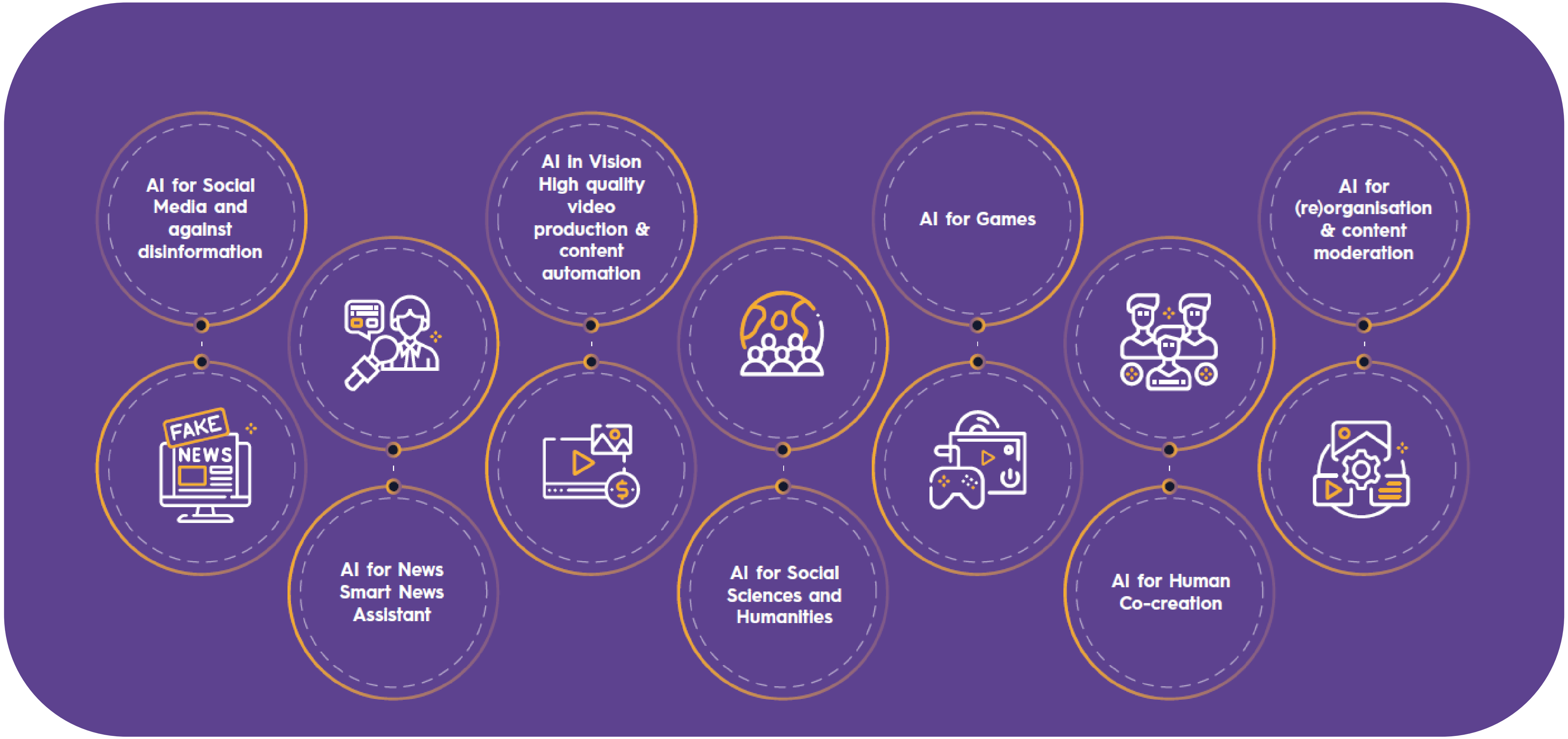


Example: New AI technologies in the fight against disinformation

- Synthetic content detection – tools for deepfake and manipulation detection
    - Exploring multiple modalities: video, image, audio, text
    - Able to generalize well and offer robustness against adversarial attacks
    - Already integrated in existing applications used by journalists and fact-checkers
- Online political debate analysis – tools for healthier debate
    - Measuring a variety of indicators: ephemerality, offensiveness, topic causality, botness, propaganda, fallacious argumentation
    - New datasets of a) Covid-19 tweets, b) Greek politics tweets, c) multilingual sentiment analysis (8 languages) news dataset focused on European politics
- Analysis of hyper-local news
    - Analyse factors involved in perception of news credibility and news consumption
    - Beyond topic modelling -> frame analysis of newspaper articles



Real-world applications of AI for the media - 7 use cases



https://www.ai4media.eu/use-cases/

AI Media Observatory

- A knowledge platform for AI and media
  - Your AI Media Feed -> monitors and curates relevant content and events on AI in media
  - Let's talk AI and Media -> provides expert perspectives on the potentials and challenges that AI poses for the media sector (video format)
  - Find an AI Media expert -> allows stakeholders to easily get in touch with relevant experts in the field via an expert directory

Just launched !!

https://www.ai4media.eu/ai-media-observatory/

Bibliography

AI4media

[1] I. Pitas, “Artificial Intelligence Science and Society Part A: Introduction to AI Science and Information Technology“, Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156460?ref_=pe_3052080_397514860

[2] I. Pitas, “Artificial Intelligence Science and Society Part B: AI Science, Mind and Humans“, Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156479?ref_=pe_3052080_397514860

[3] I. Pitas, “Artificial Intelligence Science and Society Part C: AI Science and Society”, Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156487?ref_=pe_3052080_397514860

[4] I. Pitas, “Artificial Intelligence Science and Society Part D: AI Science and the Environment“, Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156495?ref_=pe_3052080_397514860

Bibliography

[PIT2023a] Ioannis Pitas, CVML short course, “AI Science and Engineering and its Impact on the Society”, https://icarus.csd.auth.gr/introduction-to-ai-science-and-engineering-and-its-impact-on-the-society-and-the-environment/

[PIT2022] Ioannis Pitas, “AI Science and Engineering: A new scientific discipline?”, https://icarus.csd.auth.gr/chatgtp-in-education/

[PIT2023b] Ioannis Pitas, “ChatGPT in education”, http://icarus.csd.auth.gr/ai-science-and-engineering-a-new-scientific-discipline/

[PIT2023c] I. Pitas, “Artificial intelligence is not the new Tower of Babel. We must beware of technophobia instead”, Euronews, 8/5/2023, https://www.euronews.com/2023/05/08/artificial-intelligence-is-not-the-new-tower-of-babel-we-should-beware-of-technophobia-ins

[CHO2023] N. Chomsky, I. Roberts, J. Watumull, “The false promise of ChatGPT”, The New York Times, 8th March 2023.

[FUT2023] “Pause Giant AI Experiments: An Open Letter”, https://futureoflife.org/open-letter/pause-giant-ai-experiments/, 2023.

[NYT2023] The Godfather of A.I.’ Leaves Google and Warns of Danger Ahead, The NY Times, https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html

Bibliography

[OPE2023] OpenAI. “CHATGPT: Optimizing Language Models for Dialogue.” OpenAI, 2 Feb. 2023, https://openai.com/blog/chatgpt/.

[OP2023] OpenAI. “GPT-4 Technical Report.” OpenAI, 15 March 2023

[VAS2017] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[RAD2018] Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. "Improving language understanding by generative pre-training." (2018).

[RAD2019] Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. "Language models are unsupervised multitask learners." OpenAI blog 1, no. 8 (2019): 9.

[BRO2020] Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877-1901.

Bibliography

[BAN2023] Bang, Yejin, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia et al. "A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity." arXiv preprint arXiv:2302.04023 (2023).

[ZHU2015] Y. Zhu et al., "Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 19-27, doi: 10.1109/ICCV.2015.11.

[FOU] The Common Crawl Foundation. Common crawl. URL http://commoncrawl.org.

[ALE2019] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners, 2019.

[JAR2020] Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. 2015 IEEE International Conference on Computer Vision (ICCV), Dec 2015. doi:10.1109/iccv.2015.11.

Bibliography

[CFTE] CHATGPT outperforms 85% of the 4 million programmers assessed on python on linkedin (2022) CFTE. Available at: https://courses.cfte.education/chatgpt-outperforms-4mprogrammers/#:~:text=Some%20of%20its%20abilities%20include,large%20portion%20of%20our%20work

[SCA2022] Scao, T.L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A.S., Yvon, F., Gallé, M. and Tow, J., 2022. Bloom: A 176b-parameter open-access multilingual language model. arXiv preprint arXiv:2211.05100.

[CAH2022] Cahyawijaya, S., Lovenia, H., Aji, A.F., Winata, G.I., Wilie, B., Mahendra, R., Wibisono, C., Romadhony, A., Vincentio, K., Koto, F. and Santoso, J., 2022. NusaCrowd: Open Source Initiative for Indonesian NLP Resources. arXiv preprint arXiv:2212.09648.

Acknowledgements

- This lecture has received funding from the European Union's European Union Horizon 2020 research and innovation programme under grant agreement 951911 (AI4Media).

- It also contains some material from the Horizon Europe AI4Europe CSA (project 101070000) as well as from work funded by AUTH.

- This publication reflects only the authors' views. The European Commission is not responsible for any use that may be made of the information it contains.

Q & A

Thank you very much for your attention!

Contact: Prof. I. Pitas
pitas@csd.auth.gr