

10 YEARS OF UNIVERSITY RECOGNITION
20 YEARS OF ACADEMIC EXCELLENCE

REVA UNIVERSITY
Bengaluru, India

Webinar on
Self-Supervised Learning for
Large-scale AI adoption
use cases with ChatGPT

19th February, 2023 | 12:00 PM

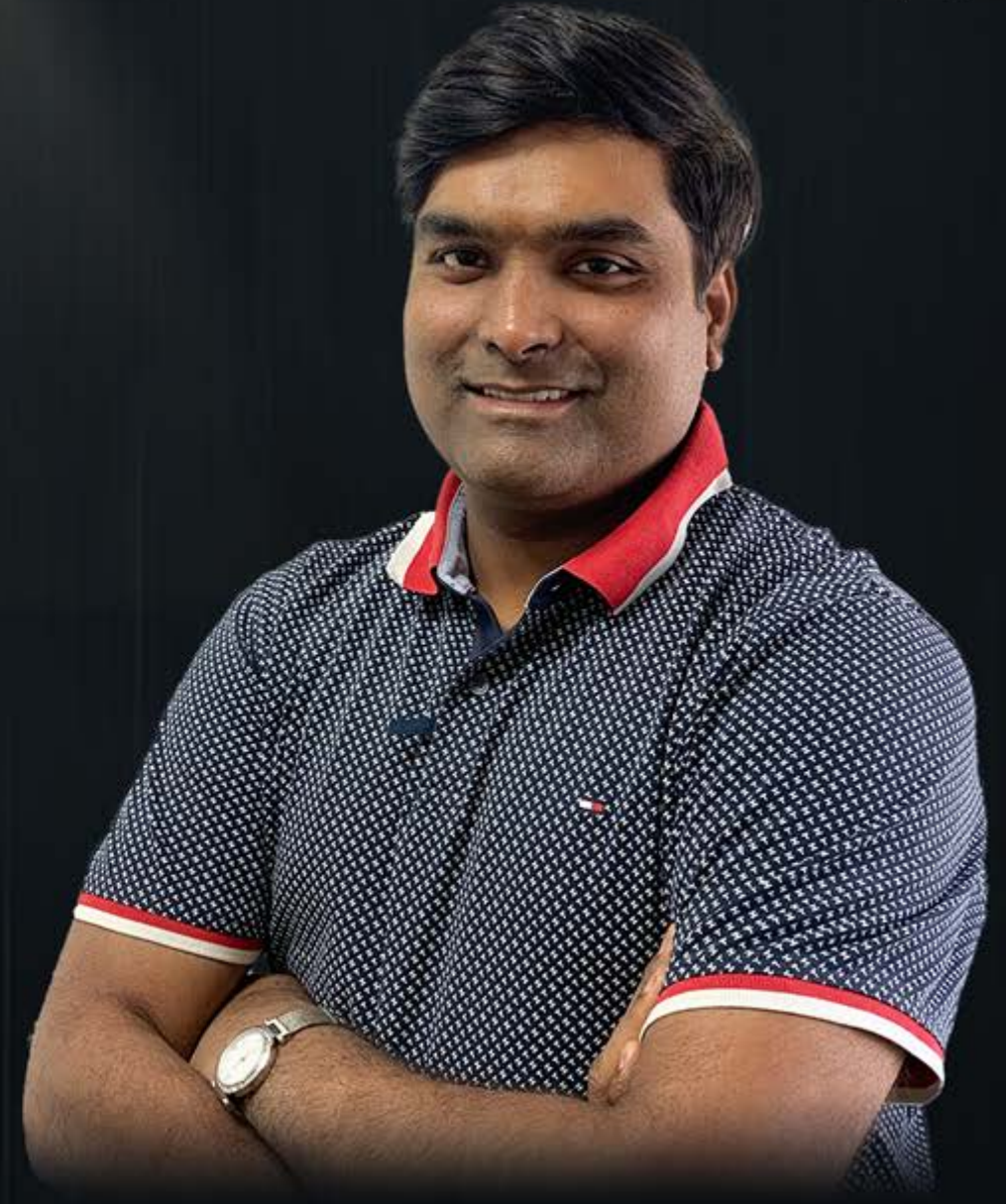
REVA Academy for Corporate Excellence - RACE
race.reva.edu.in

Who Am I

Dr. Sai Hareesh

Innovator, [Maestro Technologies Inc.](#) |
Chief Executive Officer, [Free](#) | AI Advisor, [LokiBots](#) &
[eMedicalPractice.com](#) |
Mentor – AI, [RACE](#), [REVA University](#)

Putting
Self Supervised Learning
 to Action



10 YEARS OF UNIVERSITY RECOGNITION
20 YEARS OF ACADEMIC EXCELLENCE



REVA
UNIVERSITY
Bengaluru, India

Knowledge Partners





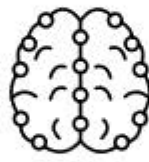
A platform for Human Bot collaboration



Do it yourself automation



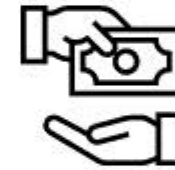
Cloud native platform



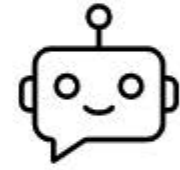
Powered by Deep-learning



Enterprise grade security



Pay-as-you-go pricing



Unlimited Bot & Users

Let humans do what they do best. For everything else, there is a bot.



What do I do?

Provide ounce of

Insights

from ton of

information

Outline – Agenda

Large Scale Models

Found on Internet

AI Adoption

My Understanding and Perspective, not on Internet



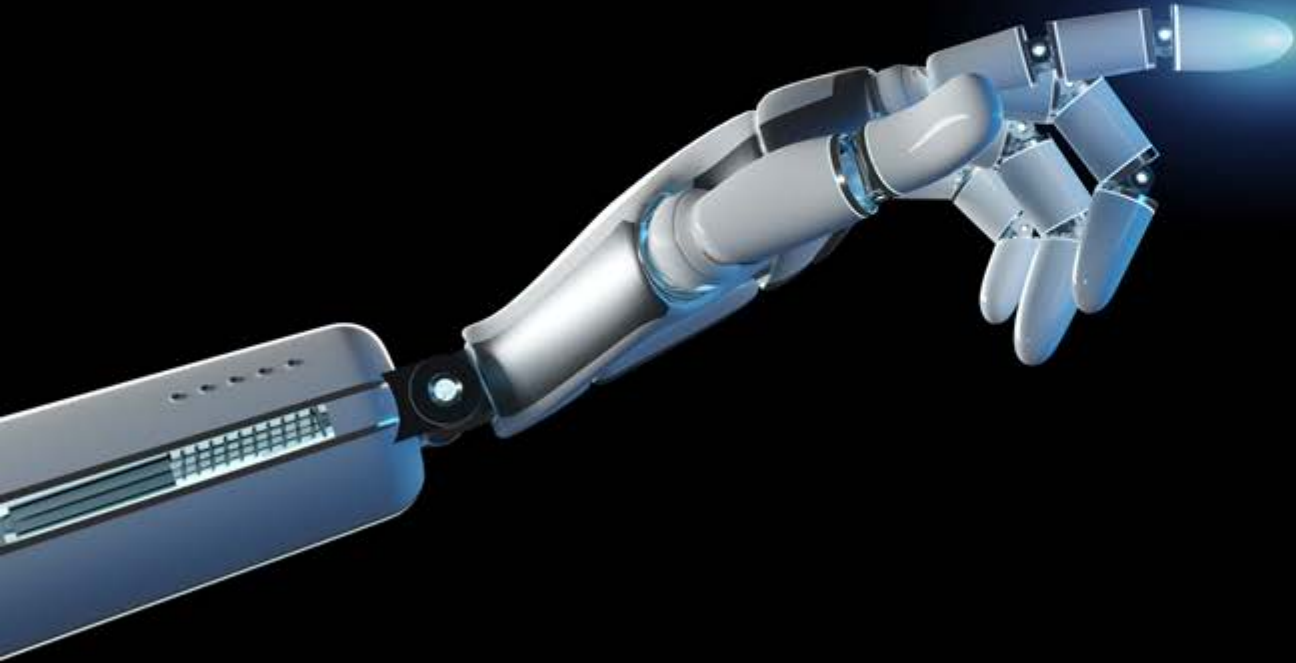
Large Scale Models Highlights

Size of the Model

Billions of Parameters

In terms of space and compute

Super Computers Needed



Basic Elements of **GPT**

How much Information is the **Machine Given during Learning?**



"Pure" Reinforcement Learning (cherry)

The machine predicts a scalar reward given once in a while.

- A few bits for some samples

Supervised Learning (icing)

The machine predicts a category or a few numbers for each input

- Predicting human-supplied data
- 10→10,000 bits per sample

Self-Supervised Learning (cake génoise)

The machine predicts any part of its input for any observed part.

- Predicts future frames in videos
- Millions of bits per sample

Self-Supervised Learning

In general, withhold some part of the data, and task the network with predicting it

The task defines a proxy loss, and the network is forced to learn what we really care about.
E.g. a semantic representation, in order to solve it

Predict everything from everything else



Pretext **Tasks**

The pretext tasks and pseudo labels share two common properties:

Pseudo labels for the pretext task can be automatically generated based on the attributes of images or videos.
(Without any human supervision.)

Visual features of images or videos need to be captured by ConvNets to solve the pretext tasks.
(Solving the task is equivalent to learning the feature.)



Self-Supervised Learning of Video

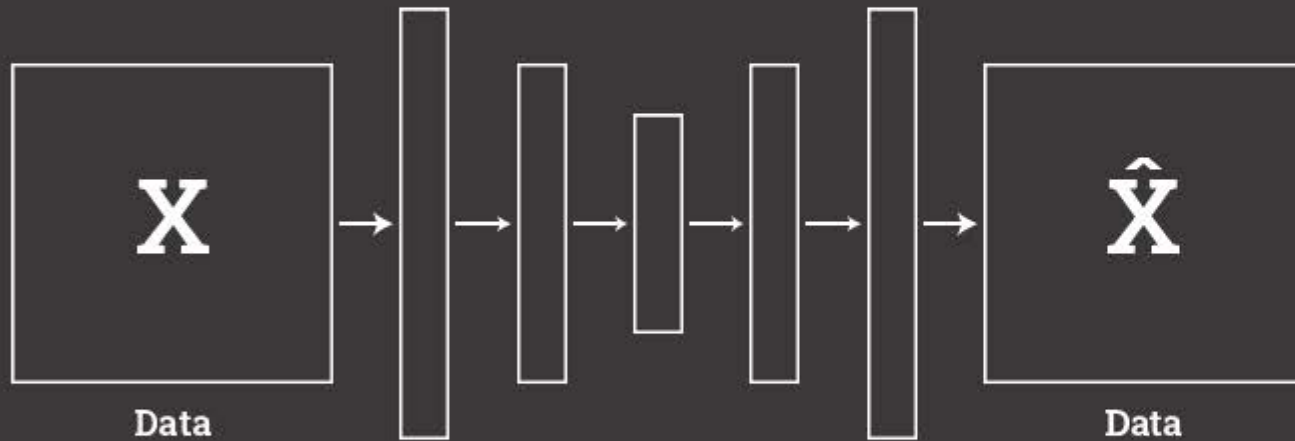
Video is a temporal sequence of frames



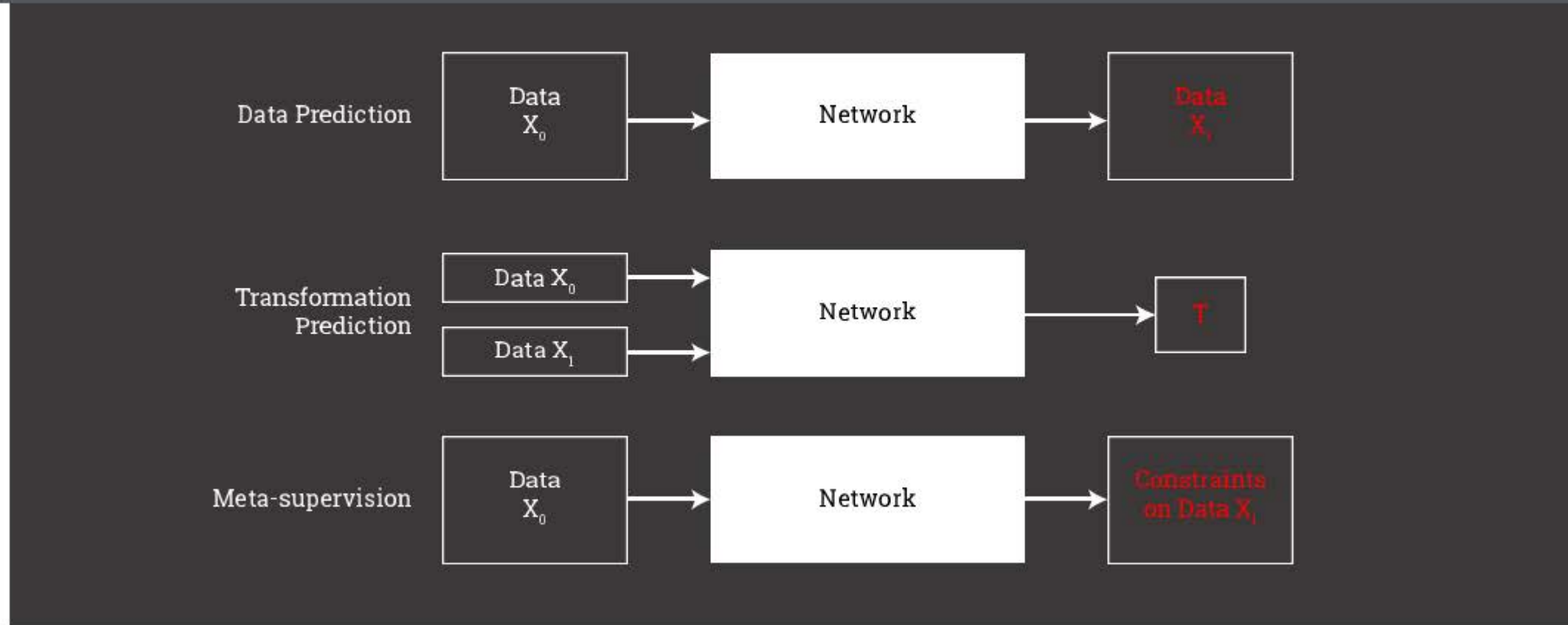
What can we define as a proxy loss?

- Nearby(in time) frames are strongly correlated, further away may not be
- Temporal order of the frames
- Motion of objects (via optical flow)
- Temporal coherence of objects their colors, their shapes, ...

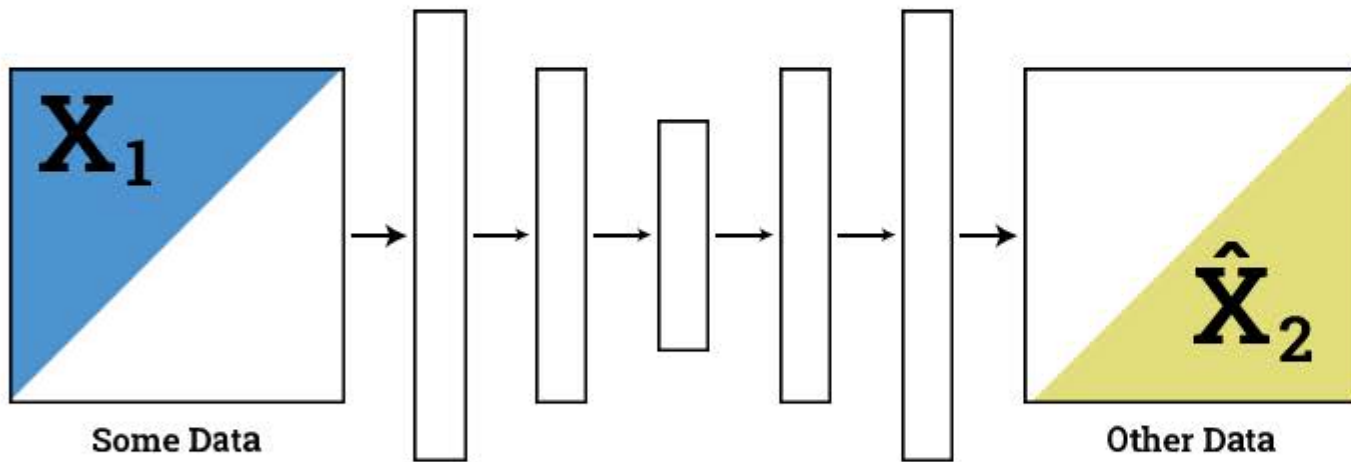
Autoencoder: **Self-Supervision as Data Compression** (Hinton 2009)



Partial Taxonomy of Self-Supervised Learning



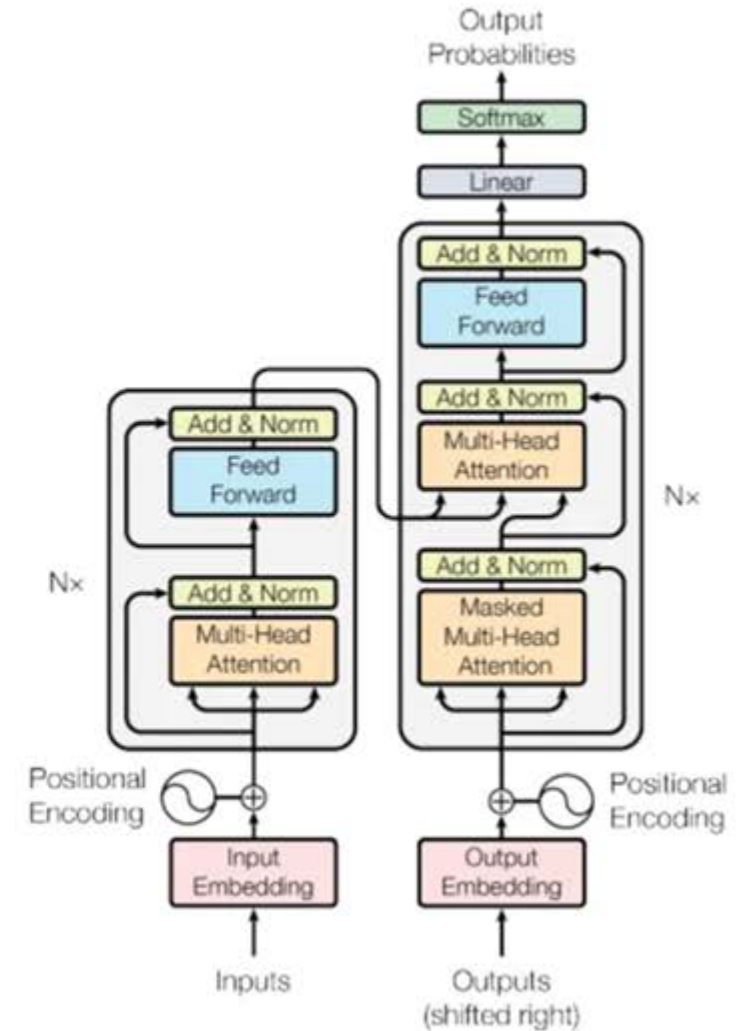
Self-Supervision as Data Prediction



Attention is **all you need**



Self-attention is a sequence-to-sequence operation: a sequence of vectors goes in, and a sequence of vectors comes out. Let's call the input vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t$, and the corresponding output vectors $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t$. The vectors all have dimension \mathbf{k} . To produce output vector \mathbf{y}_i , the self-attention operation simply takes a weighted average over all the input vectors, the simplest option is the dot product.



Attention dot product

query,

$$\mathbf{q}_i = \mathbf{W}_q \mathbf{x}_i$$

key

$$\mathbf{k}_i = \mathbf{W}_k \mathbf{x}_i$$

value

$$\mathbf{v}_i = \mathbf{W}_v \mathbf{x}_i$$

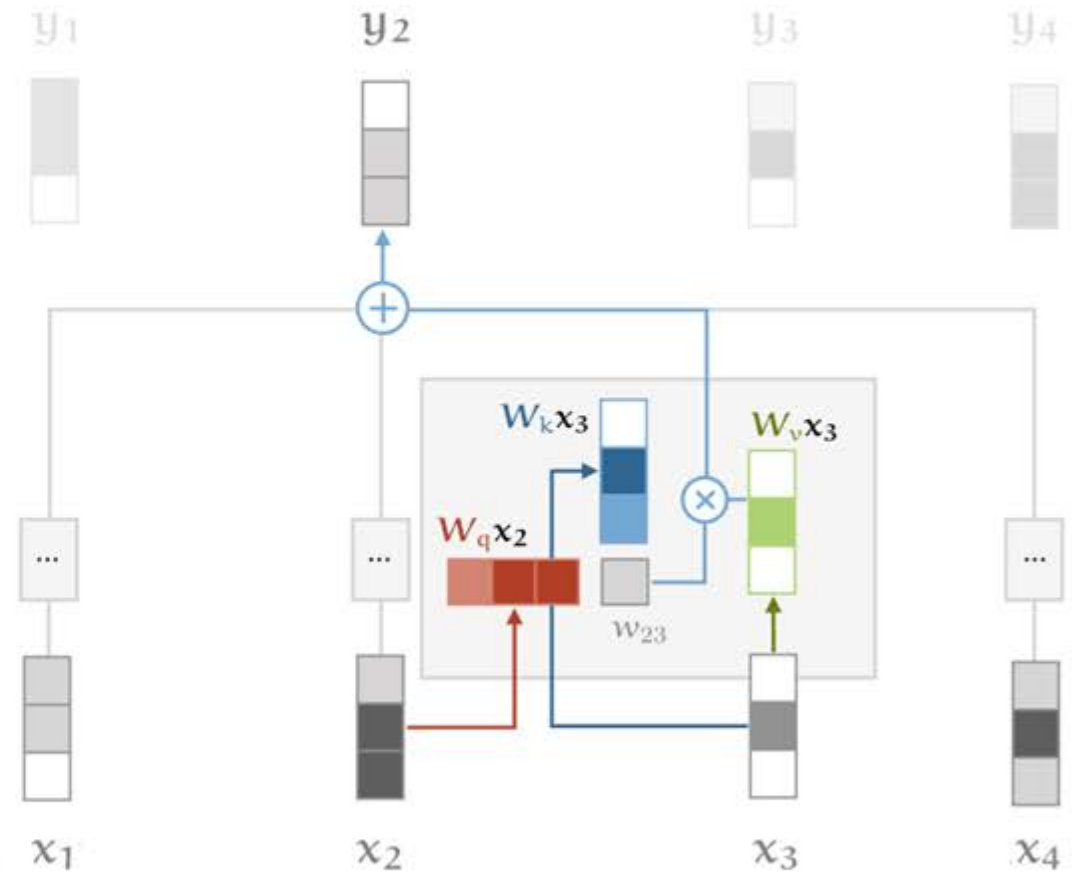
$$w'_{ij} = \mathbf{q}_i^T \mathbf{k}_j$$

$$w_{ij} = \text{softmax}(w'_{ij})$$

$$\mathbf{y}_i = \sum_j w_{ij} \mathbf{v}_j$$



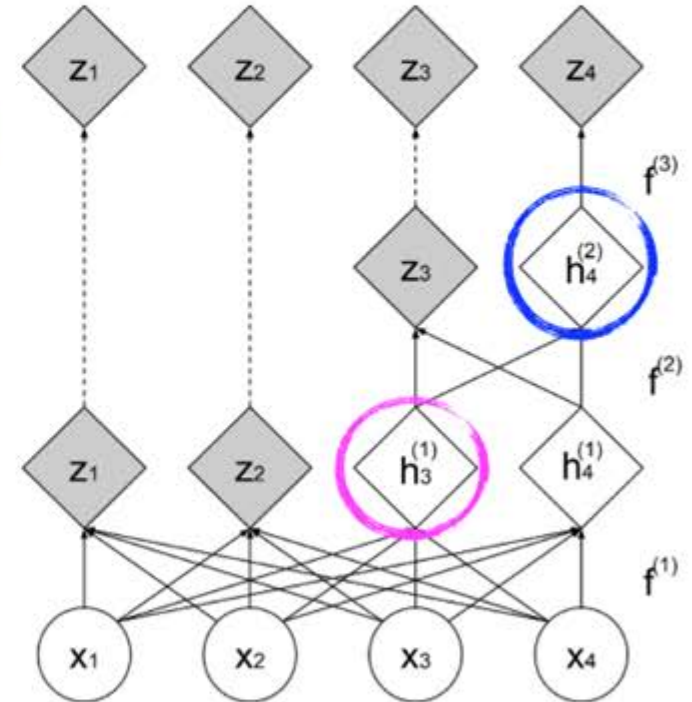
Basic **self attention**



ARM

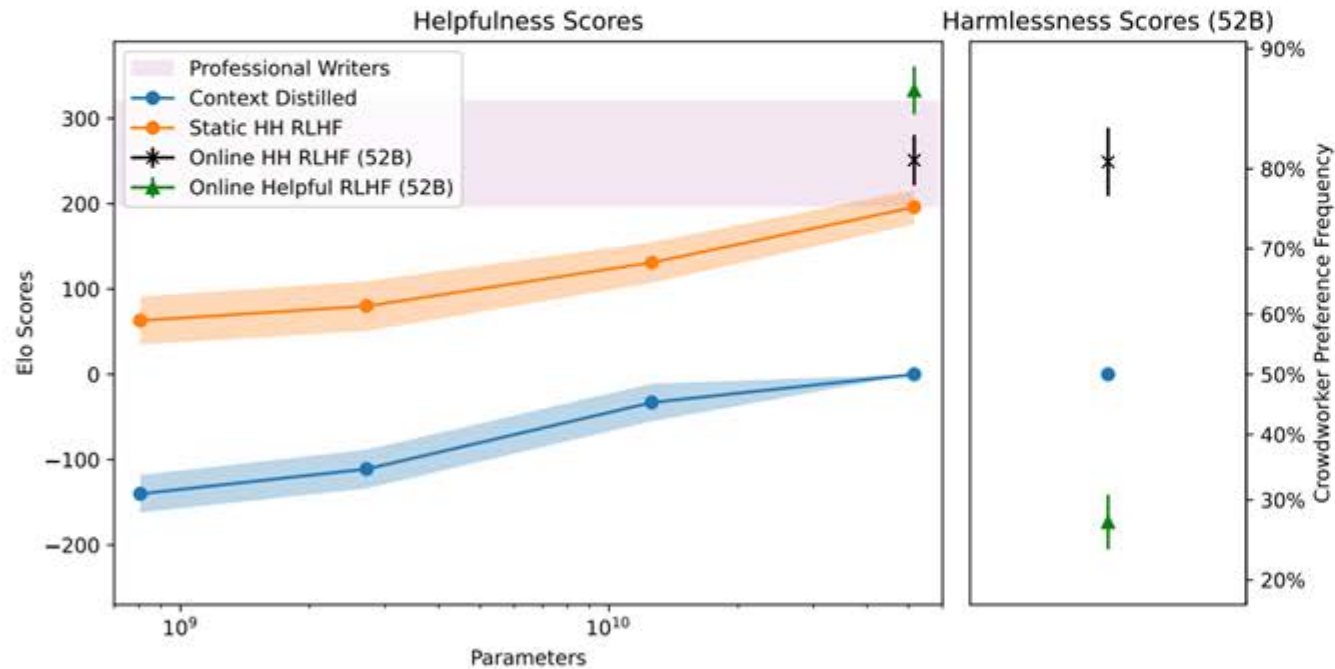
Autoregressive Models

$$p(z_1)p(z_2)p(z_3 | z_1, z_2)p(z_4 | z_1, z_2, z_3)$$

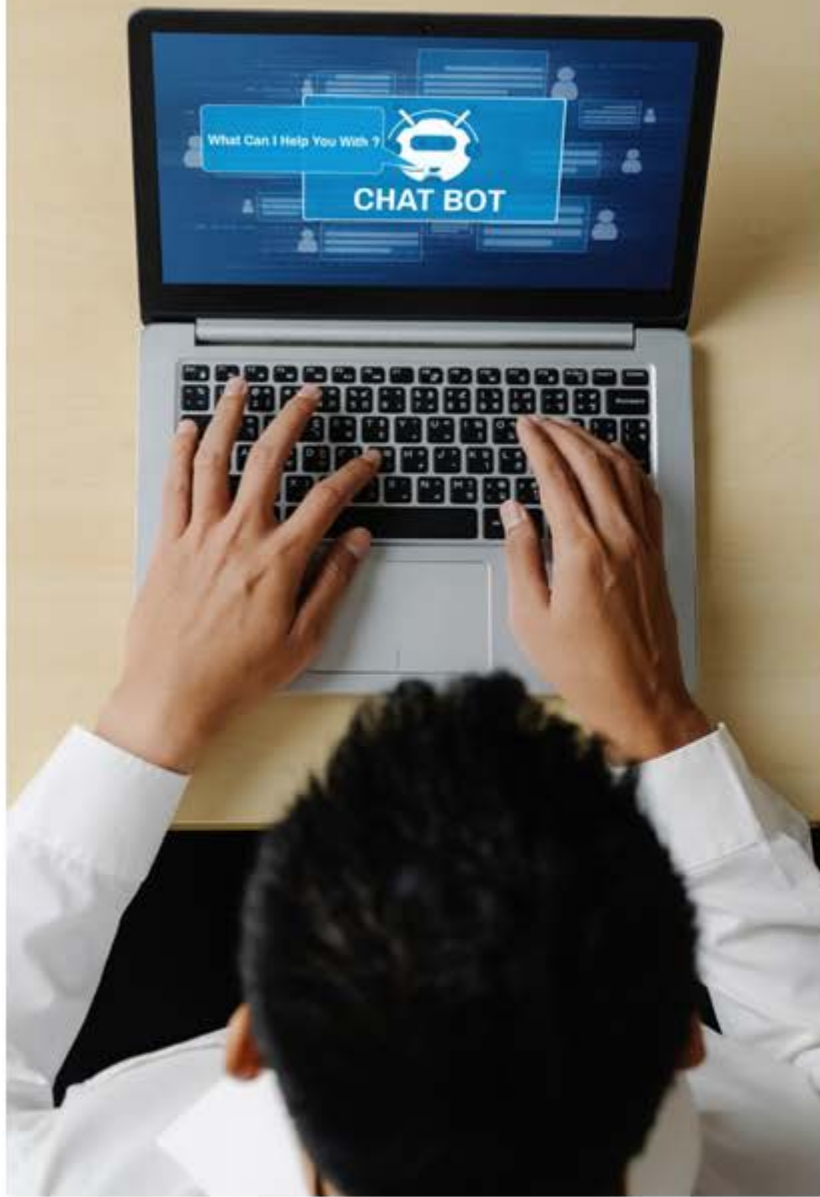
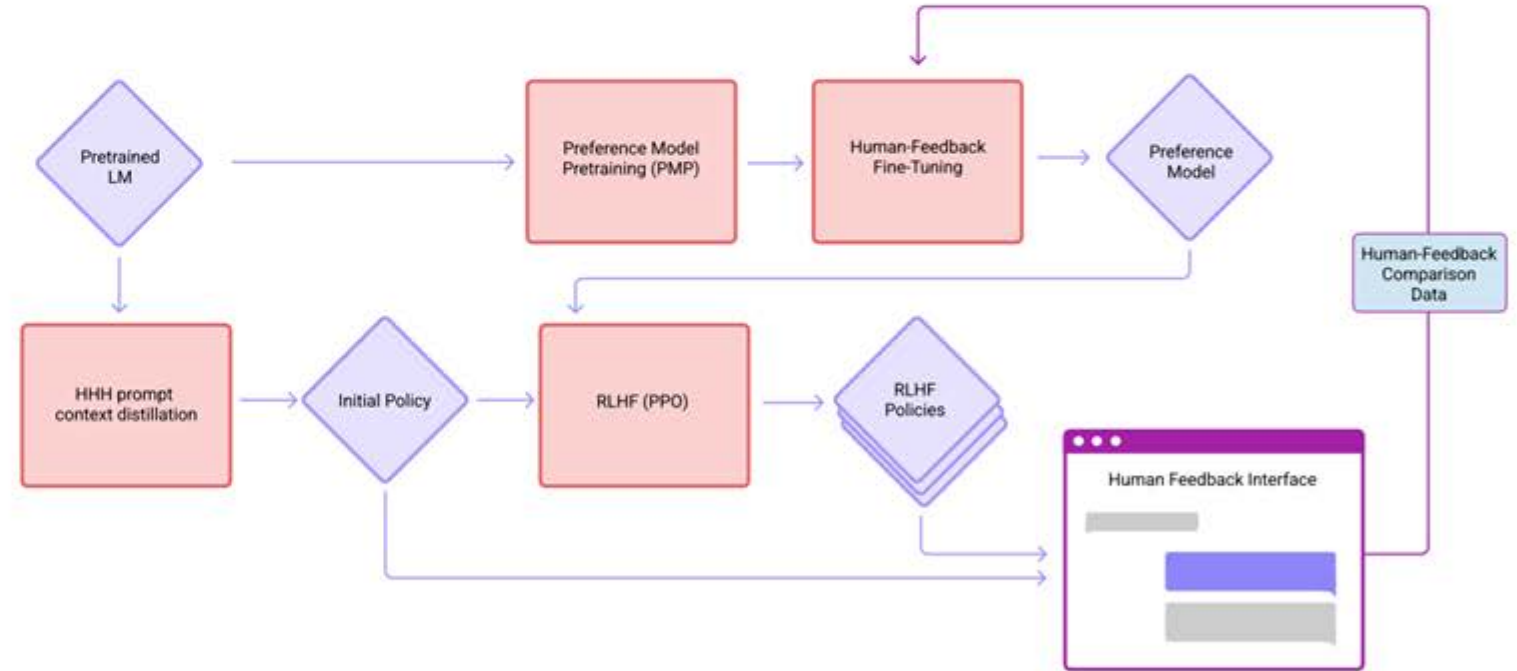


Reinforcement Learning from **Human Feedback (RLHF)**

Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback



Reinforcement Learning from Human Feedback (RLHF)



Transformers

```
mirror_mod.use_z = False  
elif _operation == "MIRROR_Y":  
    mirror_mod.use_x = False  
    mirror_mod.use_y = True  
    mirror_mod.use_z = False  
elif _operation == "MIRROR_Z":  
    mirror_mod.use_x = False  
    mirror_mod.use_y = False  
    mirror_mod.use_z = True
```

```
def reflection_of_the_mirror_obj:  
    mirror_obj = self.mirror_obj  
    modifier = self.mirror_obj
```



Google

T5

Partnership for Anthropic for Claude

- Anthropic Forges Partnership With Google Cloud to Help Deliver Reliable and Responsible AI

OpenAI used a smaller version of GPT-3 for its first popular RLHF model, InstructGPT. Anthropic used transformer models from 10 million to 52 billion parameters trained for this task. DeepMind used their 280 billion parameter model Gopher.

BARD

Courtesy: <https://www.googlecloudpresscorner.com/2023-02-03-Anthropic-Forges-Partnership-With-Google-Cloud-to-Help-Deliver-Reliable-and-Responsible-AI>

OpenAI

GPT series

DALL·E 2

new AI system that can create realistic images and art from a description in natural language.

Davinci

excels is in understanding the intent of text.

For applications requiring deep understanding of the content, like summarization for a specific audience and creative content generation

Use for: Complex intent, cause and effect, summarization for audience

Microsoft **OpenAI** partnership

Naming convention

Azure OpenAI's model names typically correspond to the following standard naming convention:

`{family}-{capability}[-{input-type}]-{identifier}`

Element	Description
<code>{family}</code>	The model family of the model. For example, GPT-3 models uses <code>text</code> , while Codex models use <code>code</code> .
<code>{capability}</code>	The relative capability of the model. For example, GPT-3 models include <code>ada</code> , <code>babbage</code> , <code>curie</code> , and <code>davinci</code> .
<code>{input-type}</code>	(Embeddings models only) The input type of the embedding supported by the model. For example, text search embedding models support <code>doc</code> and <code>query</code> .
<code>{identifier}</code>	The version identifier of the model.

For example, our most powerful GPT-3 model is called `text-davinci-003`, while our most powerful Codex model is called `code-davinci-002`.



Microsoft

Bing vs Google Search

I think **every app**
is going to be an **AI app.**

– Satya Nadella

With power fueled by GPT and other latest AI advancements mostly in Generative AI



GPT - 1

Improving Language Understanding by Generative Pre-training

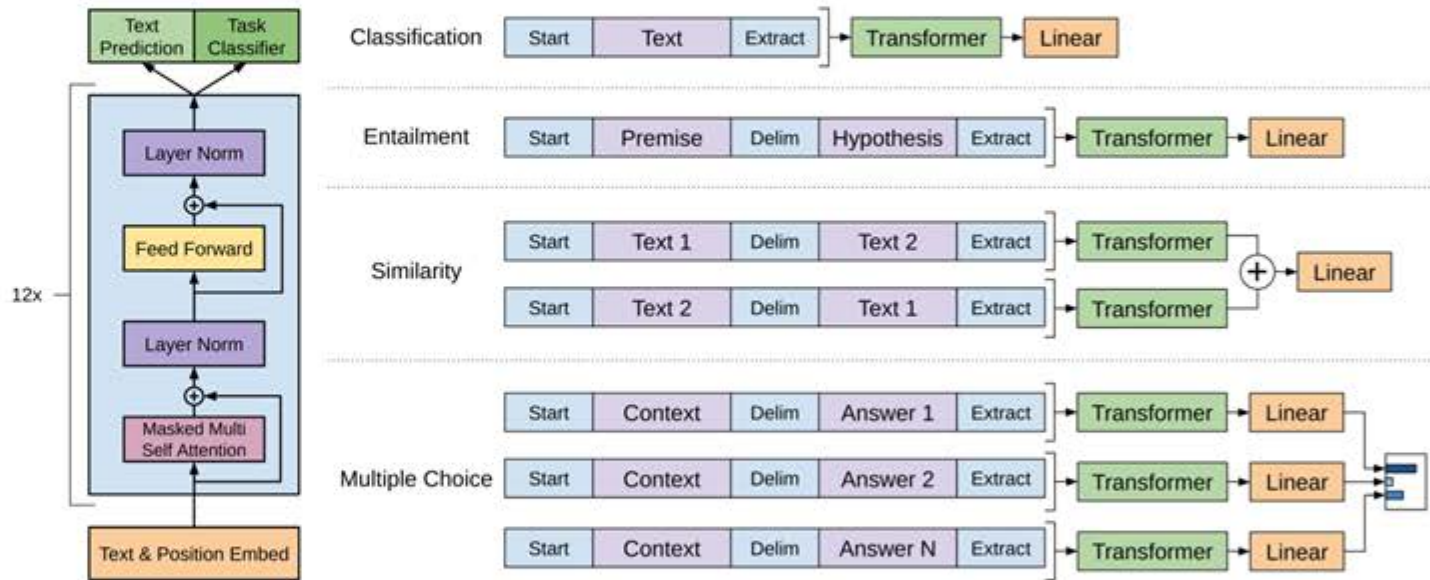
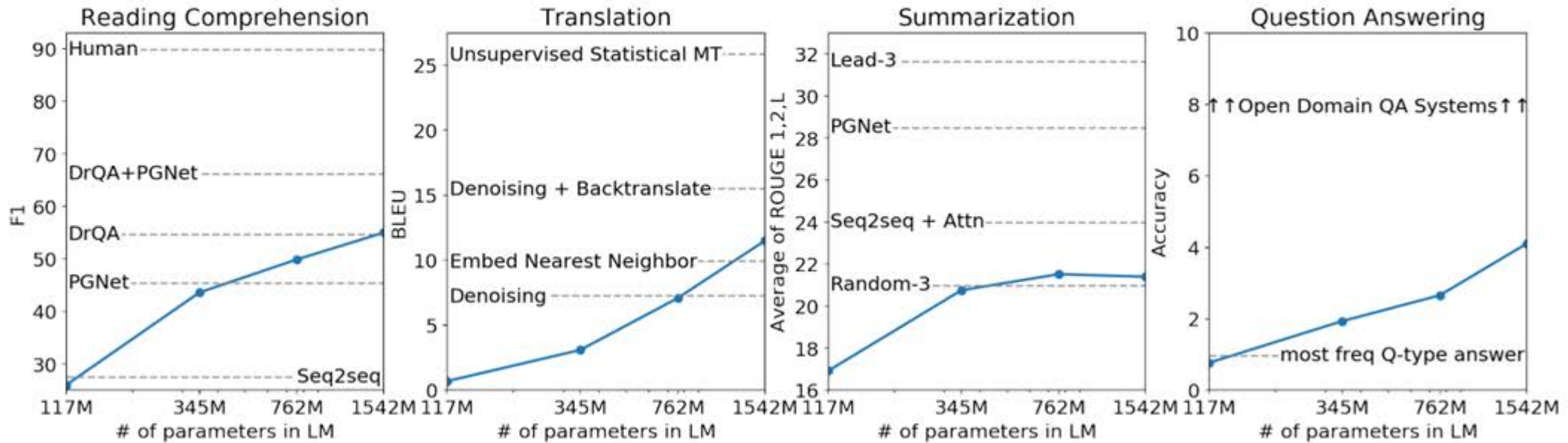


Figure 1: **(left)** Transformer architecture and training objectives used in this work. **(right)** Input transformations for fine-tuning on different tasks. We convert all structured inputs into token sequences to be processed by our pre-trained model, followed by a linear+softmax layer.

GPT - 2

Language_models_are_unsupervised_multitask_learners.



GPT - 3

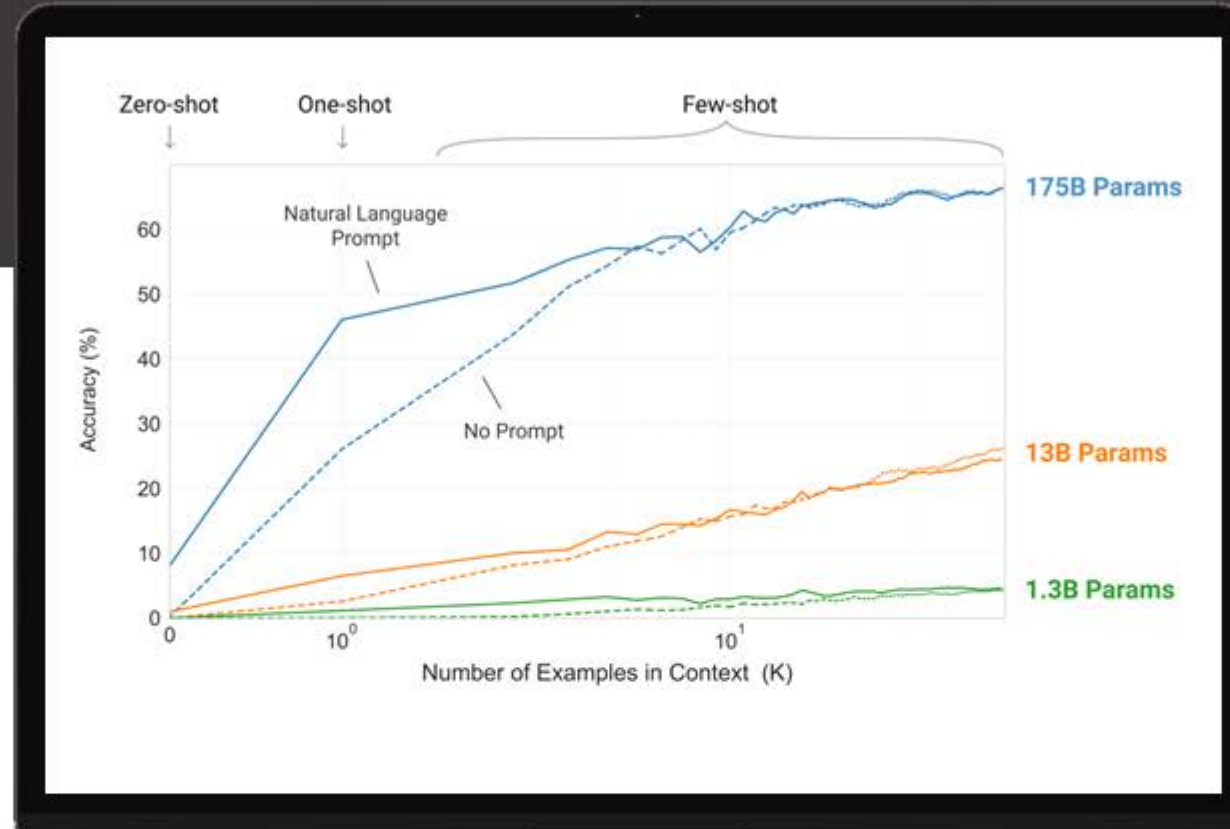


Figure 1.2: Larger models make increasingly efficient use of in-context information. We show in-context learning performance on a simple task requiring the model to remove random symbols from a word, both with and without a natural language task description (see Sec. 3.9.2). The steeper “in-context learning curves” for large models demonstrate improved ability to learn a task from contextual information. We see qualitatively similar behavior across a wide range of tasks.

GPT-J

Mesh Transformer JAX

JAX is Autograd and XLA, brought together for high-performance machine learning research.

Implementation of model & data-parallel

compute generously provided by the TPU Research Cloud with assistance from EleutherAI.

A 6 billion parameter, autoregressive text generation model trained on The Pile

Autograd can automatically differentiate native Python and Numpy code.

XLA – Accelerated Linear Algebra.



Meta

RoBERTa – optimization of parameters to BERT setting

Focused on Metaverse

**ChatGPT is 'not particularly innovative,'
and 'nothing revolutionary', says Meta's
chief AI scientist**



Amazon AlexaTM20B

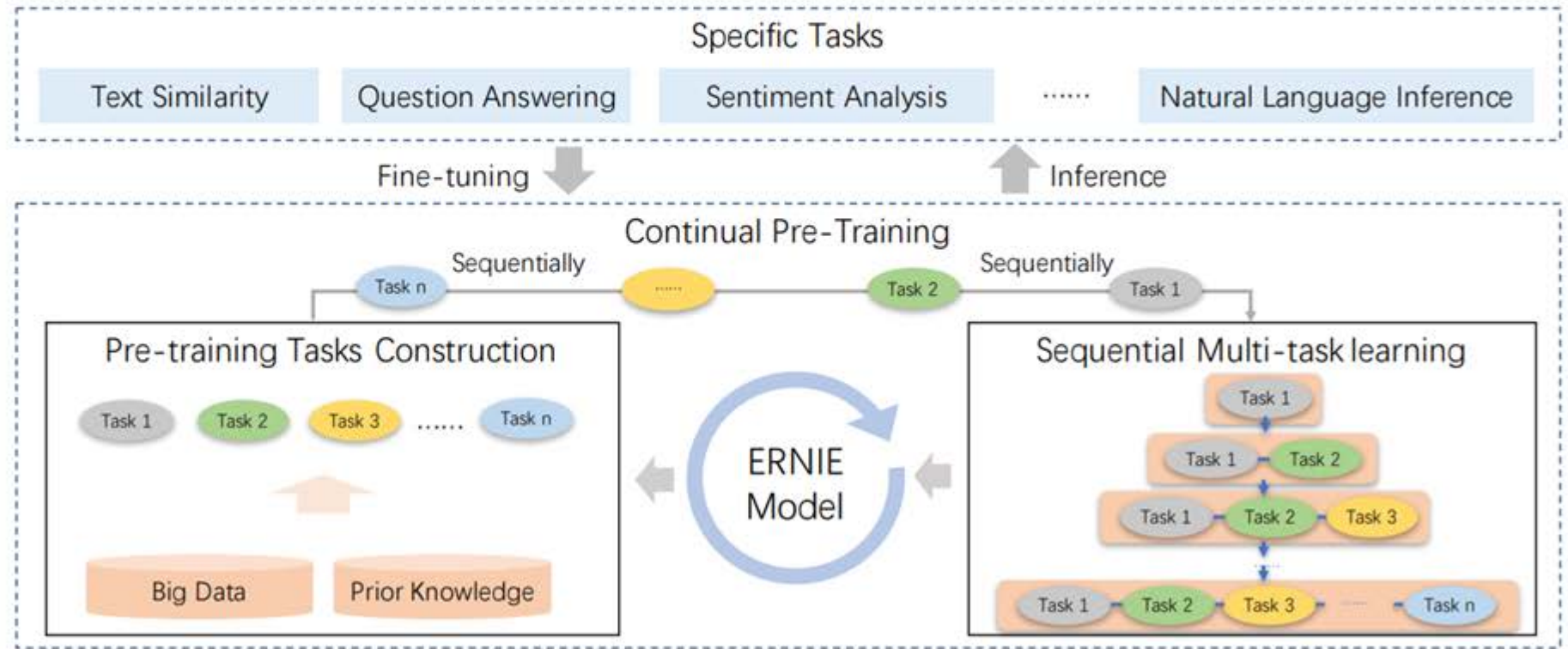
AlexaTM 20B is a 20B-Parameter sequence-to-sequence transformer model created by the Alexa Teacher Model (AlexaTM) team at Amazon. The model was trained on a **mixture of Common Crawl (mC4)** and Wikipedia data across 12 languages using **denoising and Causal Language Modeling (CLM)** tasks.

AlexaTM 20B can be used for in-context learning. "In-context learning," also known as "**prompting**," refers to a method for using NLP models in which no fine tuning is required per task.

Training examples are provided to the model only as part of the prompt given as inference input, a paradigm known as "few-shot in-context learning." In some cases, the model can perform well without any training data at all, a paradigm known as "**zero-shot in-context learning**."



Baidu ERNIE-VILG



ChatGPT – pretext Process

PRETRAIN

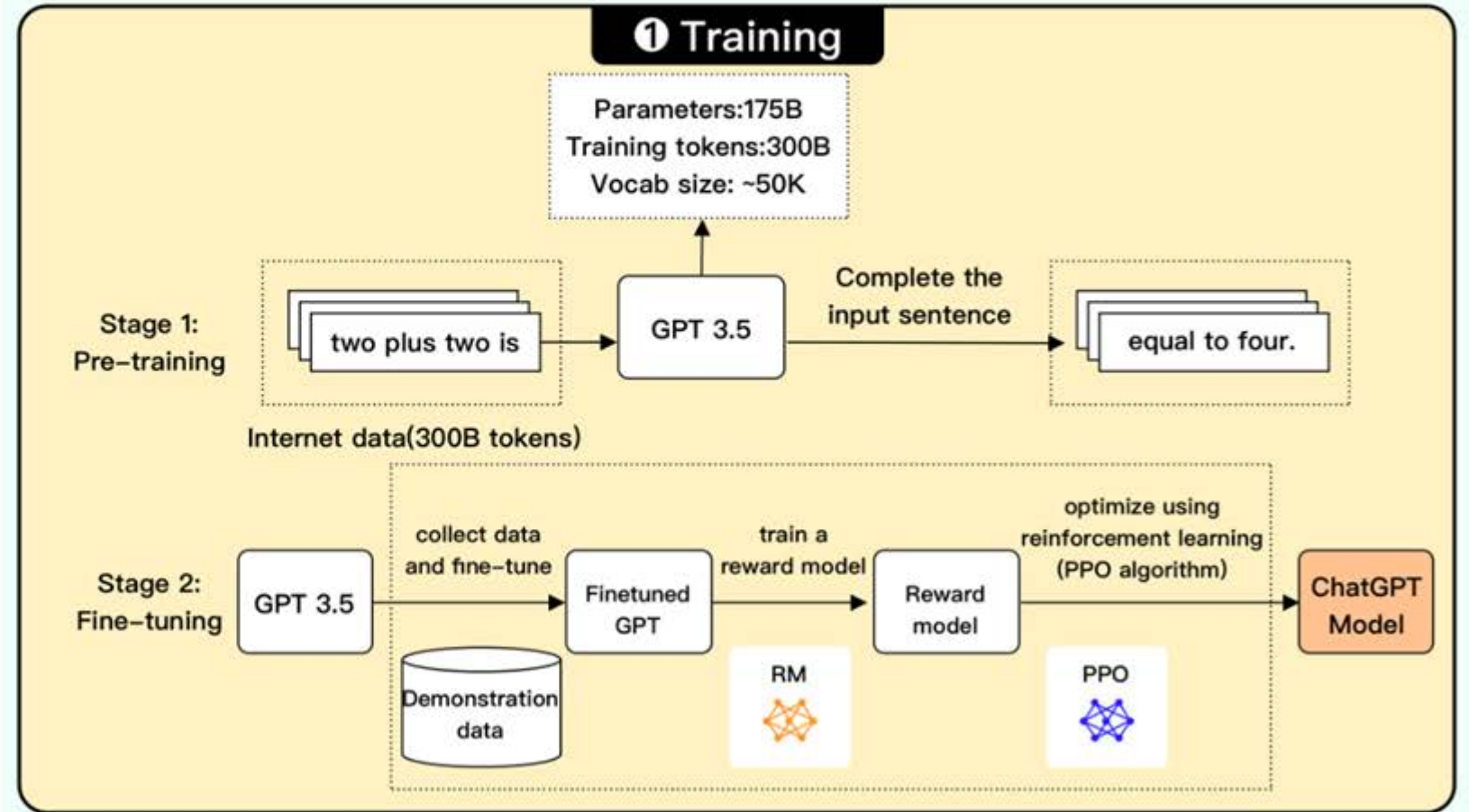
Pretrain: Train a GPT model (decoder) that can predict future words that is grammatically correct and semantically meaningful

END OF PRETRAIN

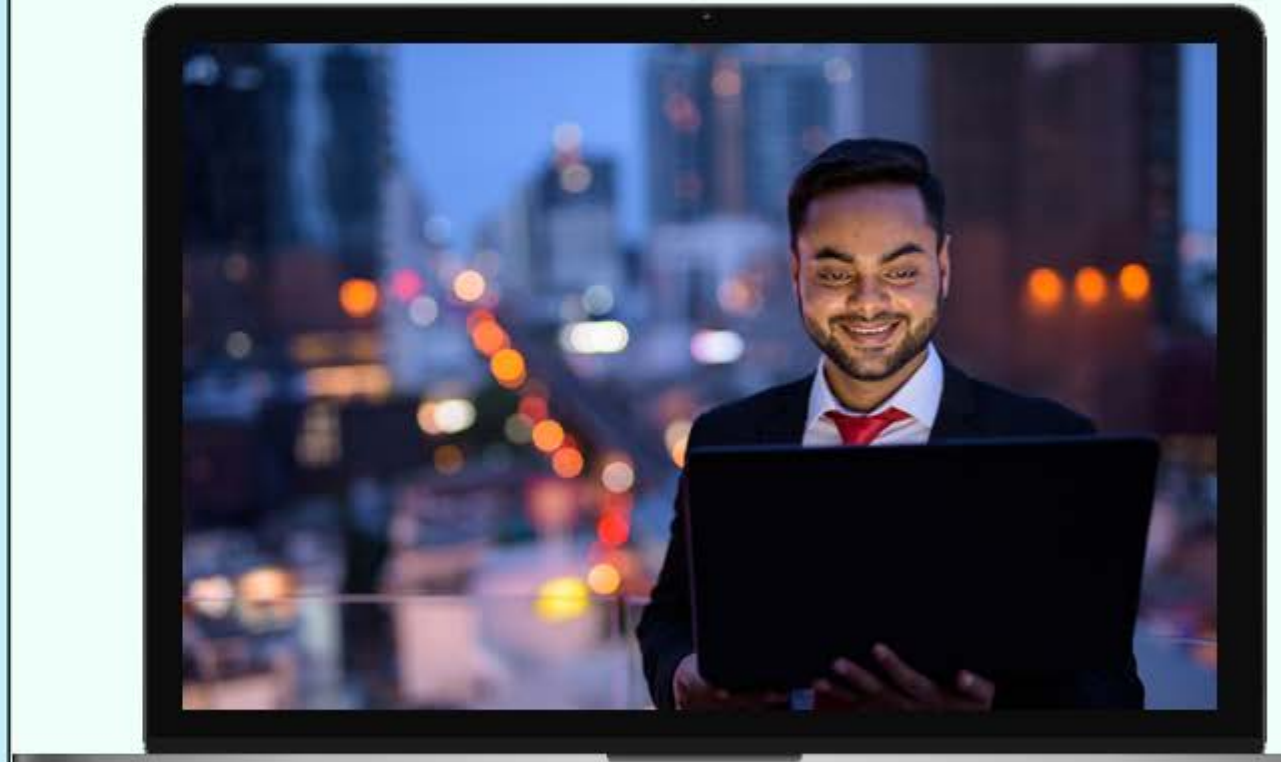
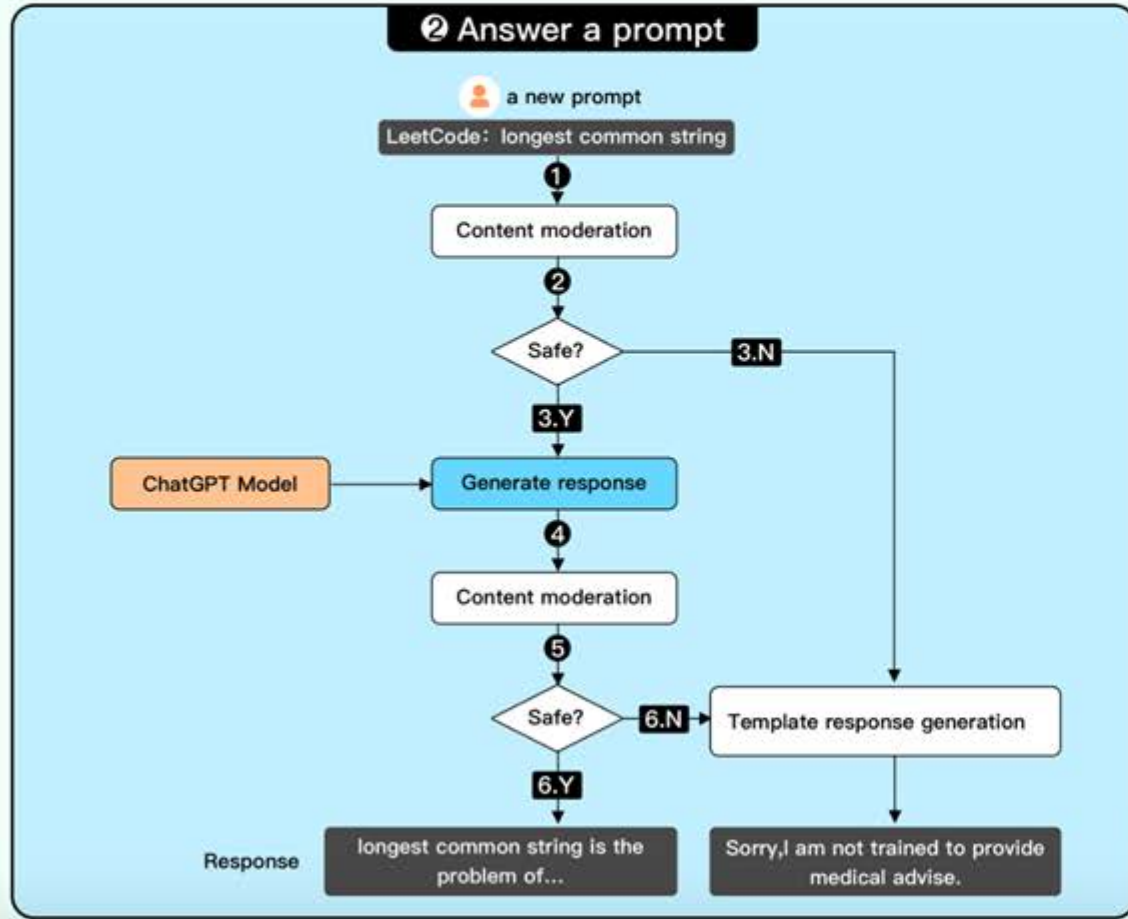
After the pre-training stage, the model can complete given sentences, but it is not capable of responding to questions.



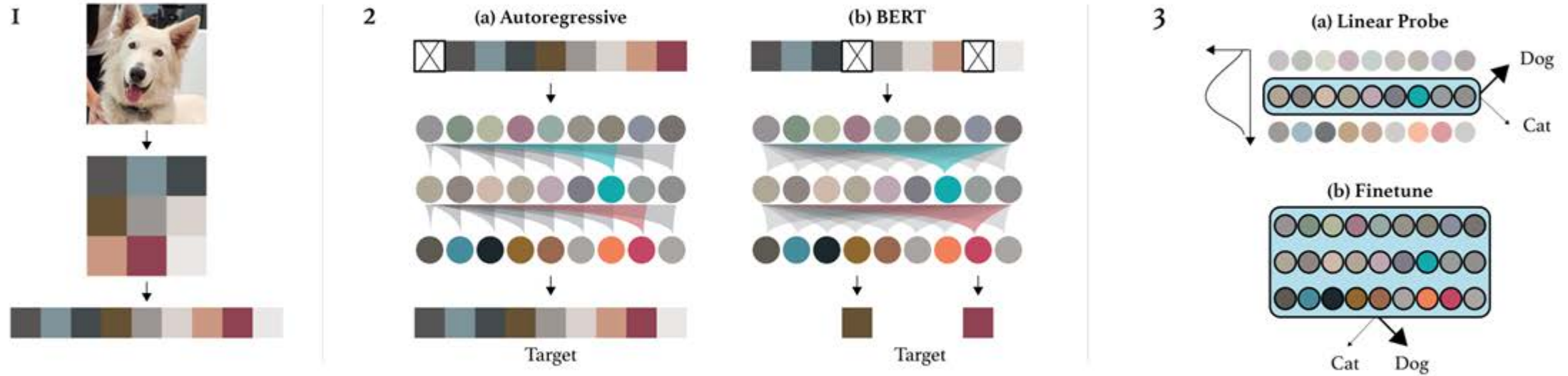
How is ChatGPT trained



Working of chatGPT



GPT-Image



iGPT-S, iGPT-M, and iGPT-L, transformers containing 76M, 455M, and 1.4B parameters respectively, on ImageNet.

iGPT-XL , a 6.8 billion parameter transformer, on a mix of ImageNet and images from the web.

GPT- Image



EVALUATION	DATASET	OUR RESULT	BEST NON-iGPT RESULT
Logistic regression on learned features (linear probe)	CIFAR-10	96.3 iGPT-L 32x32 w/ 1536 features	95.3 SimCLR ¹² w/ 8192 features
	CIFAR-100	82.8 iGPT-L 32x32 w/ 1536 features	80.2 SimCLR w/ 8192 features
	STL-10	95.5 iGPT-L 32x32 w/ 1536 features	94.2 AMDIM ¹³ w/ 8192 features
	ImageNet	72.0 iGPT-XL ^a 64x64 w/ 15360 features	76.5 SimCLR w/ 8192 features
Full fine-tune	CIFAR-10	99.0 iGPT-L 32x32, trained on ImageNet	99.0^b GPipe, ¹⁵ trained on ImageNet
	ImageNet 32x32	66.3 iGPT-L 32x32	70.2 Isometric Nets ¹⁶

10 YEARS OF UNIVERSITY RECOGNITION
20 YEARS OF ACADEMIC EXCELLENCE

 **REVA**
UNIVERSITY
Bengaluru, India



AI Adoption Pros

Unfolding use-cases of
AI adoption

What Next

(Customizable AI)

Ready to Use AI

What Dall-E and Chat GPT are AIs enable anyone to benefit

Ready to plug AI

Like Yolo Object Detection (Model available through a API)

It is hard to see in Healthcare, supply chain data it is hard to see one model fitting all

Are you part of this AI democratization shaping in the industry?



Marketing **Fraternity**

They are currently using ChatGPT to:

- Inspire ad copy or subject line versions (about one in five are somewhat on target).
- Create thought leadership content (provides some good ideas and direction).
- Supplement or enhance job postings (came up with some stuff that help round out a job rec).
- Write alternate code versions (surprisingly good).
- Power website chatbots (still testing, but this should be a layup).



But it's something marketers are paying attention to.

Improvement in Customer Experience

- Expect much experimentation in 2023 and new products emerging to add business value to ChatGPT.
- how support agents respond to consumers, either using automated bots or quickly getting an answer by leveraging ChatGPT on their own dataset
- Companies that adopt these models to improve their existing chatbot's ability to recognize and respond to emotions in interactions and other capabilities will be well-positioned to provide improved customer support and experience
- Current developments are just the tipping point for adopting web chat solutions with generative AI abilities.
- These will be ubiquitous across **B2B and B2C**



AI Adoption Cons

Unfolding use-cases of AI adoption



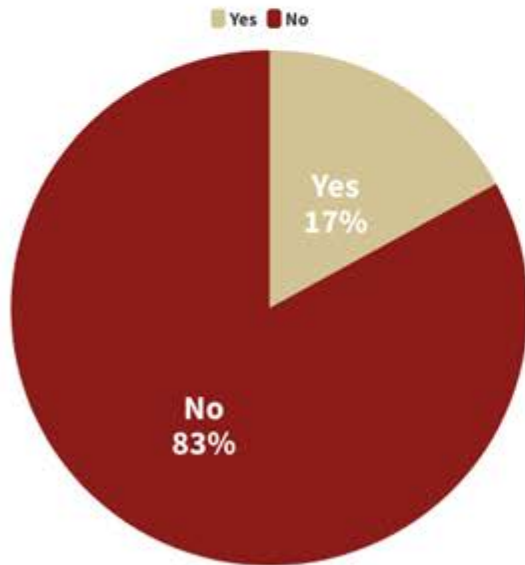
Stanford use-case

Chat GPT - to complete assignments and exams

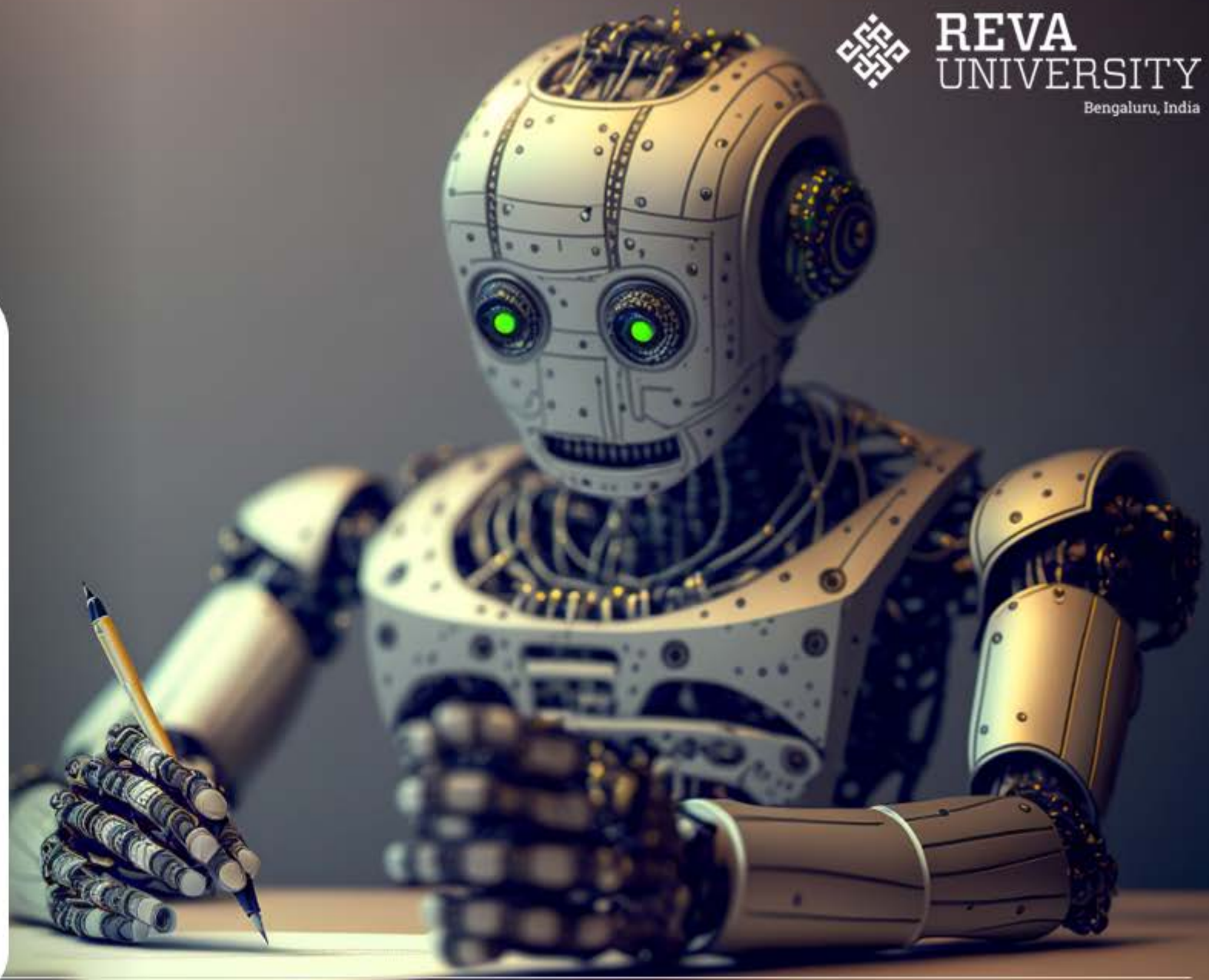
The Stanford Daily

ChatGPT Use in Fall Quarter 2022 Finals

Did you use ChatGPT on any fall quarter final assignments or exams?



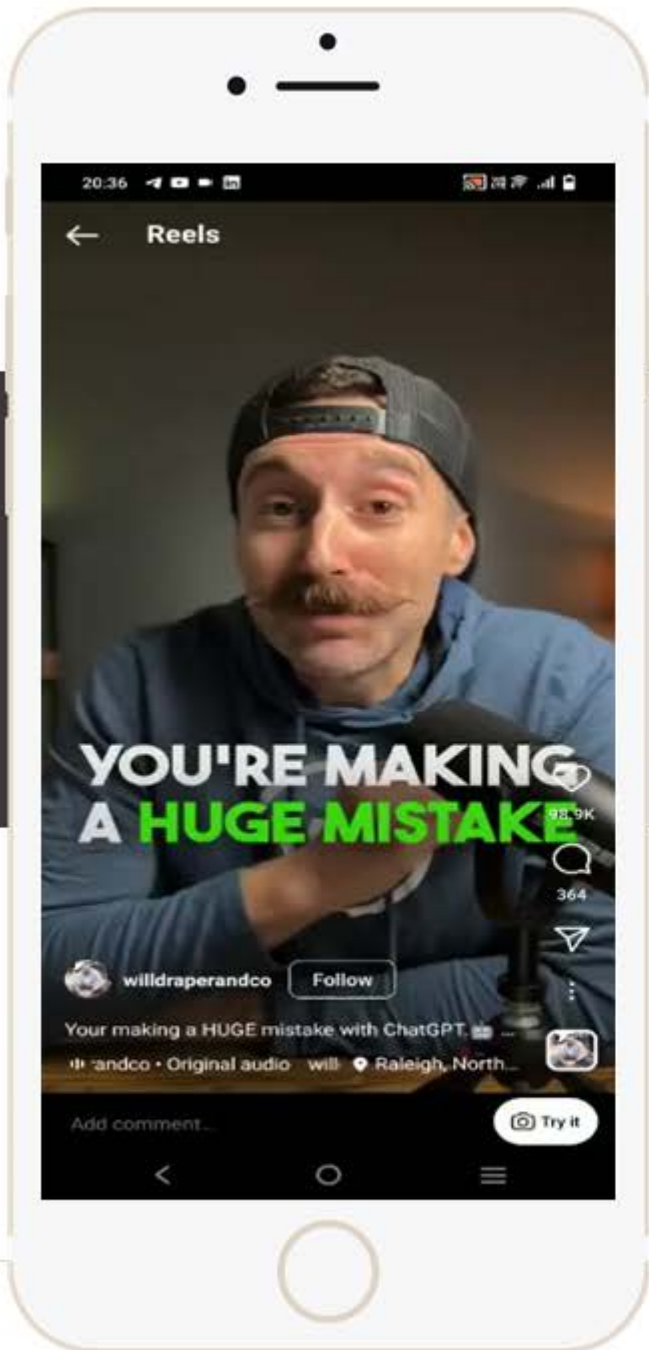
Source: [Fizz Post](#) • In an anonymous poll uploaded on Fizz, 4,497 Stanford students answered whether or not they used ChatGPT to help with their Fall final assignments.
Graphic: MARK ALLEN CU/The Stanford Daily



AI Plagarism

No end to it

ChatGPT is **not yet at a tool** that can run without close supervision.



AI Copyright Lawsuit



POLICY / TECH / LAW

Microsoft, GitHub, and OpenAI ask court to throw out AI copyright lawsuit



/ The three companies want to dismiss a complaint that alleges that the AI-powered Copilot relies on 'software piracy on an unprecedented scale.'

By **EMMA ROTH**

Jan 29, 2023, 5:32 AM GMT+5:30 | □



Microsoft, GitHub, and OpenAI claim the lawsuit doesn't hold up. Photo: Jens Büttner/dpa-Zentralbild/ZB (Photo by Jens Büttner/picture alliance via Getty Images)

My Experience



Hi everyone, 🌸



We cordially invite you to join us for an upcoming free webinar on the topic of Self-Supervised Learning (SSL) and its potential for large-scale AI adoption.

Our esteemed speaker, Dr Sai Hareesh, will dive into the use cases of SSL, specifically highlighting its implementation in ChatGPT models.

The free webinar will take place on February 19, 2023, at 12:00 PM IST. To Register for the event, please follow the link:

👉 <https://bit.ly/3YmKPmj> 👈

Upon successful registration, you will receive a confirmation email with instructions on how to join the webinar.

We look forward to your participation and engagement in this informative session.



Thank you for the invitation. As a language model developed by OpenAI, I am unable to attend the webinar in a physical sense. However, I would be happy to answer any questions you may have about Self-Supervised Learning (SSL) and its potential for large-scale AI adoption. Please feel free



🔄 Regenerate response

My Experience



Can you talk about sai hareesh's talk on self supervised learning on youtube



I'm sorry, but I am not aware of any specific talk by Dr. Sai Hareesh on Self-Supervised Learning (SSL) on YouTube. As a language model, I do not have access to current or future events unless they have taken place and been recorded and publicly available at my knowledge cutoff of 2021. If you have a specific question about SSL or



Local context from Global context is taken care by finetuning to your subsystem

PAST. PRESENT. FUTURE. - > Predict present from past or future from present is a predictive analytics

GPT is a good first step towards -> **Preventive analytics**, especially in cyber security and health care

Open source solution replicates ChatGPT training process! Ready to go with only 1.6GB GPU memory and gives you 7.73 times faster training!

Step 1

Collect demonstration data and train a supervised policy.

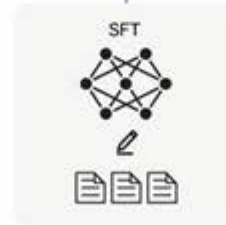
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



This data is used to fine-tune GPT-3.5 with supervised learning.



Step 2

Collect comparison data and train a reward model.

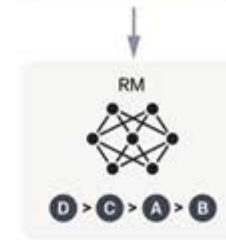
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



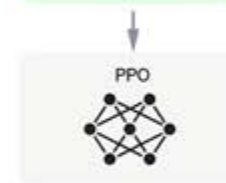
Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.



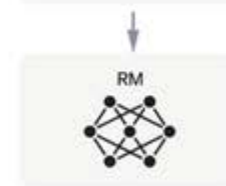
The PPO model is initialized from the supervised policy.



The policy generates an output.



The reward model calculates a reward for the output.

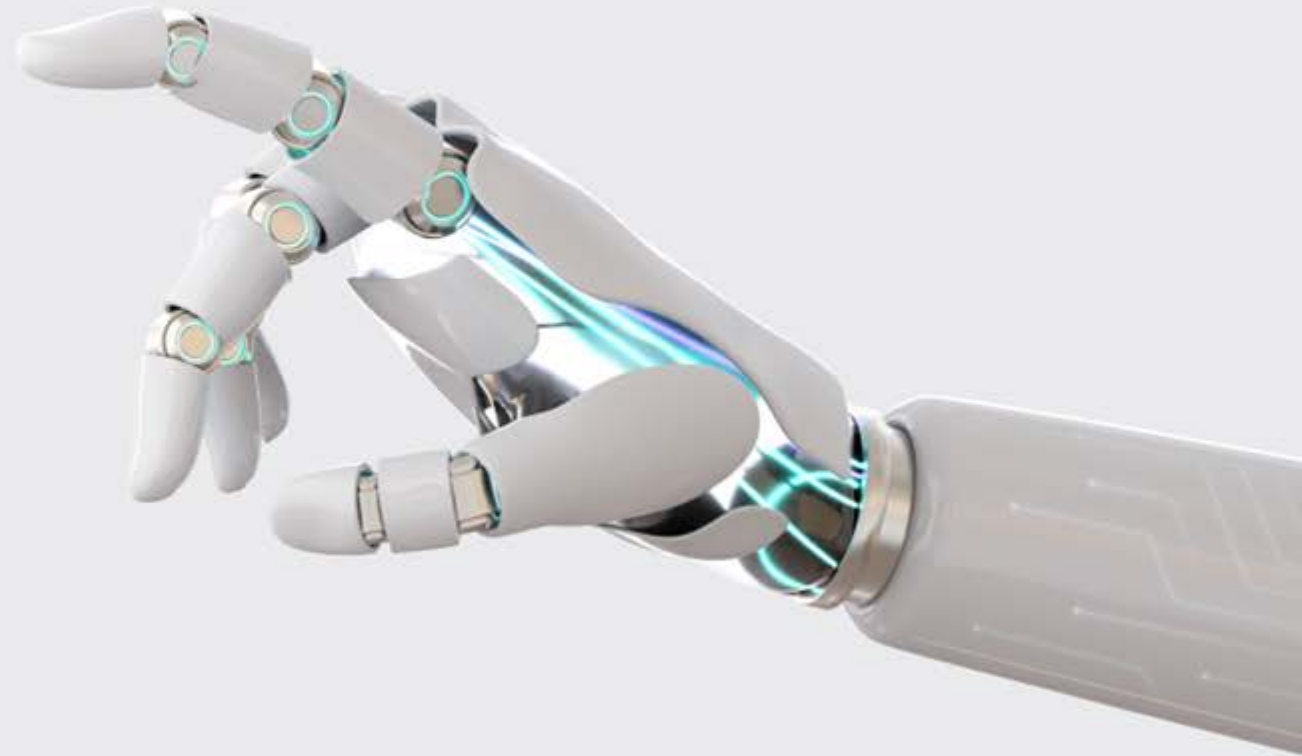


The reward is used to update the policy using PPO.



Explanability is always our Responsibility

- Whom will be responsible if chat GPT gives out sensitive information
- AI adoption will definitely improve but be used with caution
- Needs to explain causality -> what caused this response





Not on Internet Food for thought

- Converting one form of GPT to other
- For example Image GPT data synthesized into a Chat GPT or summary what the image conveys
- Always the use-cases are looking at the same modality till now,
- Integrated system
- Use-cases:
 - can create Instagram memes/reels easily
 - AI for storytelling



**To know more
about our Programs,
Please contact**

+91 89040 58866

+91 95388 74441 | +91 76250 69676

+91 99025 46881 | +91 99453 95881

For Research and
Consulting opportunities, please contact

Dr. Shinu Abhi

Director, Corporate Training

REVA Academy for Corporate Excellence (RACE)

P: +91 99729 16030

E: shinuabhi@reva.edu.in

10 YEARS OF UNIVERSITY RECOGNITION
20 YEARS OF ACADEMIC EXCELLENCE

REVA UNIVERSITY
Bengaluru, India

Thank you

