Research in AI

Daochang Liu

- 2022 - Now, Postdoc Researcher at University of Sydney
- 2017 - 2022, Ph.D. at Peking University, China
- 2013 - 2017, B.E at Tongji University, China

Research interests:

Generative learning using diffusion models

Computer vision in surgeries

Video understanding and action analysis

Roadmap

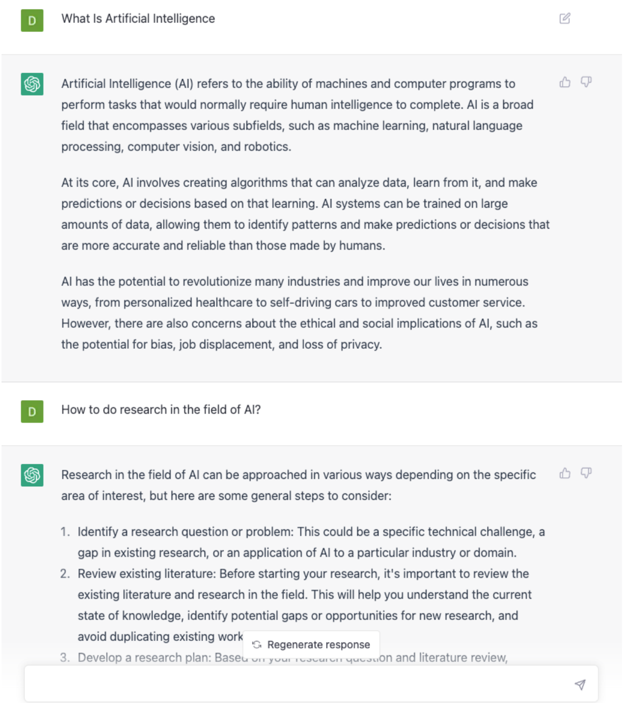
- Introduction to AI and Deep Learning
- Research Methods in AI and Deep Learning
- Recent Large Models and Paradigm Shift

Roadmap

- **Introduction to AI and Deep Learning**
- Research Methods in AI and Deep Learning
- Recent Large Models and Paradigm Shift

Exciting Time …

ChatGPT



D What Is Artificial Intelligence

Artificial Intelligence (AI) refers to the ability of machines and computer programs to perform tasks that would normally require human intelligence to complete. AI is a broad field that encompasses various subfields, such as machine learning, natural language processing, computer vision, and robotics.

At its core, AI involves creating algorithms that can analyze data, learn from it, and make predictions or decisions based on that learning. AI systems can be trained on large amounts of data, allowing them to identify patterns and make predictions or decisions that are more accurate and reliable than those made by humans.

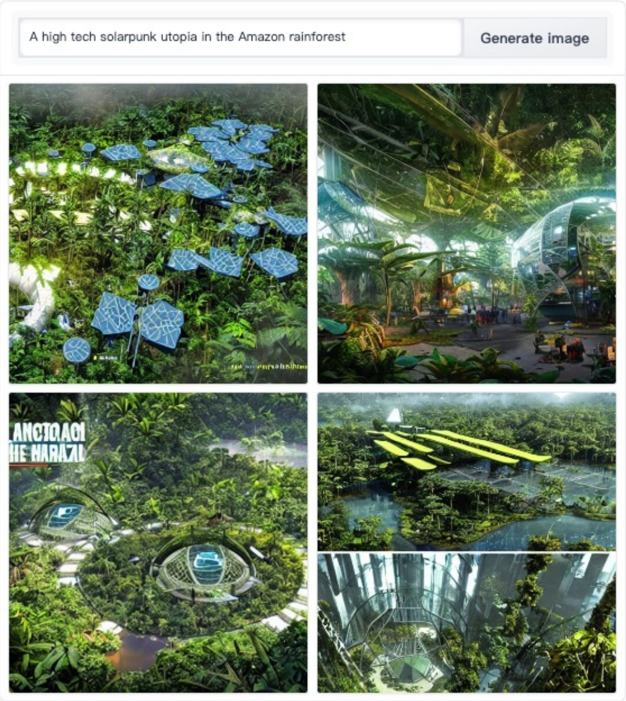
AI has the potential to revolutionize many industries and improve our lives in numerous ways, from personalized healthcare to self-driving cars to improved customer service. However, there are also concerns about the ethical and social implications of AI, such as the potential for bias, job displacement, and loss of privacy.

D How to do research in the field of AI?

Research in the field of AI can be approached in various ways depending on the specific area of interest, but here are some general steps to consider:

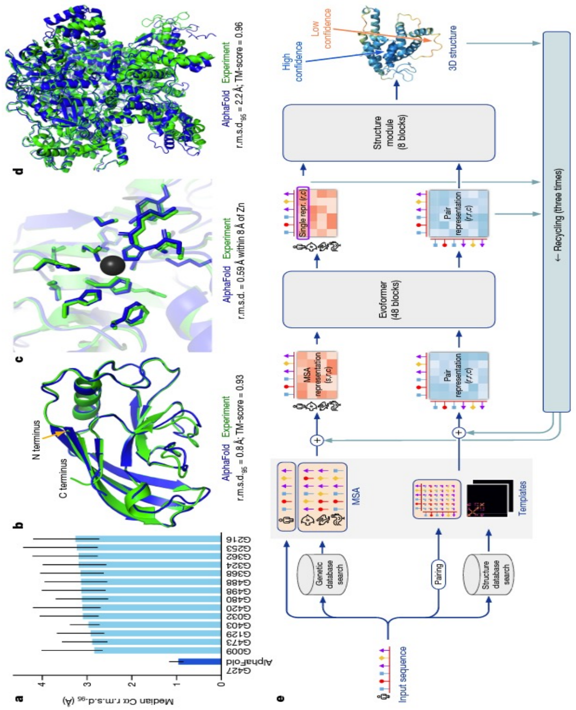
1. Identify a research question or problem: This could be a specific technical challenge, a gap in existing research, or an application of AI to a particular industry or domain.
2. Review existing literature: Before starting your research, it's important to review the existing literature and research in the field. This will help you understand the current state of knowledge, identify potential gaps or opportunities for new research, and avoid duplicating existing work
3. Develop a research plan: Based on your research question and literature review,

Stable Diffusion



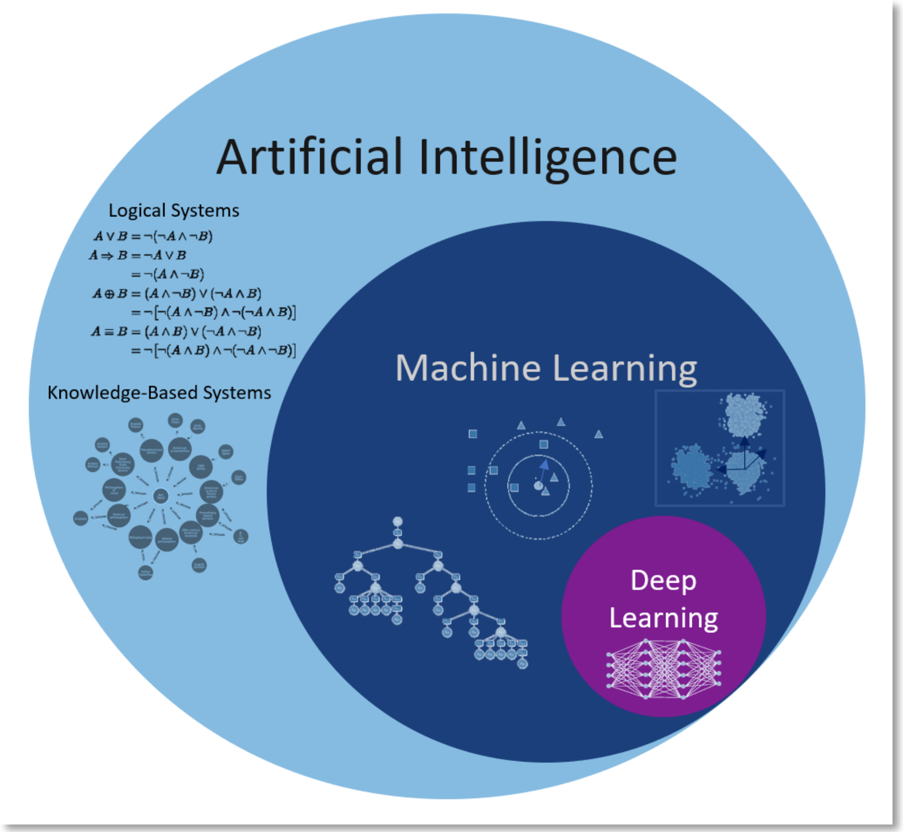
A high tech solarpunk utopia in the Amazon rainforest

AlphaFold



What is Artificial Intelligence (AI)?

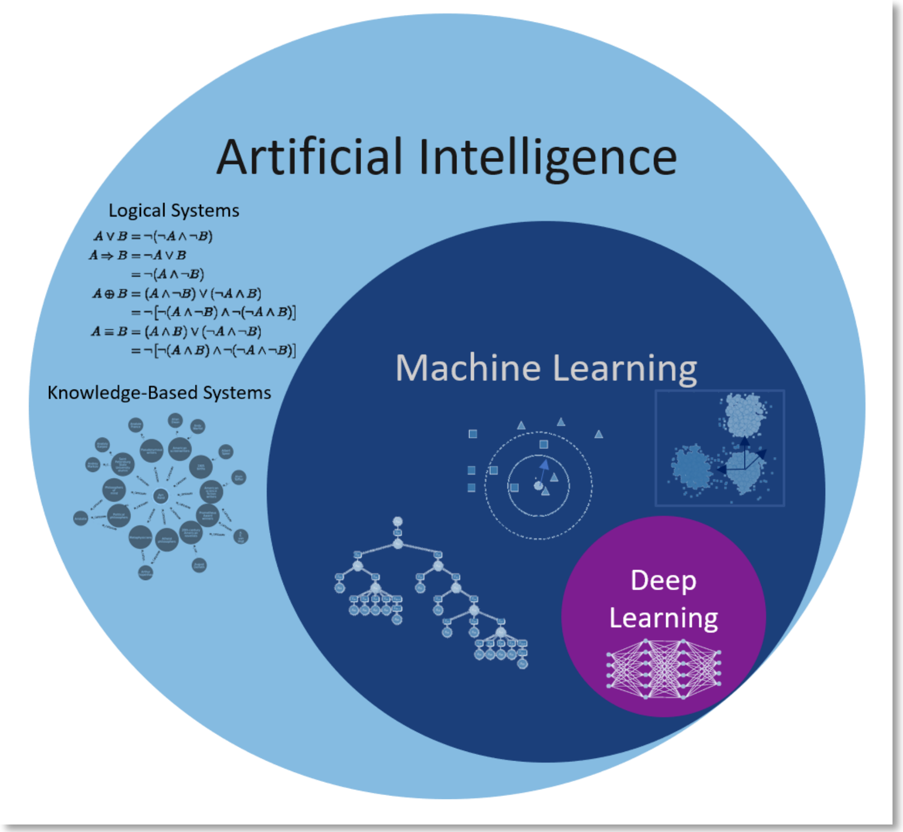
- **Artificial intelligence**
  - Human intelligence exhibited by machines
- **Machine learning**
  - An approach to achieve artificial intelligence
- **Deep learning**
  - A technique for implementing machine learning


Artificial Intelligence
Logical Systems
$A \vee B = \neg(\neg A \wedge \neg B)$
$A \Rightarrow B = \neg A \vee B$
$= \neg(A \wedge \neg B)$
$A \oplus B = (A \wedge \neg B) \vee (\neg A \wedge B)$
$= \neg[\neg(A \wedge \neg B) \wedge \neg(\neg A \wedge B)]$
$A \equiv B = (A \wedge B) \vee (\neg A \wedge \neg B)$
$= \neg[\neg(A \wedge B) \wedge \neg(\neg A \wedge \neg B)]$
Knowledge-Based Systems
Machine Learning
Deep
Learning

What is Artificial Intelligence (AI)?

***ChatGPT**: "Artificial Intelligence (AI) refers to the ability of machines to perform tasks that would normally require human intelligence to complete"*

*Artificial Intelligence (AI) refers to the ability of machines to mimic or surpass human **intelligence in different ways** by **learning from different sources.***


Artificial Intelligence
Logical Systems
$A \vee B = \neg(\neg A \wedge \neg B)$
$A \Rightarrow B = \neg A \vee B$
$= \neg(A \wedge \neg B)$
$A \oplus B = (A \wedge \neg B) \vee (\neg A \wedge B)$
$= \neg[\neg(A \wedge \neg B) \wedge \neg(\neg A \wedge B)]$
$A \equiv B = (A \wedge B) \vee (\neg A \wedge \neg B)$
$= \neg[\neg(A \wedge B) \wedge \neg(\neg A \wedge \neg B)]$
Knowledge-Based Systems
Machine Learning
Deep Learning

Intelligence from Different Sources

Human Learning

Learn with existing knowledge

Find patterns without guidance

Learn from iteraction with environment

…

Intelligence from Different Sources

Human Learning		**Machine Learning**
Learn with existing knowledge	→	Supervised learning
Find patterns without guidance	→	Unsupervised learning
Learn from iteraction with environment	→	Reinforcement learning
…		…

Intelligence in Different Ways

Human Learning

Learn by asking questions

Learn with only a few examples

Learn how to learn better

Update your knowledge over time

Apply learned knowledge to new cases

…

Intelligence in Different Ways

Human Learning	**Machine Learning**
Learn by asking questions	Active learning
Learn with only a few examples	Few-shot learning
Learn how to learn better	Meta learning
Update your knowledge over time	Continuous learning
Apply learned knowledge to new cases	Transfer learning
…	…

Intelligence in Different Ways

Natural Language Processing

Read like human

Computer Vision

See like human

Robotics

Act like human

Artifical Intelligence

Machine Learning

Think like human

Audio and Speech

Listen/Speak like human

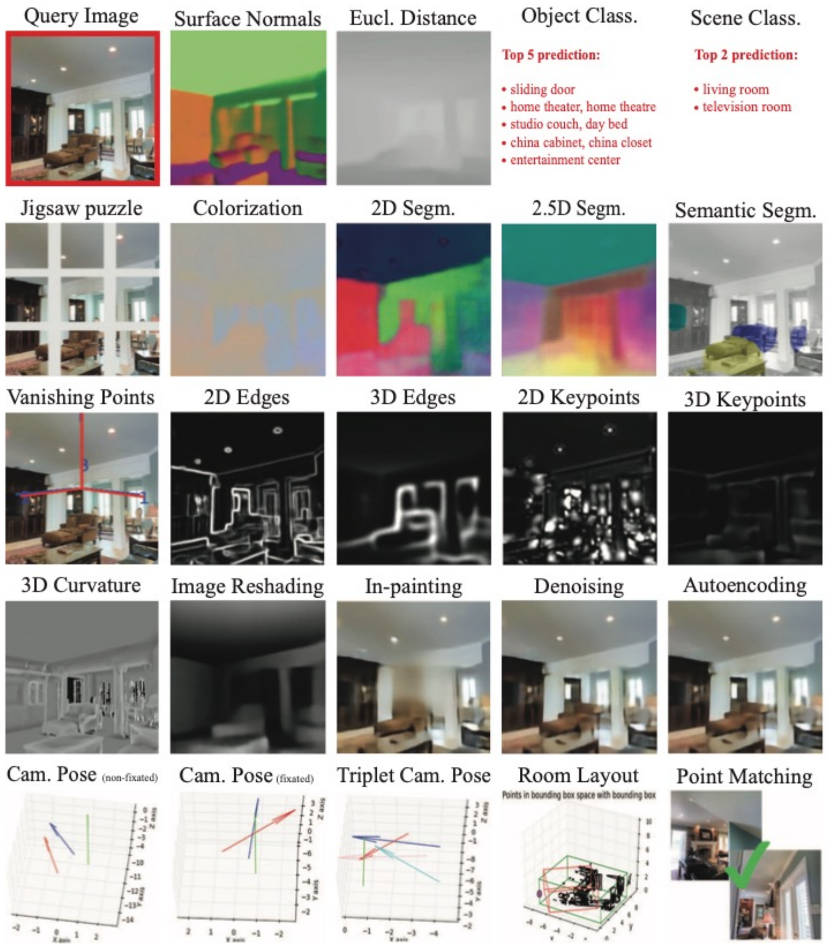
Topics in Computer Vision

High-level tasks:

- Image classification
- Object detection
- Semantic segmentation
- Image captioning
- …

Low-level tasks:

- Super-resolution
- Denosing
- Depth estimation
- …



[Taskonomy: Disentangling Task Transfer Learning]

Topics in Natural Language Processing

`Token1 Token2 Token3` → `Label`

Document classification, Sentiment Analysis, …

`Token1 Token2 Token3` → `Label1 Label2 Label3`

Sentence Tagging, Named Entity Recognition, …

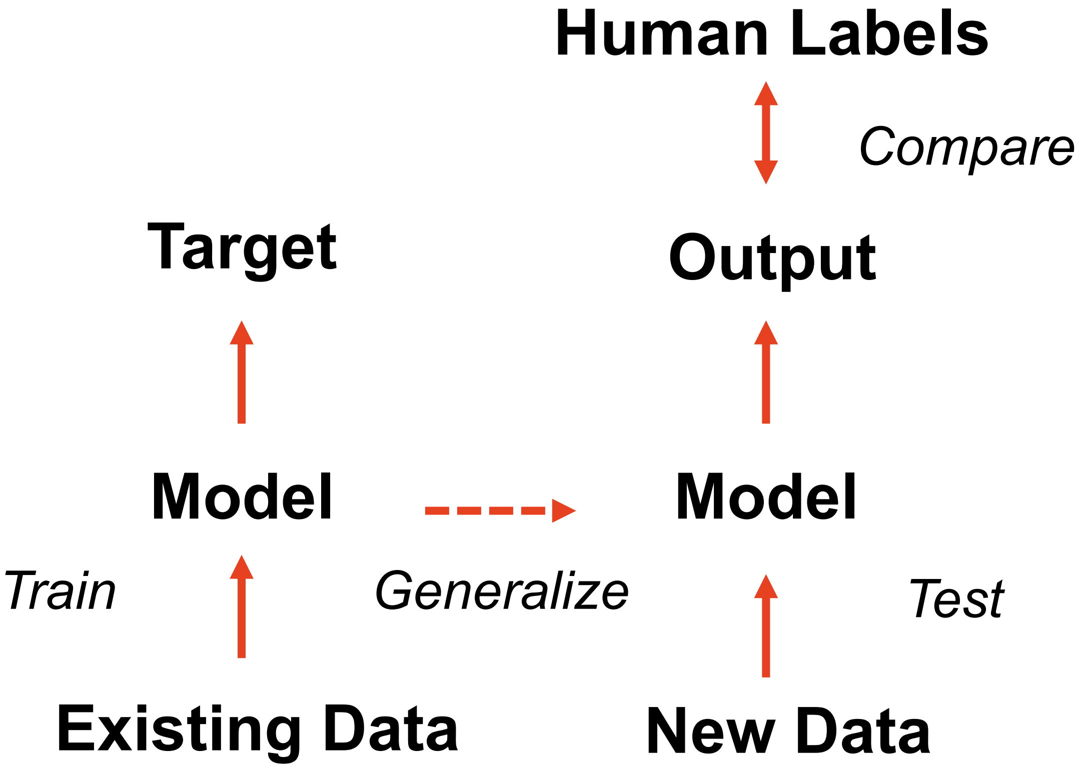
`Token1 Token2 Token3` → `Token4 Token5 Token6`

Machine Translation, Text Summarization, …

`Token1 Token2 Token3`
`Token1 Token2 Token3` → `Label`

Natural Language Inference, Extraction-Based Question Answering, …

A Common Pipeline



How to process the data

How to design the model

How to define the target

How to optimize the model

Roadmap

- Introduction to AI and Deep Learning
- **Research Methods in AI and Deep Learning**
- Recent Large Models and Paradigm Shift

Research Methods from My Experience

- An AI Research Cycle
- Think Like a Machine
- Find a Topic on Grid

Research Methods from My Experience

- **An AI Research Cycle**
- Think Like a Machine
- Find a Topic on Grid

An AI Research Cycle

1. Choose a topic
2. Literature review
3. Identify a baseline
4. Find codes and reproduce baseline
5. Make some improvement

…

6. Debug and tune
7. Spend some GPU hours
8. Debug and tune
9. Spend more GPU hours

…

10. Design experiments
11. Write a paper and submit
12. Response or rebuttal
13. Present the outcome, or go back to 11 / 10 / 5 / 3 / 1

An AI Research Cycle

1. Choose a topic: *Trade-offs: impact-competition, significance-risk, novelty-feasibility*
2. Literature review
3. Identify a baseline: *Need to be recent SOTA, well-recognized, adaptable, easy to use*
4. Find codes and reproduce baseline: *Github, PaperWithCodes, Benchmarks*
5. Make some improvement
…
6. Debug and tune
7. Spend some GPU hours
8. Debug and tune
9. Spend more GPU hours
…
10. Design experiments
11. Write a paper and submit: Story-telling, exploration is bottom-up, story is top-down
12. Response or rebuttal: Clarify misunderstandings, provide new information
13. Present the outcome, or go back to 11 / 10 / 5 / 3 / 1

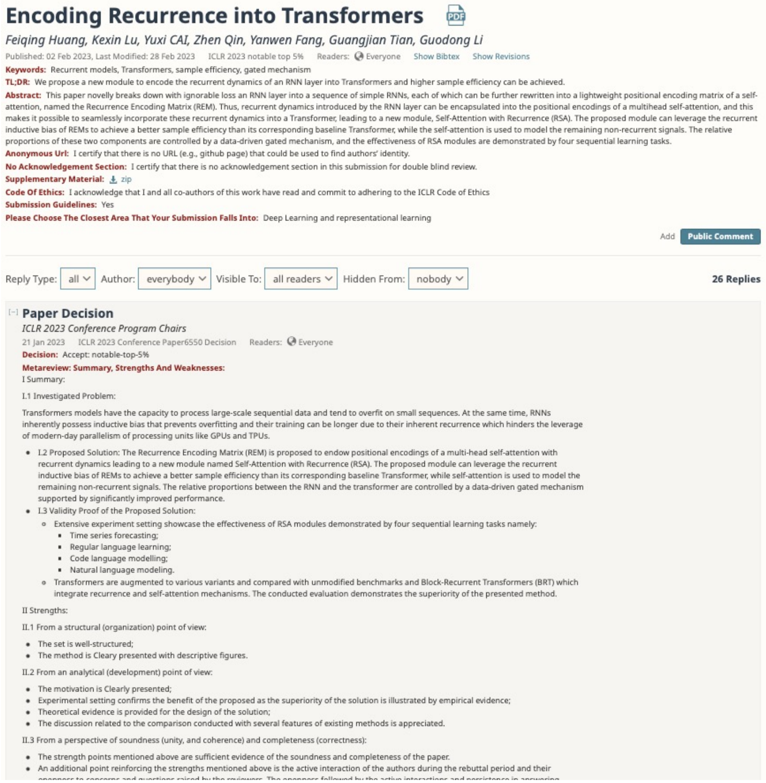
OpenReview

A good learning source

To know how it works

Accepted papers:
Promising ideas
Experiment designs
Successful rebuttal

Rejected papers:
Things to avoid



Encoding Recurrence into Transformers

Feiqing Huang, Kexin Lu, Yuxi CAI, Zhen Qin, Yanwen Fang, Guangjian Tian, Guodong Li

Published: 02 Feb 2023, Last Modified: 28 Feb 2023 ICLR 2023 notable top 5% Readers: Everyone Show Bibtex Show Revisions

Keywords: Recurrent models, Transformers, sample efficiency, gated mechanism

TL;DR: We propose a new module to encode the recurrent dynamics of an RNN layer into Transformers and higher sample efficiency can be achieved.

Abstract: This paper novelly breaks down with ignorable loss an RNN layer into a sequence of simple RNNs, each of which can be further rewritten into a lightweight positional encoding matrix of a self-attention, named the Recurrence Encoding Matrix (REM). Thus, recurrent dynamics introduced by the RNN layer can be encapsulated into the positional encodings of a multihead self-attention, and this makes it possible to seamlessly incorporate these recurrent dynamics into a Transformer, leading to a new module, Self-Attention with Recurrence (RSA). The proposed module can leverage the recurrent inductive bias of REMs to achieve a better sample efficiency than its corresponding baseline Transformer, while the self-attention is used to model the remaining non-recurrent signals. The relative proportions of these two components are controlled by a data-driven gated mechanism, and the effectiveness of RSA modules are demonstrated by four sequential learning tasks.

Anonymous Url: I certify that there is no URL (e.g., github page) that could be used to find authors' identity.

No Acknowledgement Section: I certify that there is no acknowledgement section in this submission for double blind review.

Supplementary Material: zip

Code Of Ethics: I acknowledge that I and all co-authors of this work have read and commit to adhering to the ICLR Code of Ethics

Submission Guidelines: Yes

Please Choose The Closest Area That Your Submission Falls Into: Deep Learning and representational learning

Reply Type: all Author: everybody Visible To: all readers Hidden From: nobody

26 Replies

Paper Decision

ICLR 2023 Conference Program Chairs

21 Jan 2023 ICLR 2023 Conference Paper6550 Decision Readers: Everyone

Decision: Accept: notable-top-5%

Metareview: Summary, Strengths And Weaknesses:

I Summary:

I.1 Investigated Problem:

Transformers models have the capacity to process large-scale sequential data and tend to overfit on small sequences. At the same time, RNNs inherently possess inductive bias that prevents overfitting and their training can be longer due to their inherent recurrence which hinders the leverage of modern-day parallelism of processing units like GPUs and TPUs.

- I.2 Proposed Solution: The Recurrence Encoding Matrix (REM) is proposed to endow positional encodings of a multi-head self-attention with recurrent dynamics leading to a new module named Self-Attention with Recurrence (RSA). The proposed module can leverage the recurrent inductive bias of REMs to achieve a better sample efficiency than its corresponding baseline Transformer, while self-attention is used to model the remaining non-recurrent signals. The relative proportions between the RNN and the transformer are controlled by a data-driven gated mechanism supported by significantly improved performance.
- I.3 Validity Proof of the Proposed Solution:
  - Extensive experiment setting showcase the effectiveness of RSA modules demonstrated by four sequential learning tasks namely:
    - Time series forecasting;
    - Regular language learning;
    - Code language modelling;
    - Natural language modeling.
  - Transformers are augmented to various variants and compared with unmodified benchmarks and Block-Recurrent Transformers (BRT) which integrate recurrence and self-attention mechanisms. The conducted evaluation demonstrates the superiority of the presented method.

II Strengths:

II.1 From a structural (organization) point of view:

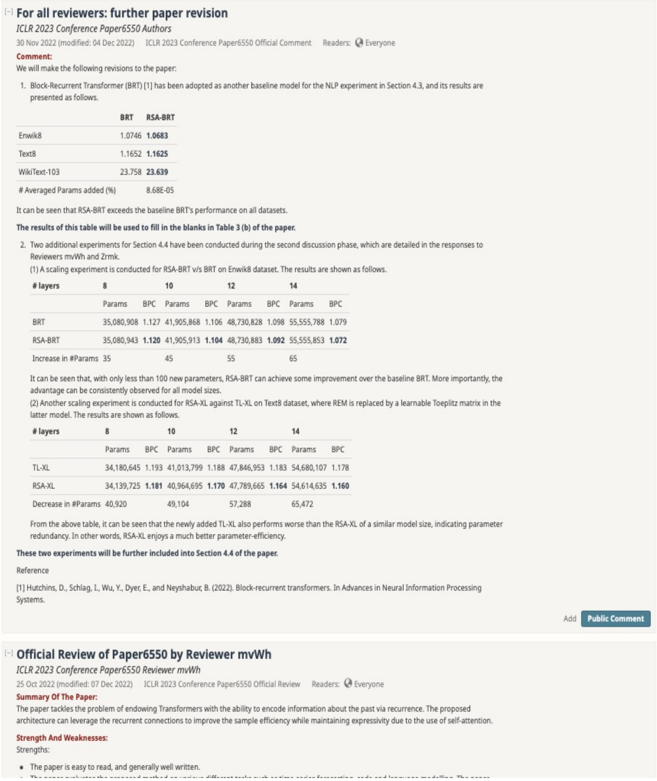
- The set is well-structured;
- The method is Cleary presented with descriptive figures.

II.2 From an analytical (development) point of view:

- The motivation is Clearly presented;
- Experimental setting confirms the benefit of the proposed as the superiority of the solution is illustrated by empirical evidence;
- Theoretical evidence is provided for the design of the solution;
- The discussion related to the comparison conducted with several features of existing methods is appreciated.

II.3 From a perspective of soundness (unity, and coherence) and completeness (correctness):

- The strength points mentioned above are sufficient evidence of the soundness and completeness of the paper.
- An additional point reinforcing the strengths mentioned above is the active interaction of the authors during the rebuttal period and their



For all reviewers: further paper revision

ICLR 2023 Conference Paper6550 Authors

30 Nov 2022 (modified: 04 Dec 2022) ICLR 2023 Conference Paper6550 Official Comment Readers: Everyone

Comment:

We will make the following revisions to the paper:

1. Block-Recurrent Transformer (BRT) [1] has been adopted as another baseline model for the NLP experiment in Section 4.3, and its results are presented as follows.

	BRT	RSA-BRT
Enwik8	1.0746	**1.0683**
Text8	1.1652	**1.1625**
WikiText-103	23.758	**23.639**
# Averaged Params added (%)		8.68E-05

It can be seen that RSA-BRT exceeds the baseline BRT's performance on all datasets.

The results of this table will be used to fill in the blanks in Table 3 (b) of the paper.

2. Two additional experiments for Section 4.4 have been conducted during the second discussion phase, which are detailed in the responses to Reviewers mvWh and Zrmk.

(1) A scaling experiment is conducted for RSA-BRT v/s BRT on Enwik8 dataset. The results are shown as follows.

# layers	8		10		12		14	
	Params	BPC	Params	BPC	Params	BPC	Params	BPC
BRT	35,080,908	1.127	41,905,868	1.106	48,730,828	1.098	55,555,788	1.079
RSA-BRT	35,080,943	**1.120**	41,905,913	**1.104**	48,730,883	**1.092**	55,555,853	**1.072**
Increase in #Params	35		45		55		65	

It can be seen that, with only less than 100 new parameters, RSA-BRT can achieve some improvement over the baseline BRT. More importantly, the advantage can be consistently observed for all model sizes.

(2) Another scaling experiment is conducted for RSA-XL against TL-XL on Text8 dataset, where REM is replaced by a learnable Toeplitz matrix in the latter model. The results are shown as follows.

# layers	8		10		12		14	
	Params	BPC	Params	BPC	Params	BPC	Params	BPC
TL-XL	34,180,645	1.193	41,013,799	1.188	47,846,953	1.183	54,680,107	1.178
RSA-XL	34,139,725	**1.181**	40,964,695	**1.170**	47,789,665	**1.164**	54,614,635	**1.160**
Decrease in #Params	40,920		49,104		57,288		65,472	

From the above table, it can be seen that the newly added TL-XL also performs worse than the RSA-XL of a similar model size, indicating parameter redundancy. In other words, RSA-XL enjoys a much better parameter-efficiency.

These two experiments will be further included into Section 4.4 of the paper.

Reference

[1] Hutchins, D., Schlag, I., Wu, Y., Dyer, E., and Neyshabur, B. (2022). Block-recurrent transformers. In Advances in Neural Information Processing Systems.

Official Review of Paper6550 by Reviewer mvWh

ICLR 2023 Conference Paper6550 Reviewer mvWh

25 Oct 2022 (modified: 07 Dec 2022) ICLR 2023 Conference Paper6550 Official Review Readers: Everyone

Summary Of The Paper:

The paper tackles the problem of endowing Transformers with the ability to encode information about the past via recurrence. The proposed architecture can leverage the recurrent connections to improve the sample efficiency while maintaining expressivity due to the use of self-attention.

Strength And Weaknesses:

Strengths:

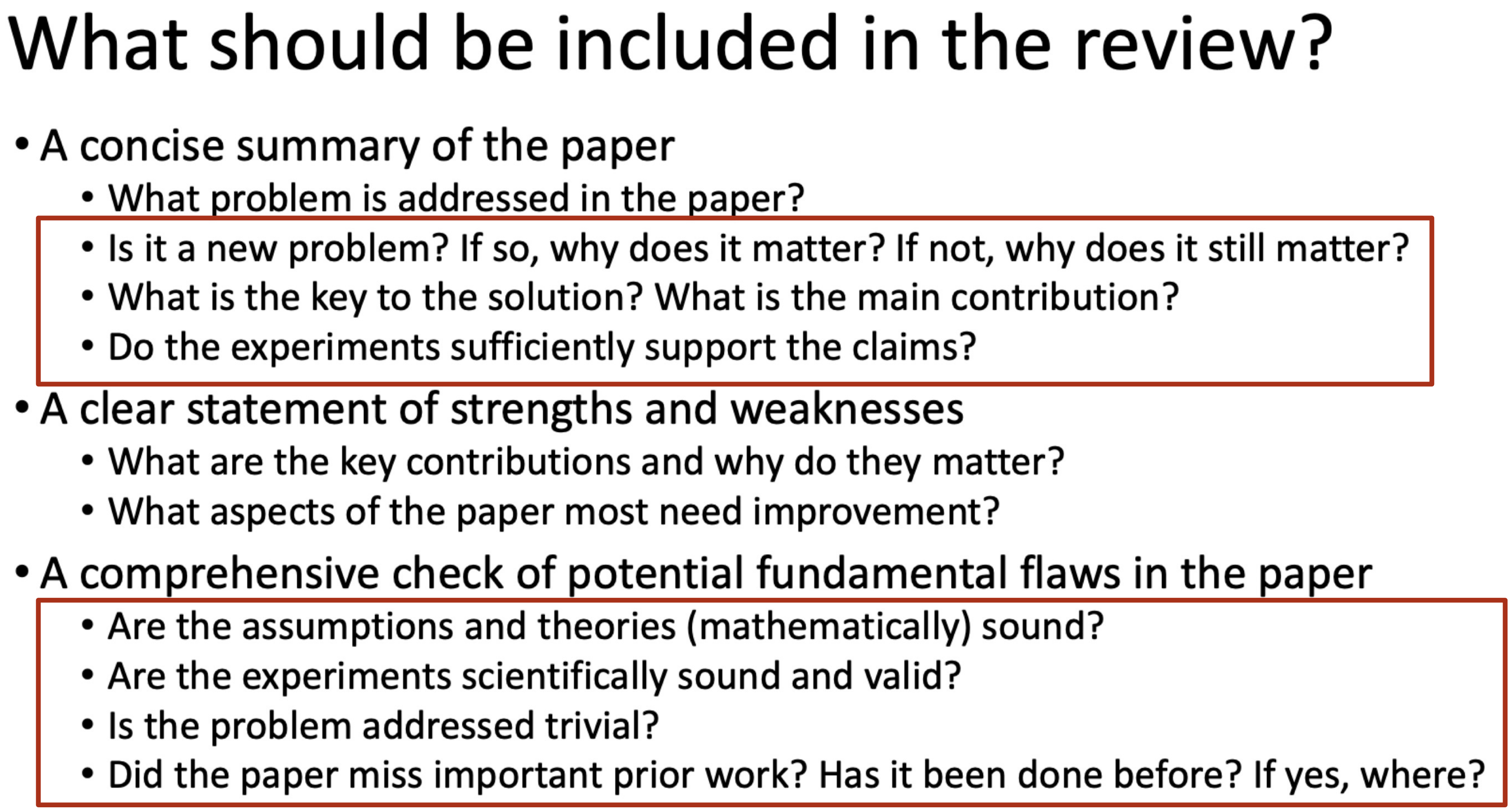
- The paper is easy to read, and generally well written.

Look at Reviewer Guideline

Reviewer guidelines of conferences or journals are also good learning sources

- CVPR
- NeurIPS
- MICCAI

CVPR Reviewer Guideline



What should be included in the review?

- A concise summary of the paper
  - What problem is addressed in the paper?
  - Is it a new problem? If so, why does it matter? If not, why does it still matter?
  - What is the key to the solution? What is the main contribution?
  - Do the experiments sufficiently support the claims?
- A clear statement of strengths and weaknesses
  - What are the key contributions and why do they matter?
  - What aspects of the paper most need improvement?
- A comprehensive check of potential fundamental flaws in the paper
  - Are the assumptions and theories (mathematically) sound?
  - Are the experiments scientifically sound and valid?
  - Is the problem addressed trivial?
  - Did the paper miss important prior work? Has it been done before? If yes, where?

NeurIPS Reviewer Guideline

Originality: Are the tasks or methods new? Is the work a novel combination of well-known techniques? (This can be valuable!) Is it clear how this work differs from previous contributions? Is related work adequately cited?
Quality: Is the submission technically sound? Are claims well supported (e.g., by theoretical analysis or experimental results)? Are the methods used appropriate? Is this a complete piece of work or work in progress? Are the authors careful and honest about evaluating both the strengths and weaknesses of their work?
Clarity: Is the submission clearly written? Is it well organized? (If not, please make constructive suggestions for improving its clarity.) Does it adequately inform the reader? (Note that a superbly written paper provides enough information for an expert reader to reproduce its results.)
Significance: Are the results important? Are others (researchers or practitioners) likely to use the ideas or build on them? Does the submission address a difficult task in a better way than previous work? Does it advance the state of the art in a demonstrable way? Does it provide unique data, unique conclusions about existing data, or a unique theoretical or experimental approach?

https://neurips.cc/Conferences/2021/PaperInformation/PaperChecklist
https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf

for self-check

MICCAI Reviewer Guideline

MIC-based papers: **When reviewing MIC based MICCAI papers, we would like to see:**
whether the proposed methods are innovative or
whether the application is innovative.

In particular the following questions should be asked when evaluating MIC-based papers:
Is the topic of paper clinically significant?
Do the authors clearly explain data collection, processing, and division methods?
Do the data appropriately represent the range of possible patients and disease manifestations?
Are the data labels (if applicable) of sufficient quality to support the claimed performance of the algorithms?
Do the authors report a sufficient number and type of performance measures to accurately represent strengths and weaknesses of the algorithms? Are performance measures reported with confidence intervals?
Are the results and comparison with prior art placed in the context of a clinical application in terms of significance and impact? Have they performed a proper statistical significance analysis of results?
Does the work make a significant contribution to the field or the society, or is it just incremental over previous work?
Do the authors discuss limitations of their methods and directions for future research?

Different conferences or journals have different tastes

Research Methods from My Experience

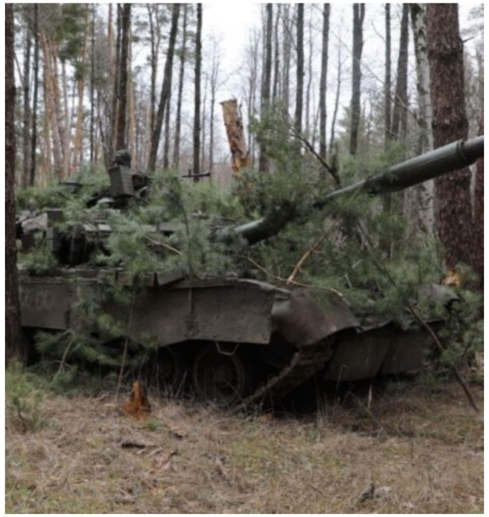
- An AI Research Cycle
- **Think Like a Machine**
- Find a Topic on Grid

Think Like a Machine

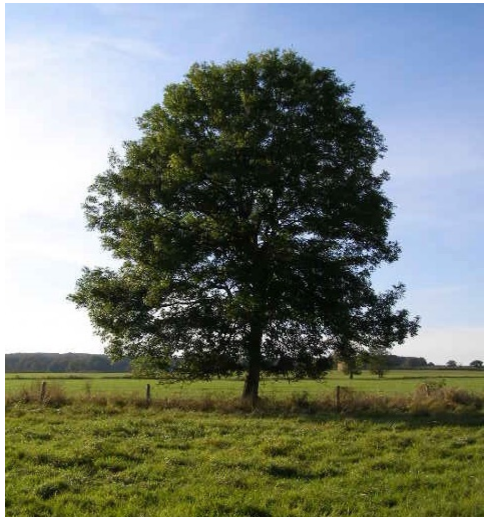
Machine may behave not as you expect.

Think Like a Machine

Machine may behave not as you expect.



Source



Source

What you expect:

Tank vs. No Tank

What the machine learn:

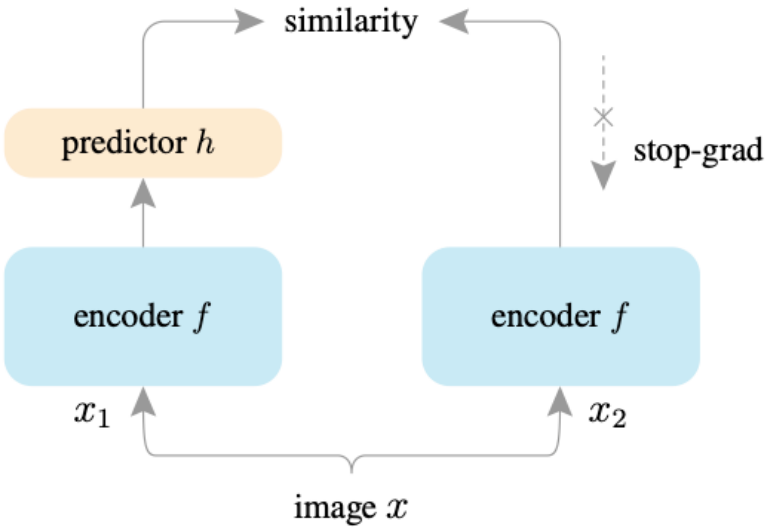
Cloudy vs. Sunny

Many Trees vs. Single Tree

…

Think Like a Machine

Machine may behave not as you expect.



[SimSiam for Self-Supervised Representation Learning](#)

What you expect:

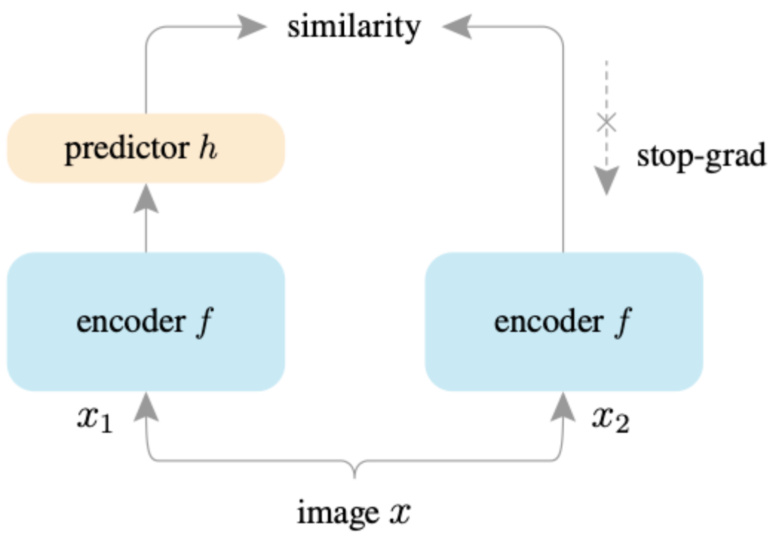
To make augmentations of the same image have similar feature representations

What the machine learn:

Output an all-zero feature all the time to take a shortcut to the learning target (Collapse)

Think Like a Machine

Machine may behave not as you expect.



SimSiam for Self-Supervised Representation Learning

What you expect:

To make augmentations of the same image have similar feature representations

What the machine learn:

Output an all-zero feature all the time to take a shortcut to the learning target (Collapse)

If you want to make the machine think like human, you need to make yourself think like the machine first.

Think Like a Machine

Imagine yourself as the AI model you are training.

- Visualize what the model sees (Input and Intermediate Results)
- How would you achieve the learning target if you are the model? Any shortcut?
- Test on some toy data or mental experiments
- Unexpected biases in the predictions?

Think Like a Machine

AI researchers are the translator between our natural mind and the digital mind

Formulate your expectations / domain knowledge in a computational way

Think Like a Machine

AI researchers are the translator between our natural mind and the digital mind

Formulate your expectations / domain knowledge in a computational way

Goal in *natural language*

Make two images more similar

Source

Source

Think Like a Machine

AI researchers are the translator between our natural mind and the digital mind

Formulate your expectations / domain knowledge in a computational way

Goal in *natural language*

Make two images more similar

Source

Source

What does 'similar' means?

Similar objects

Similar style

Similar colors

…

Think Like a Machine

AI researchers are the translator between our natural mind and the digital mind

Formulate your expectations / domain knowledge in a computational way

Goal in *natural language*

Make two images more similar

Source

Source

What does ‘similar’ means?

Similar objects

Similar style

Similar colors

…

Translated to

computational language

Perceptual loss

Gram matrix

Color histograms

…

Think Like a Machine

AI researchers are the *translator* between our natural mind and the digital mind

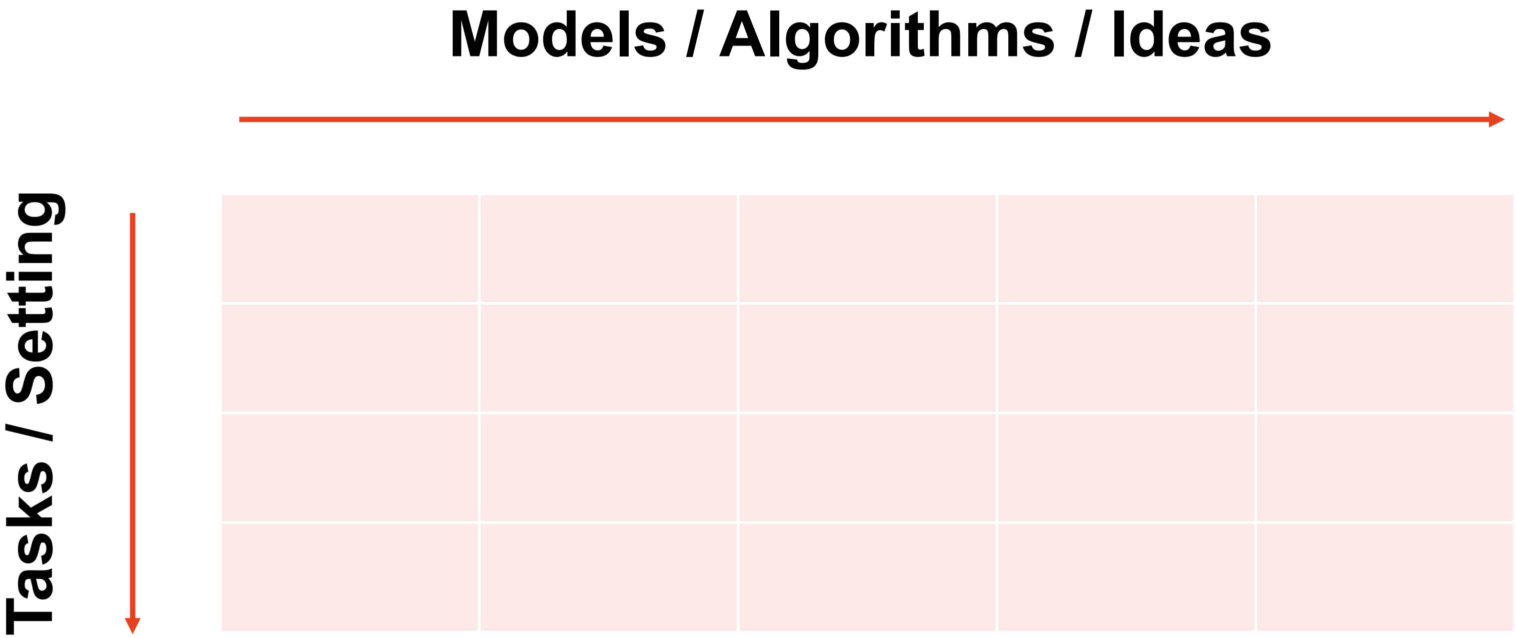
- **When interacting with machine, think like the machine.**
  - Designing the model, Debugging, Experiments, Results Analysis, …

- **When interacting with people, think like human.**
  - Reading a paper, Presenting your work, …
  - Focus more on intuitive interpretations and physical meanings

Research Methods from My Experience

- An AI Research Cycle
- Think Like a Machine
- **Find a Topic on Grid**

Find a Topic on Grid

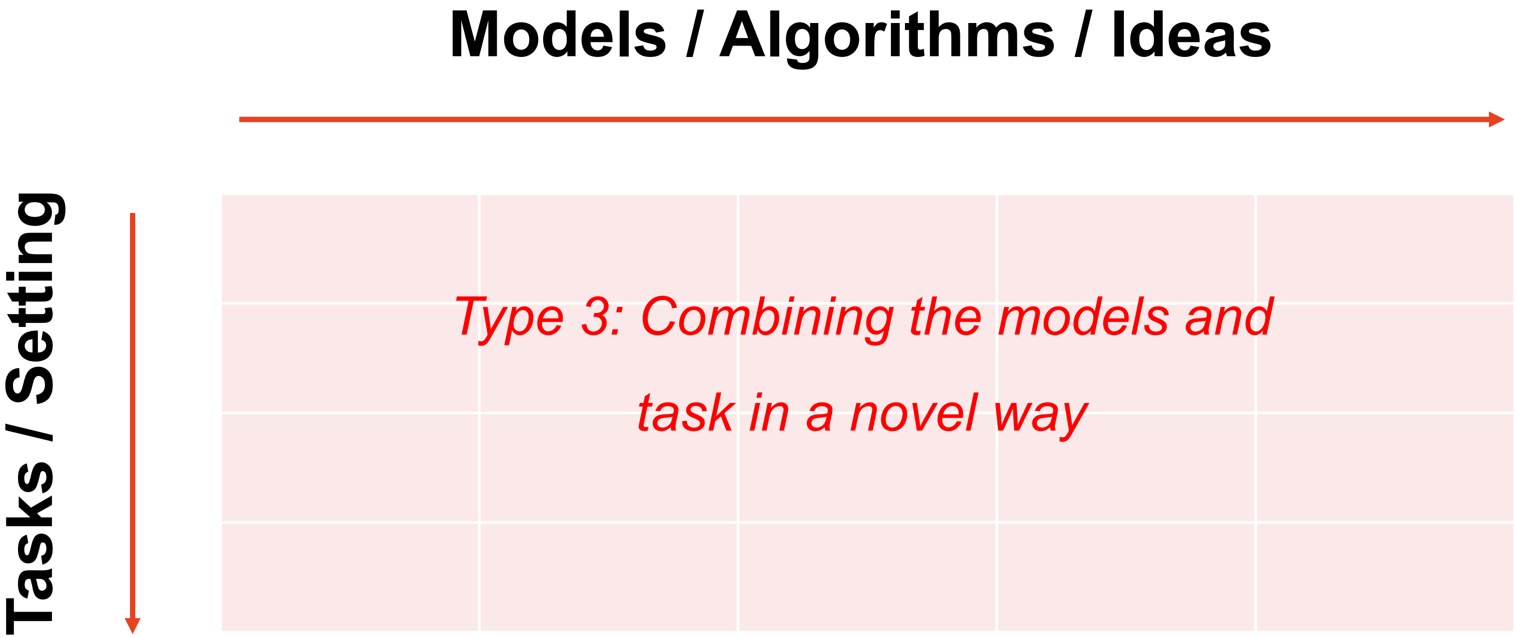


Models / Algorithms / Ideas

Tasks / Setting

Find a Topic on Grid

Type 1: Propose new models / algorithms



Models / Algorithms / Ideas

Tasks / Setting

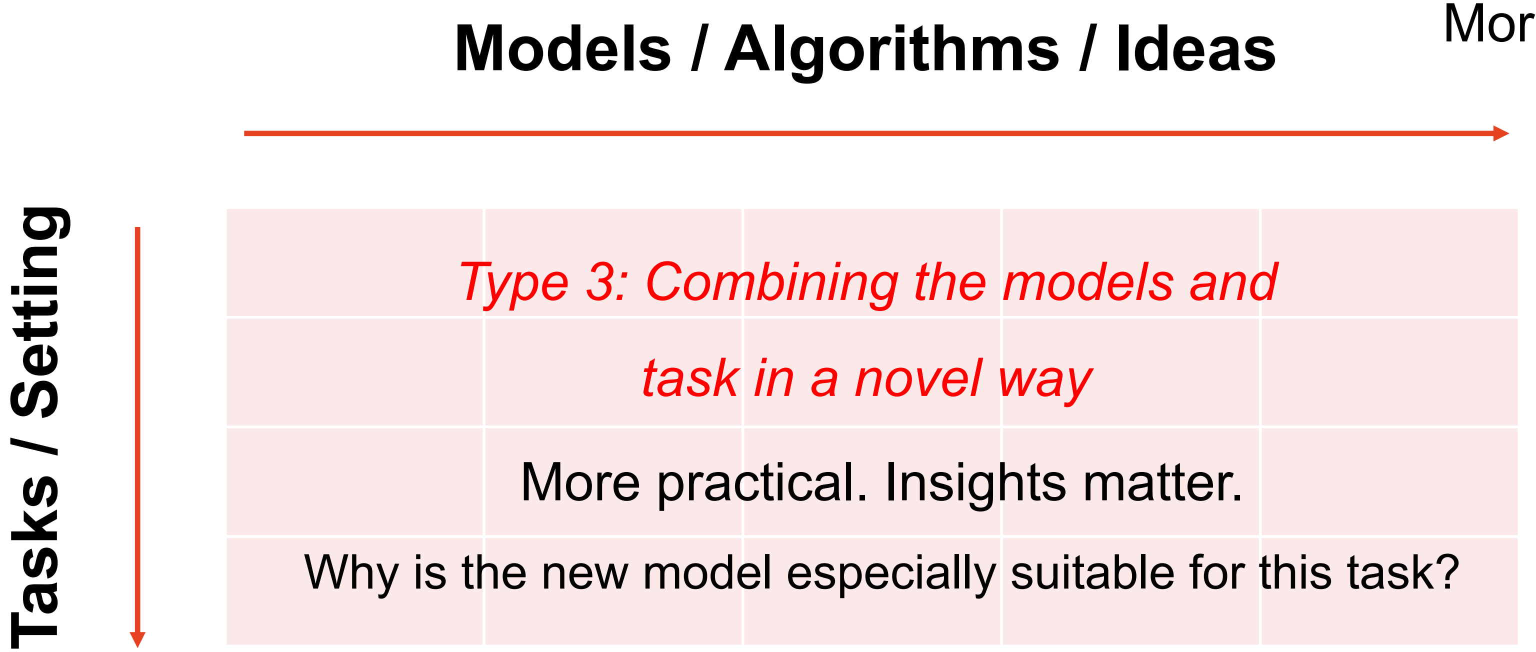
Type 3: Combining the models and task in a novel way

Type 2: Identify new tasks

Find a Topic on Grid

Type 1: Propose new models / algorithms

More significant



Models / Algorithms / Ideas

Tasks / Setting

Type 3: Combining the models and task in a novel way

More practical. Insights matter.

Why is the new model especially suitable for this task?

Type 2: Identify new tasks

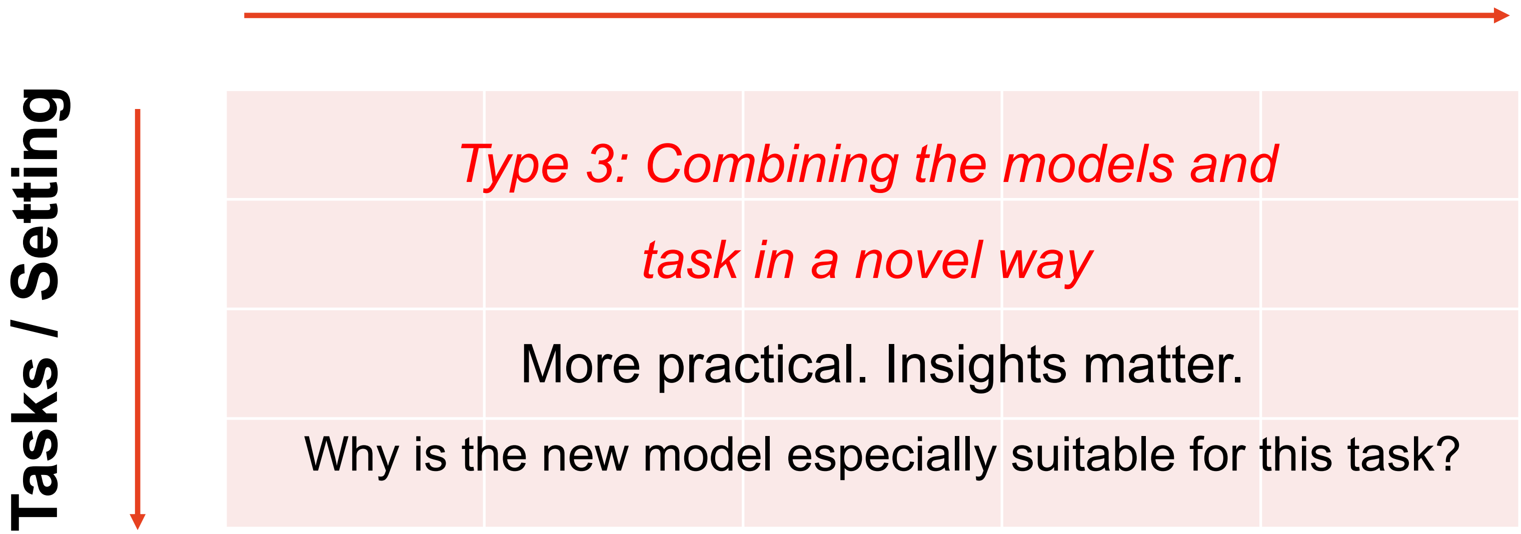
More significant: People usually care more about what you do rather than how you do

Find a Topic on Grid

Type 1: Propose new models / algorithms

More significant

Models / Algorithms / Ideas



Tasks / Setting

Type 3: Combining the models and task in a novel way

More practical. Insights matter.

Why is the new model especially suitable for this task?

Type 2: Identify new tasks

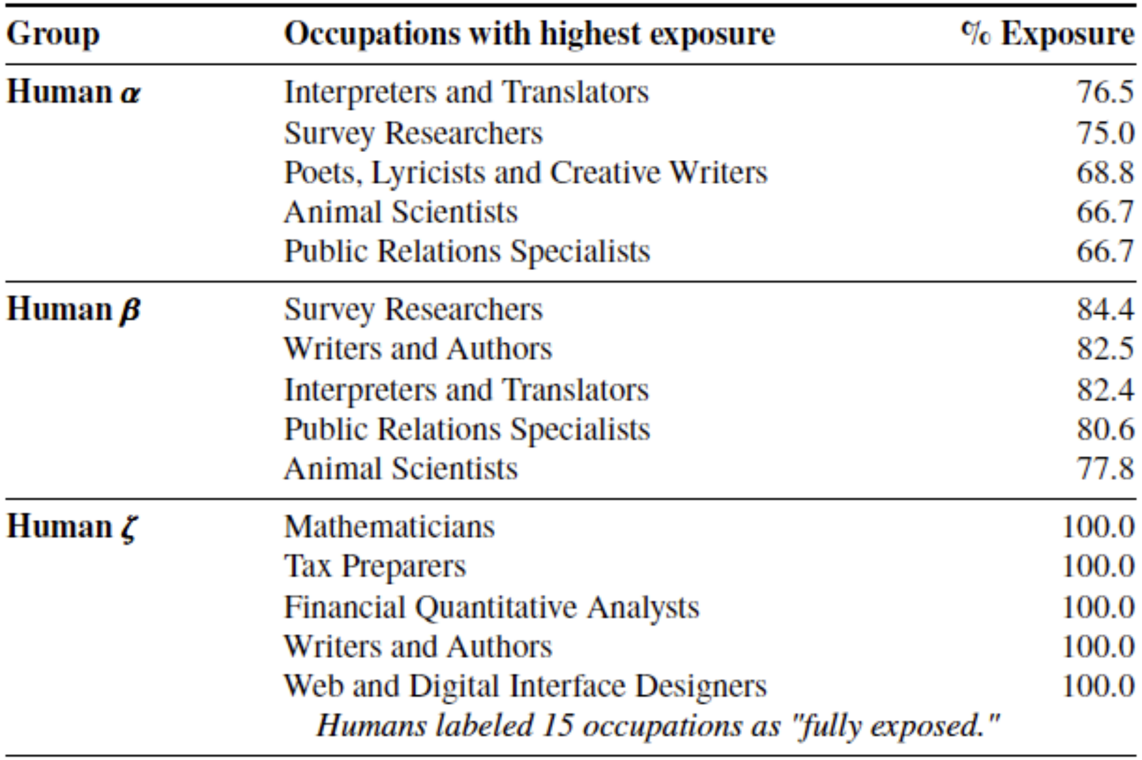
More significant

Suggestion: Read more papers outside your field to extend the grid

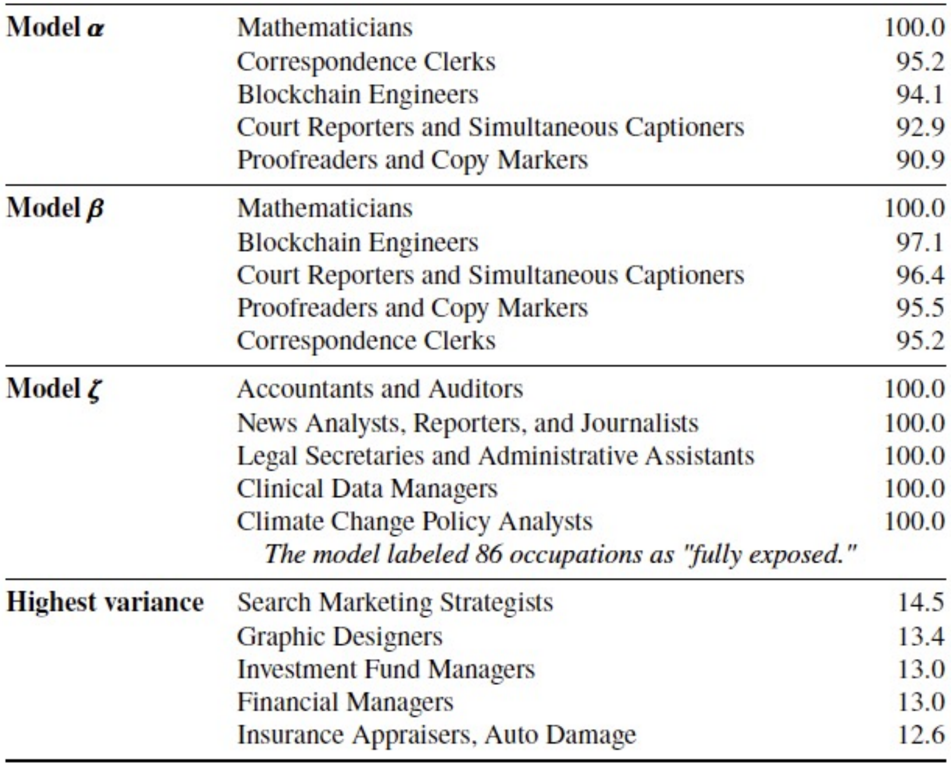
Roadmap

- Introduction to AI and Deep Learning
- Research Methods in AI and Deep Learning
- **Recent Large Models and Paradigm Shift**

Labor Market Impact of Large Models (GPTs)



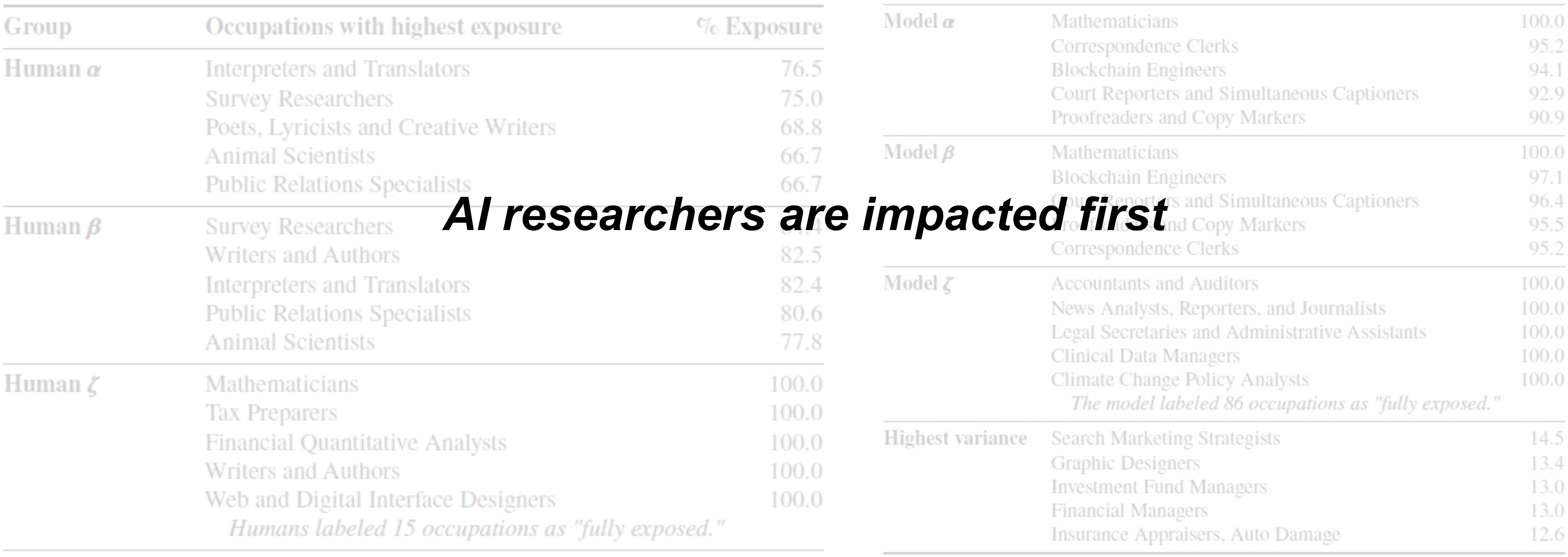
Group	Occupations with highest exposure	% Exposure
Human α	Interpreters and Translators	76.5
	Survey Researchers	75.0
	Poets, Lyricists and Creative Writers	68.8
	Animal Scientists	66.7
	Public Relations Specialists	66.7
Human β	Survey Researchers	84.4
	Writers and Authors	82.5
	Interpreters and Translators	82.4
	Public Relations Specialists	80.6
	Animal Scientists	77.8
Human ζ	Mathematicians	100.0
	Tax Preparers	100.0
	Financial Quantitative Analysts	100.0
	Writers and Authors	100.0
	Web and Digital Interface Designers	100.0
	Humans labeled 15 occupations as "fully exposed."	



Model α	Mathematicians	100.0
	Correspondence Clerks	95.2
	Blockchain Engineers	94.1
	Court Reporters and Simultaneous Captioners	92.9
	Proofreaders and Copy Markers	90.9
Model β	Mathematicians	100.0
	Blockchain Engineers	97.1
	Court Reporters and Simultaneous Captioners	96.4
	Proofreaders and Copy Markers	95.5
	Correspondence Clerks	95.2
Model ζ	Accountants and Auditors	100.0
	News Analysts, Reporters, and Journalists	100.0
	Legal Secretaries and Administrative Assistants	100.0
	Clinical Data Managers	100.0
	Climate Change Policy Analysts	100.0
	The model labeled 86 occupations as "fully exposed."	
Highest variance	Search Marketing Strategists	14.5
	Graphic Designers	13.4
	Investment Fund Managers	13.0
	Financial Managers	13.0
	Insurance Appraisers, Auto Damage	12.6

An Early Look at the Labor Market Impact Potential of Large Language Models

Labor Market Impact of Large Models (GPTs)



Group	Occupations with highest exposure	% Exposure
Human α	Interpreters and Translators	76.5
	Survey Researchers	75.0
	Poets, Lyricists and Creative Writers	68.8
	Animal Scientists	66.7
	Public Relations Specialists	66.7
Human β	Survey Researchers	84.4
	Writers and Authors	82.5
	Interpreters and Translators	82.4
	Public Relations Specialists	80.6
	Animal Scientists	77.8
Human ζ	Mathematicians	100.0
	Tax Preparers	100.0
	Financial Quantitative Analysts	100.0
	Writers and Authors	100.0
	Web and Digital Interface Designers	100.0
	Humans labeled 15 occupations as "fully exposed."	

Model α	Mathematicians	100.0
	Correspondence Clerks	95.2
	Blockchain Engineers	94.1
	Court Reporters and Simultaneous Captioners	92.9
	Proofreaders and Copy Markers	90.9
Model β	Mathematicians	100.0
	Blockchain Engineers	97.1
	Court Reporters and Simultaneous Captioners	96.4
	Proofreaders and Copy Markers	95.5
	Correspondence Clerks	95.2
Model ζ	Accountants and Auditors	100.0
	News Analysts, Reporters, and Journalists	100.0
	Legal Secretaries and Administrative Assistants	100.0
	Clinical Data Managers	100.0
	Climate Change Policy Analysts	100.0
	The model labeled 86 occupations as "fully exposed."	
Highest variance	Search Marketing Strategists	14.5
	Graphic Designers	13.4
	Investment Fund Managers	13.0
	Financial Managers	13.0
	Insurance Appraisers, Auto Damage	12.6

AI researchers are impacted first

An Early Look at the Labor Market Impact Potential of Large Language Models

Paradigm Shift in AI Research

- **One-by-one** to **All-in-one**
- **Model-centric** to **Computation-centric**
- **Decentralized** to **Centralized**

Paradigm Shift in AI Research

- **One-by-one** to **All-in-one**
- **Model-centric** to **Computation-centric**
- **Decentralized** to **Centralized**

This has happened for languages,
and is happening for images and multi-modality research

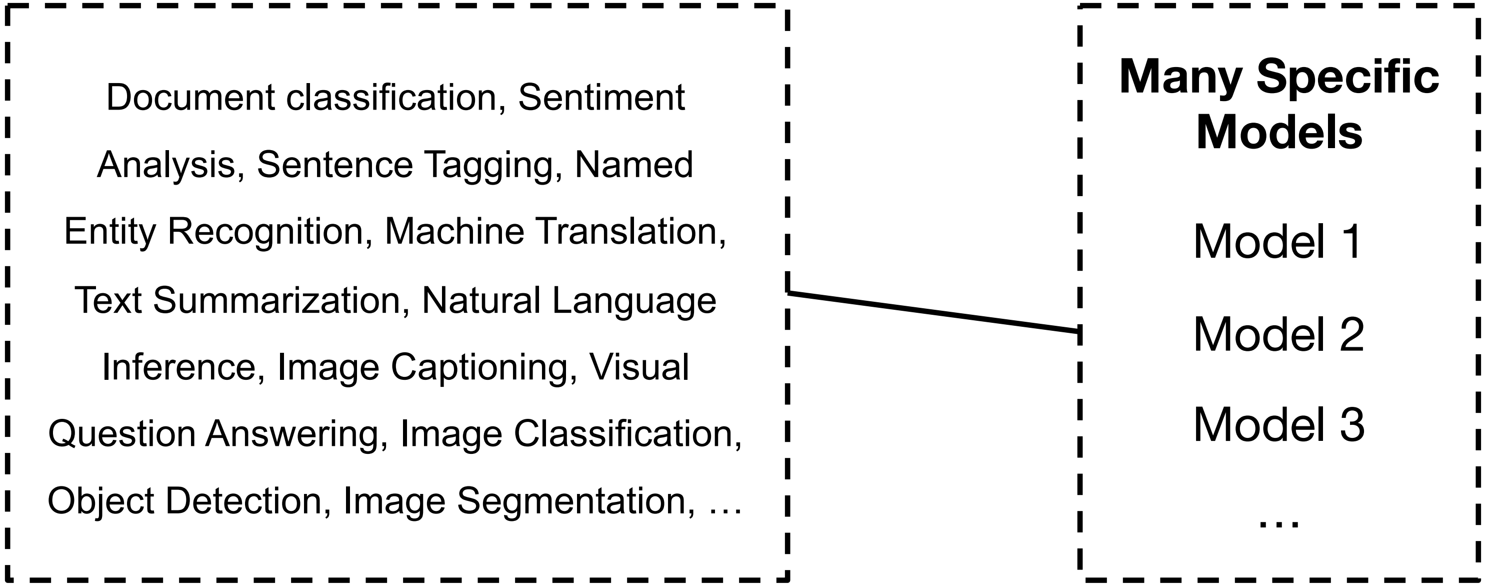
One-by-one to All-in-one

Many different tasks in CV / NLP / ML

Document classification, Sentiment Analysis, Sentence Tagging, Named Entity Recognition, Machine Translation, Text Summarization, Natural Language Inference, Image Captioning, Visual Question Answering, Image Classification, Object Detection, Image Segmentation, …

One-by-one to All-in-one

Many different tasks in CV / NLP / ML



Document classification, Sentiment Analysis, Sentence Tagging, Named Entity Recognition, Machine Translation, Text Summarization, Natural Language Inference, Image Captioning, Visual Question Answering, Image Classification, Object Detection, Image Segmentation, …

Many Specific Models

Model 1

Model 2

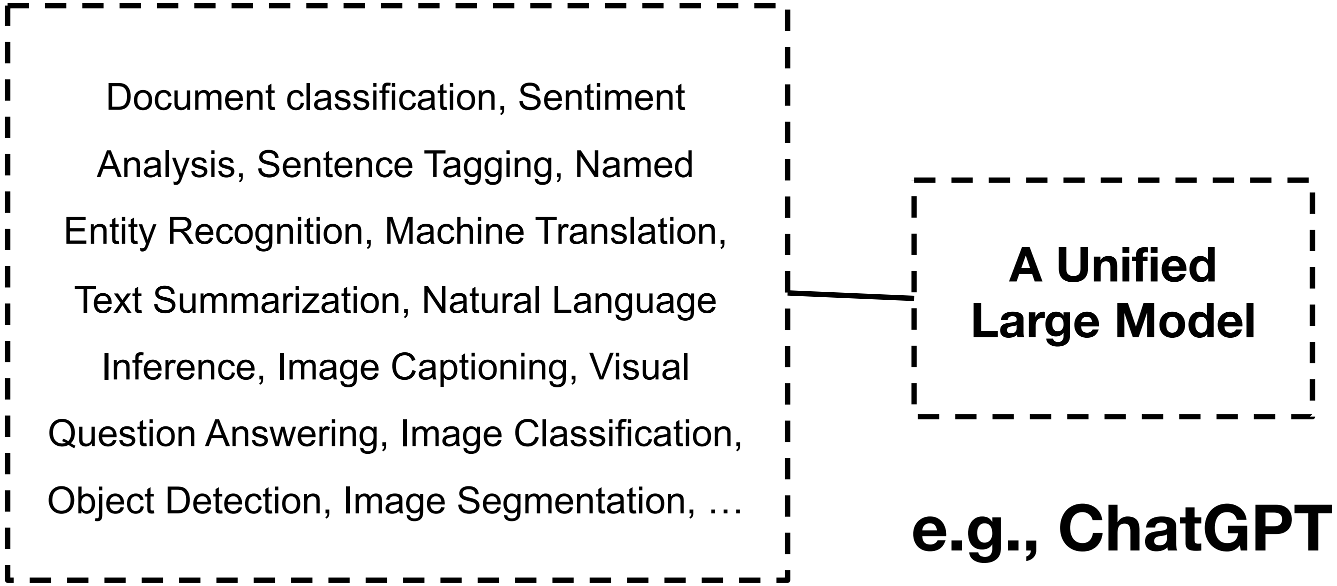
Model 3

…

Large communities in CV / NLP / ML

One-by-one to All-in-one

Many different tasks in CV / NLP / ML

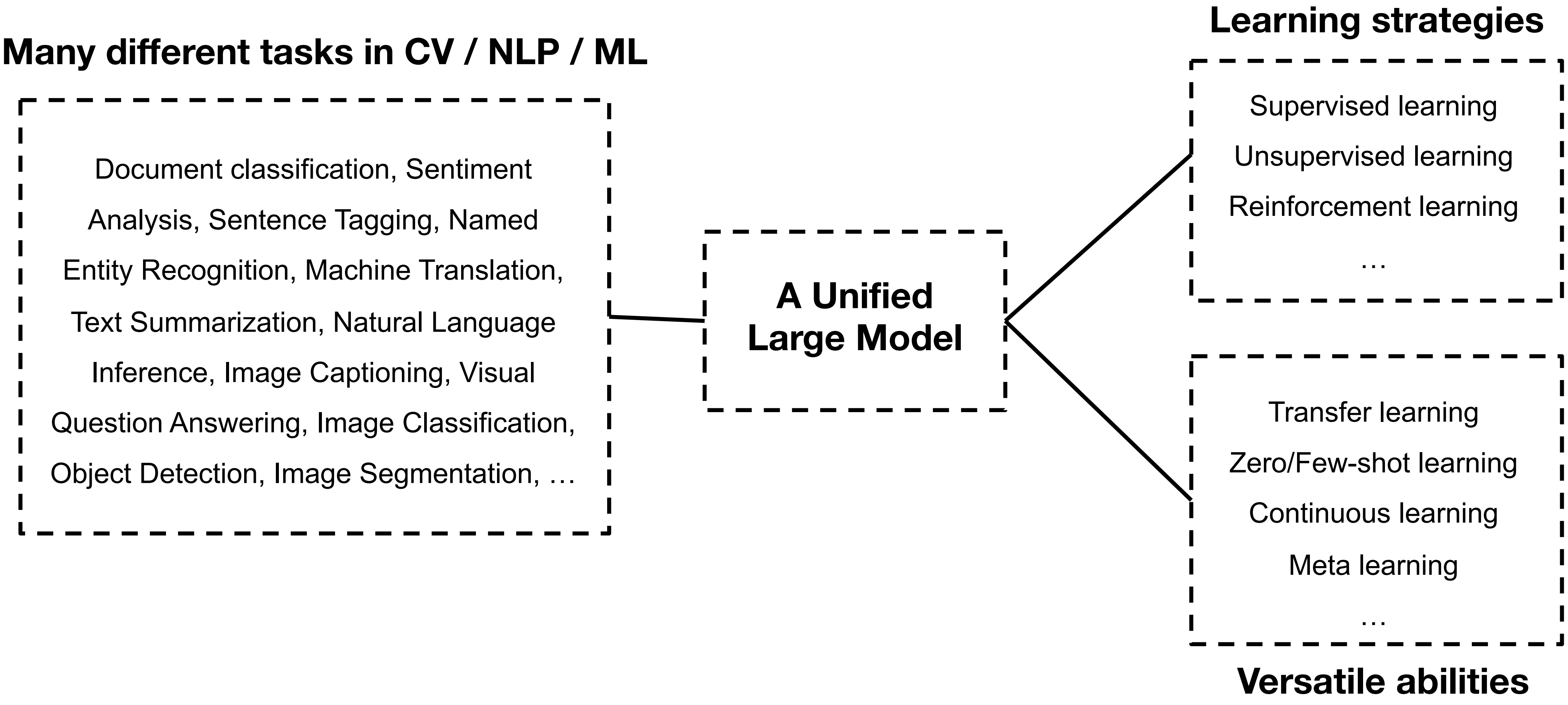


Document classification, Sentiment Analysis, Sentence Tagging, Named Entity Recognition, Machine Translation, Text Summarization, Natural Language Inference, Image Captioning, Visual Question Answering, Image Classification, Object Detection, Image Segmentation, …

A Unified Large Model

e.g., ChatGPT

One-by-one to All-in-one



Many different tasks in CV / NLP / ML

Document classification, Sentiment Analysis, Sentence Tagging, Named Entity Recognition, Machine Translation, Text Summarization, Natural Language Inference, Image Captioning, Visual Question Answering, Image Classification, Object Detection, Image Segmentation, …

A Unified Large Model

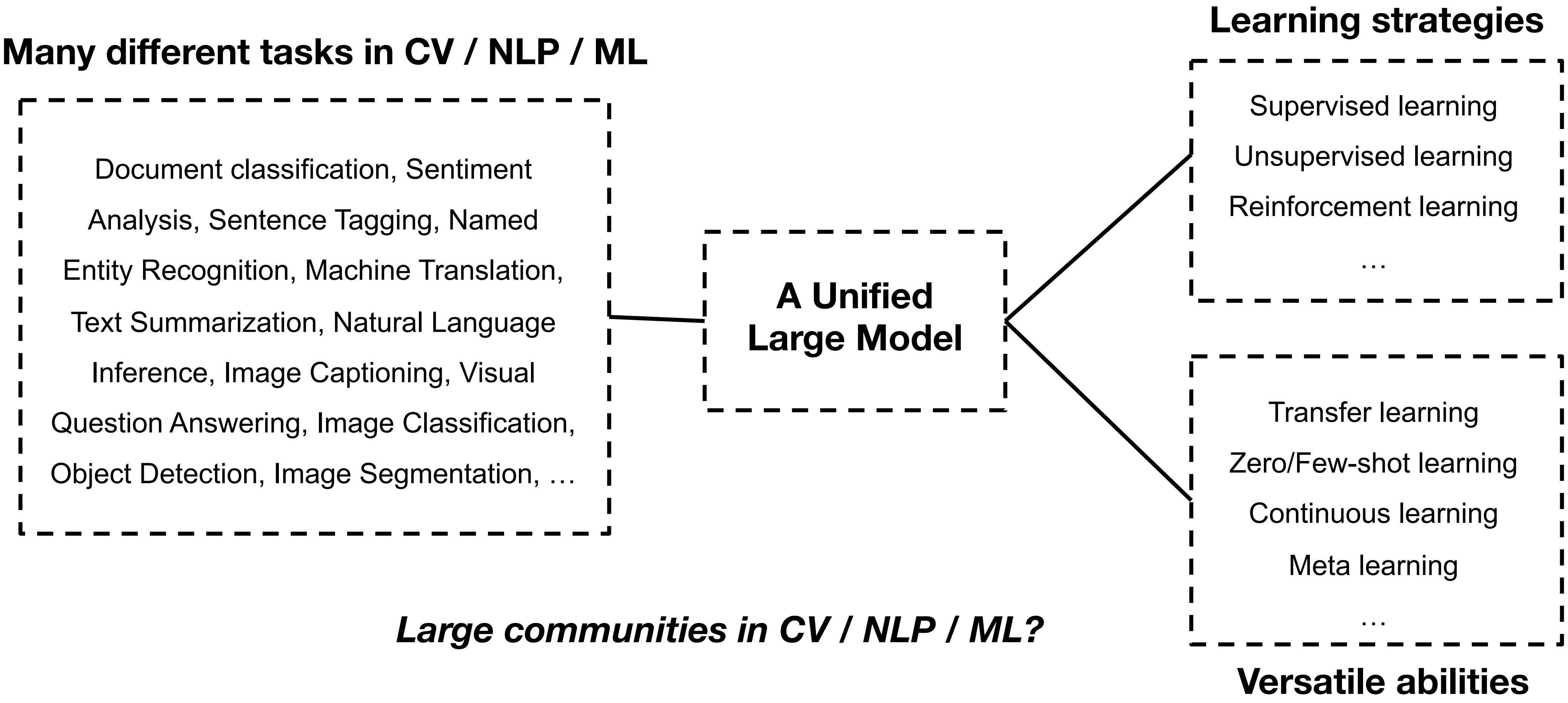
Learning strategies

Supervised learning
Unsupervised learning
Reinforcement learning
…

Transfer learning
Zero/Few-shot learning
Continuous learning
Meta learning
…

Versatile abilities

One-by-one to All-in-one



Many different tasks in CV / NLP / ML

Document classification, Sentiment Analysis, Sentence Tagging, Named Entity Recognition, Machine Translation, Text Summarization, Natural Language Inference, Image Captioning, Visual Question Answering, Image Classification, Object Detection, Image Segmentation, …

A Unified Large Model

Learning strategies

Supervised learning
Unsupervised learning
Reinforcement learning
…

Transfer learning
Zero/Few-shot learning
Continuous learning
Meta learning
…

Versatile abilities

Large communities in CV / NLP / ML?

Model-centric to Computation-centric

Techniques behind large models are not new

Attention > Transformer > Self-Supervised Learning > BERT > GPT 1 > GPT 2 > GPT 3 …

Model-centric to Computation-centric

Techniques behind large models are not new

Attention > Transformer > Self-Supervised Learning > BERT > GPT 1 > GPT 2 > GPT 3 …

Performant large models =
Existing models + More data + More Computation + More engineer

Model-centric to Computation-centric

Techniques behind large models are not new

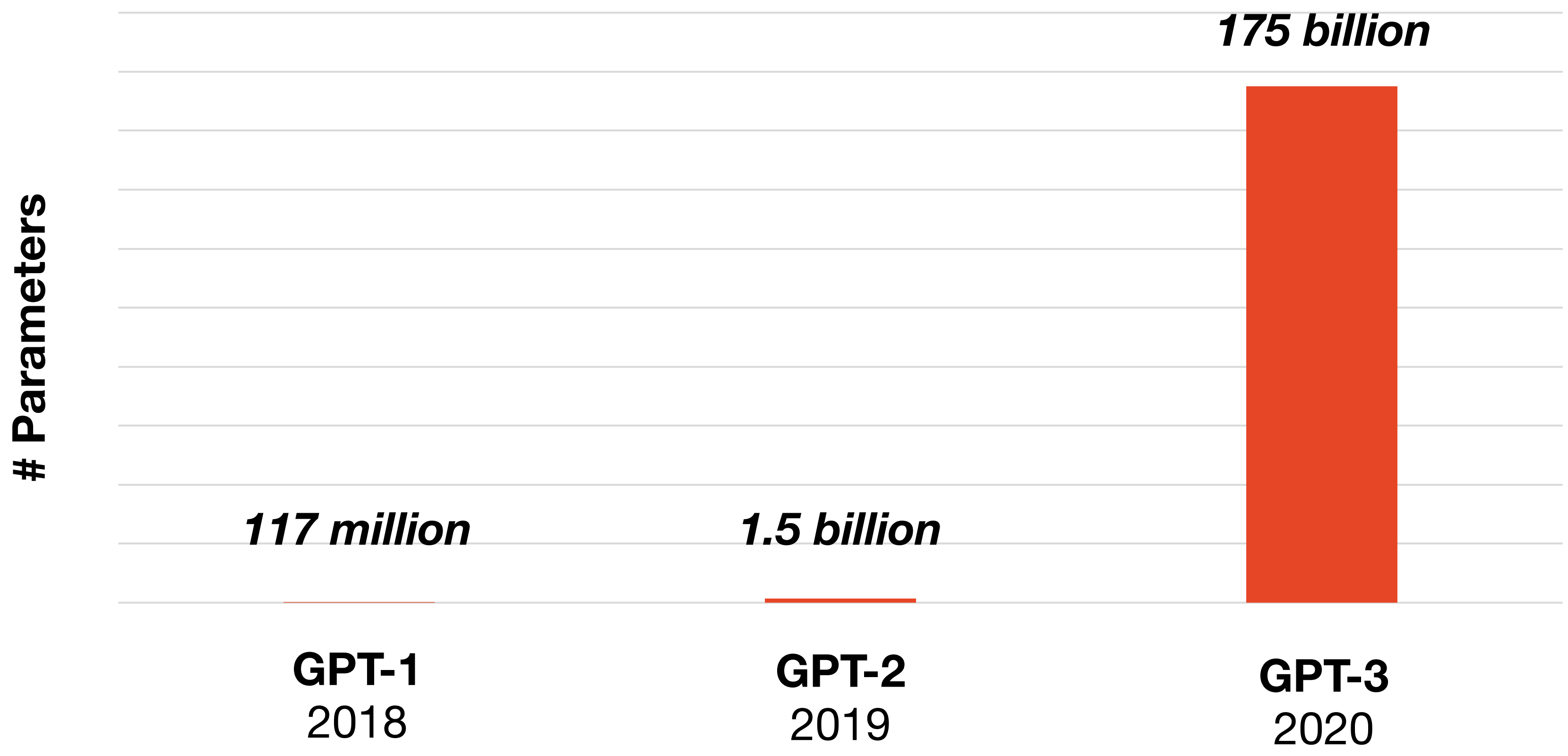
Attention > Transformer > Self-Supervised Learning > BERT > GPT 1 > GPT 2 > GPT 3 …

Performant large models =
Existing models + More data + More Computation + More engineer

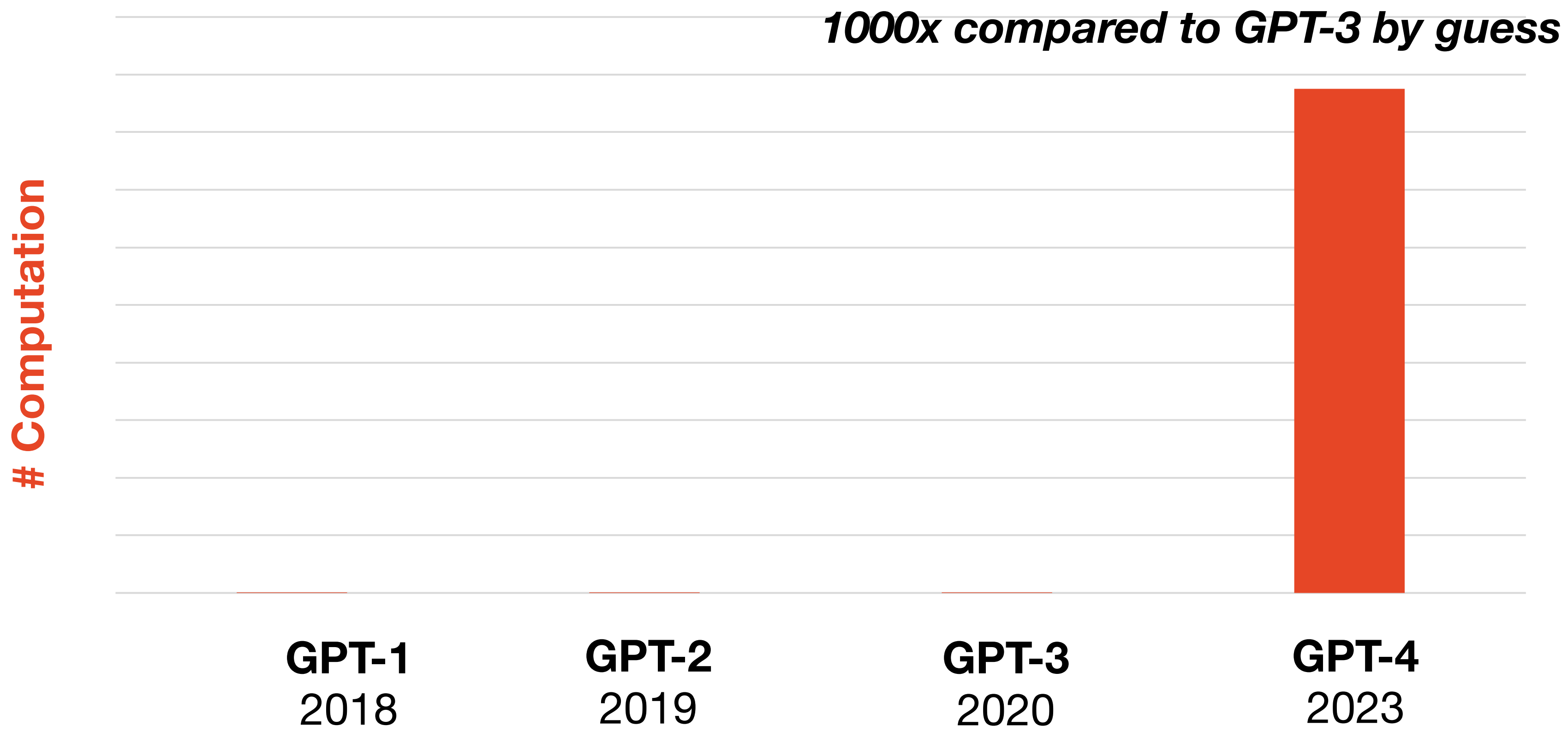
Less Important

More Important

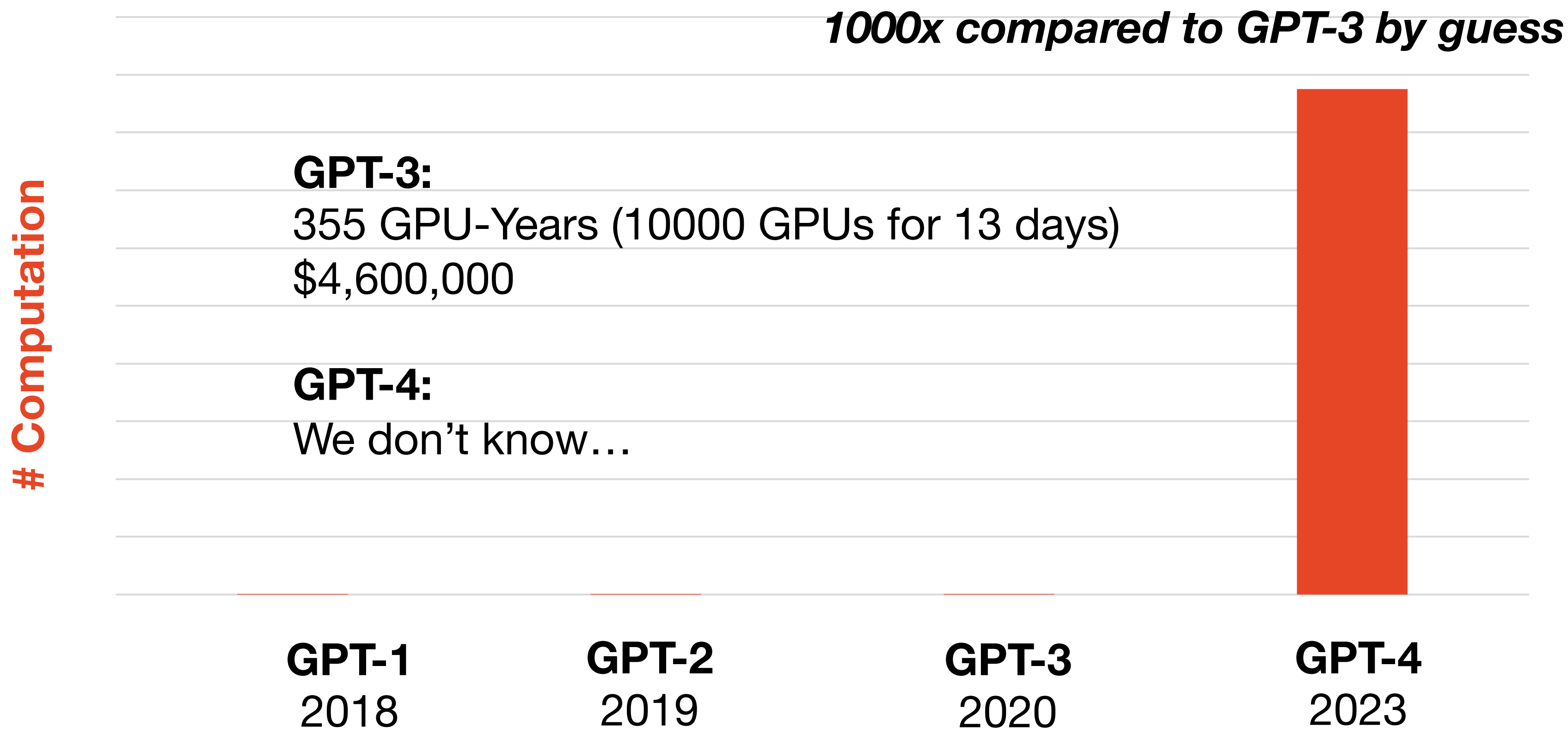
Model-centric to Computation-centric



Model-centric to Computation-centric



Model-centric to Computation-centric


1000x compared to GPT-3 by guess

GPT-3:
355 GPU-Years (10000 GPUs for 13 days)
$4,600,000

GPT-4:
We don’t know…

Decentralized to Centralized

Performant large models =
Existing models + More data + More Computation + More engineer

Very important

Decentralized to Centralized

Performant large models =
Existing models + More data + More Computation + More engineer

Private in-house data

Very expensive

Many tricks not disclosed

It is only affordable for big companies

Winner takes all

What Should We Do?

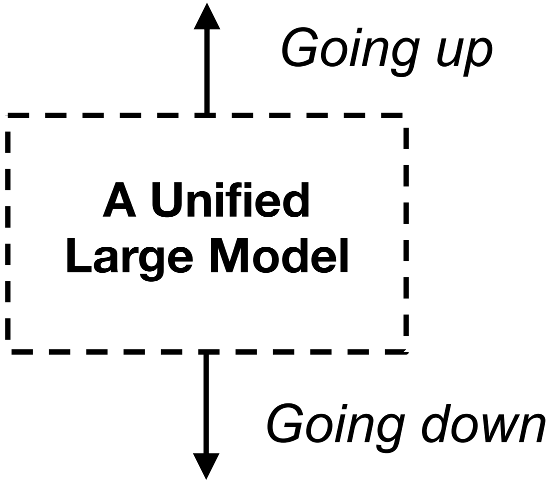
A Unified
Large Model

e.g., ChatGPT

What Should We Do?

AI + X

AI + Science, AI + Medical, AI + Social Computing, Embodied AI, …



Going up

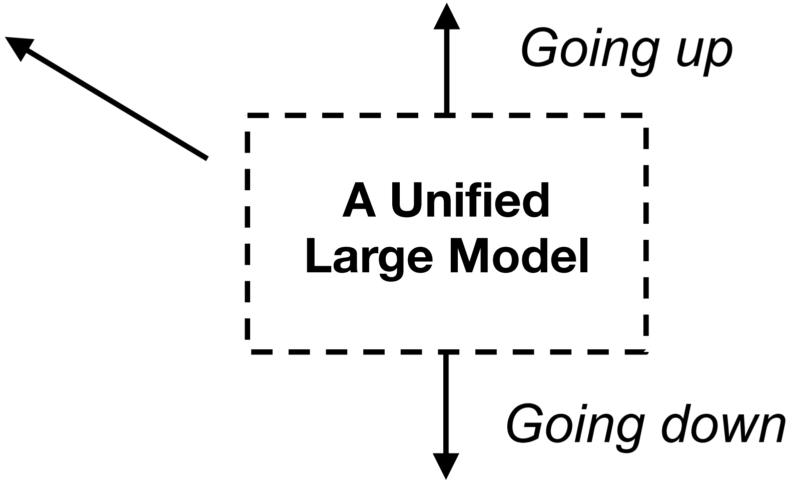
A Unified Large Model

Going down

Machine Learning

Different frameworks, Fundamental Problems, …

What Should We Do?



New Problems

Prompt, RLHF,

Understand its behaviours

Large models without

human in loop?

AI + X

AI + Science, AI + Medical, AI + Social Computing, Embodied AI, …

Going up

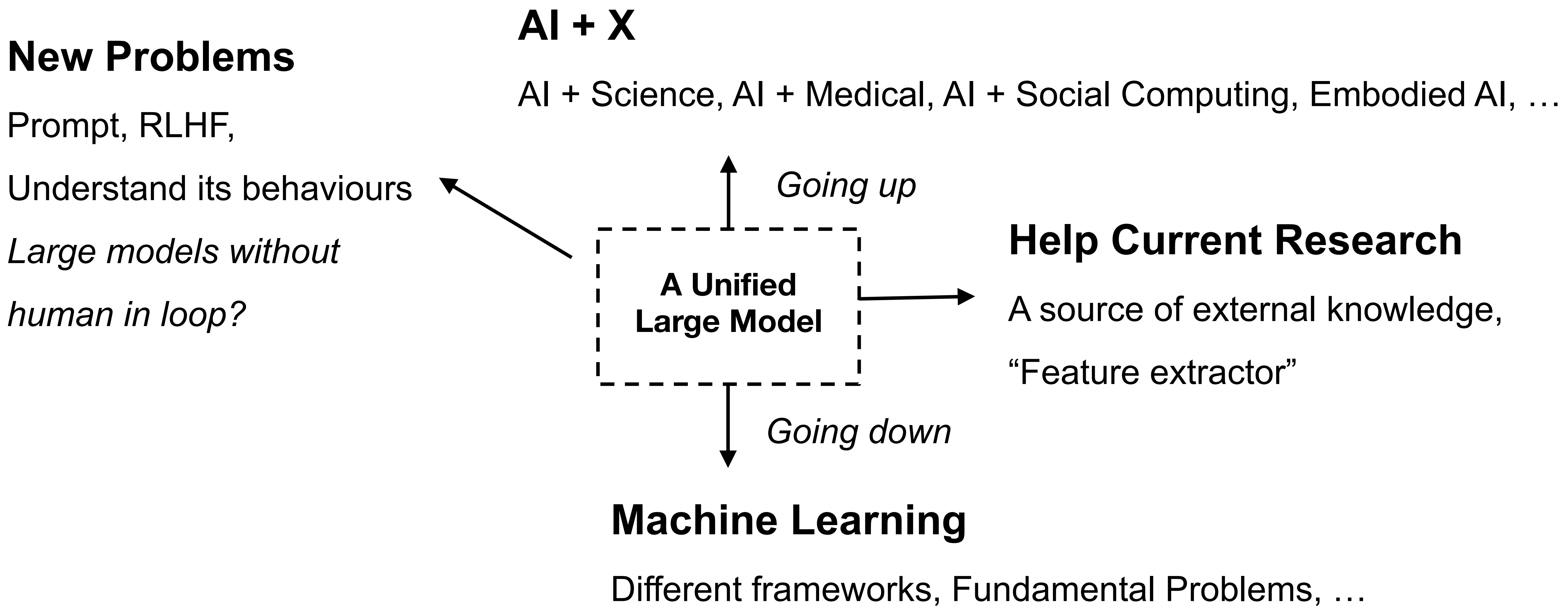
A Unified Large Model

Going down

Machine Learning

Different frameworks, Fundamental Problems, …

What Should We Do?



New Problems

Prompt, RLHF,

Understand its behaviours

Large models without

human in loop?

AI + X

AI + Science, AI + Medical, AI + Social Computing, Embodied AI, …

Going up

A Unified Large Model

Help Current Research

A source of external knowledge,

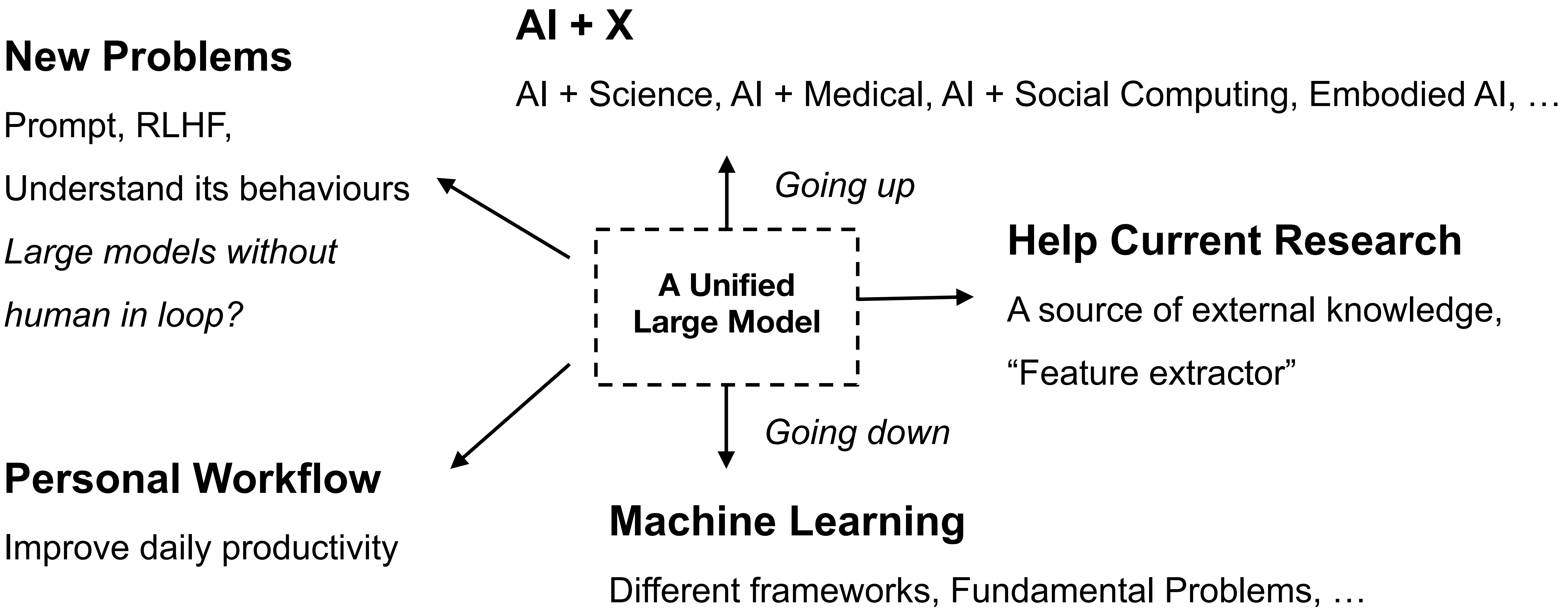
“Feature extractor”

Going down

Machine Learning

Different frameworks, Fundamental Problems, …

What Should We Do?



AI + X

AI + Science, AI + Medical, AI + Social Computing, Embodied AI, …

New Problems

Prompt, RLHF,

Understand its behaviours

Large models without human in loop?

Going up

A Unified Large Model

Help Current Research

A source of external knowledge,

“Feature extractor”

Going down

Personal Workflow

Improve daily productivity

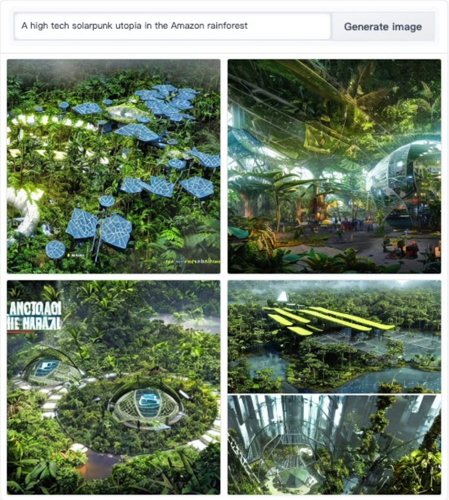
Machine Learning

Different frameworks, Fundamental Problems, …

Our Lab: Going Down

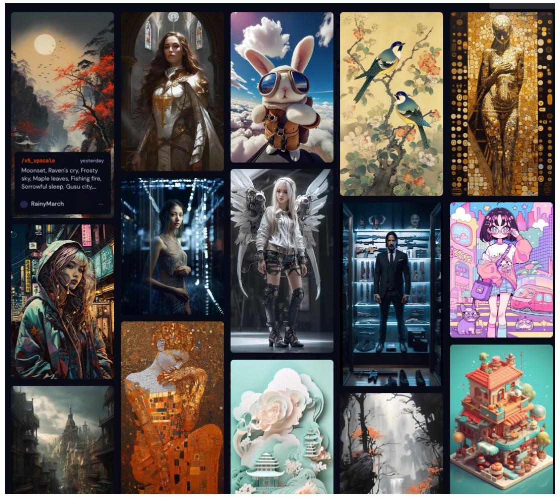
Diffusion Models: The technique behind Stable Diffusion and Midjourney

Stable Diffusion



A high tech solarpunk utopia in the Amazon rainforest

Midjourney



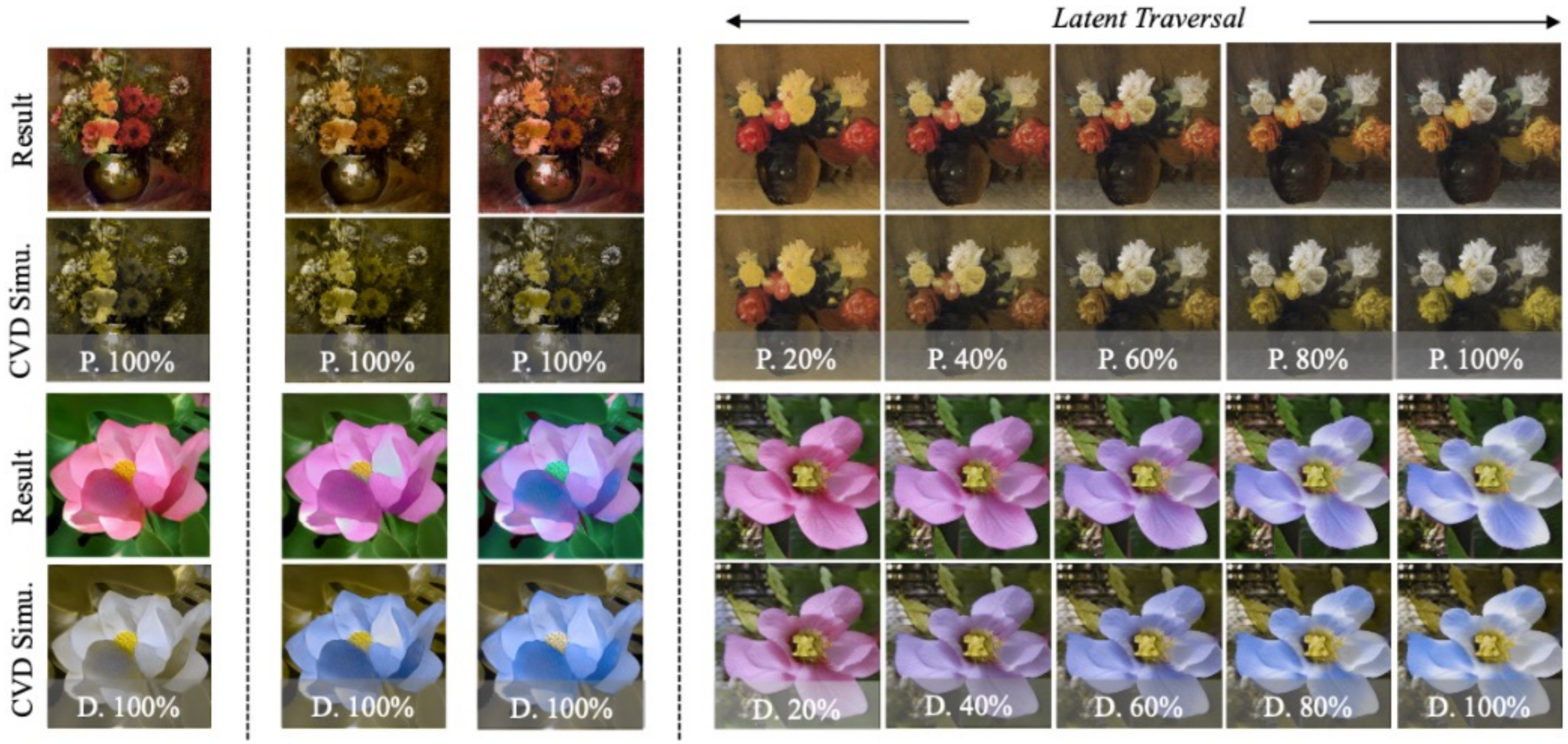
We are making diffusion models:

- Faster
- Safer
- More controllable
- More flexible
- More balanced
- …

Credit to: Dr. Chang Xu, Anh-Dung Dinh, Xiyu Wang, Junyu Zhang, Chen Chen

Our Lab: Going Up

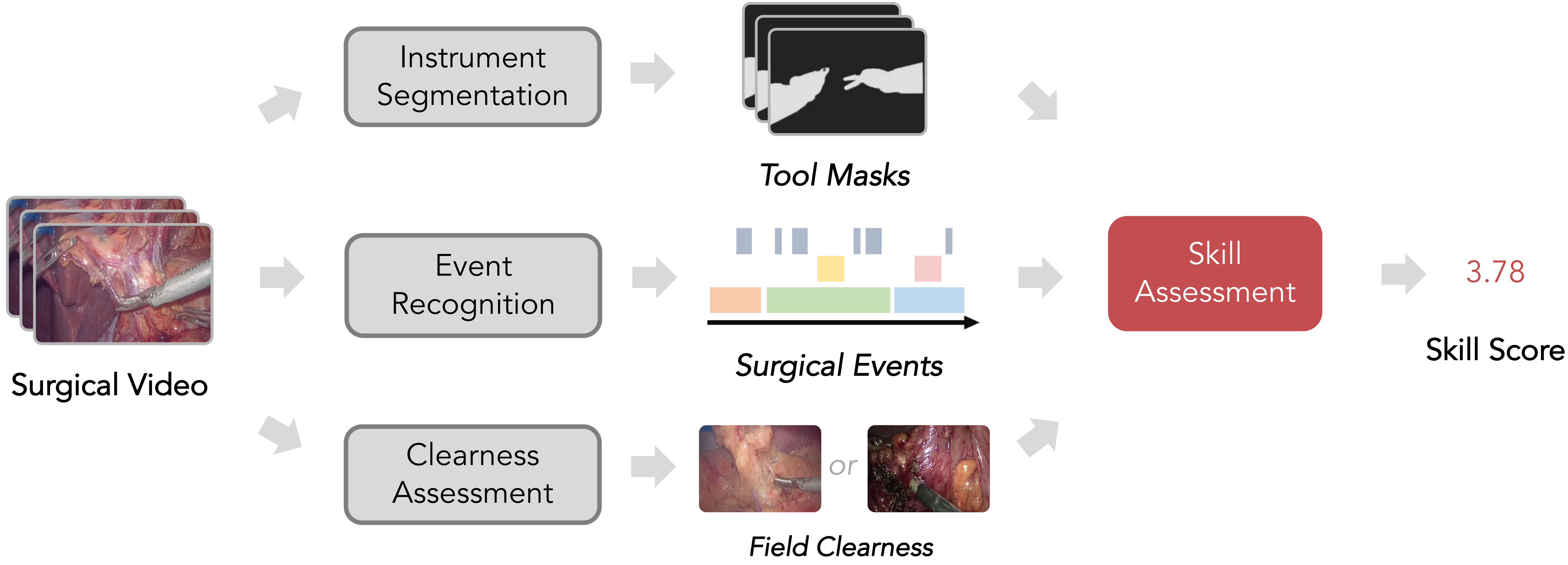
Generating images for people with color vision deficiency



Credit to: Dr. Chang Xu, Shuyi Jiang

Our Lab: Going Up

Surgical skill assessment and feedback using computer vision



Thank You!

Questions?

daochang.liu@sydney.edu.au