



جامعة الأمير سكام بن عبدالعزيز  
PRINCE SATTAM BIN ABDULAZIZ UNIVERSITY

# Prompt-based Approach for NER and Keyphrase Extraction

Subproject 4: Comparison of Prompt-based Approach with Fine-  
tuned BERT Models

2024-December

IS 661 – Text Mining  
Dr. Mohsin Bila  
Amer Basha – 445540038  
Ahmed Haddadi – 445540039  
Rakan Alqahtani - 445540040  
Omar Alotabi - 445540032

# CONTENTS

---

## 1. Introduction

- 1.1 Overview
- 1.2 Objective and Problem Statement
- 1.3 Proposed Solution

## 2. Literature Review

- 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer
- 2.2 Transformers are Meta-Reinforcement Learners

## 3. Named Entity Recognition with BERT and GPT-3.5

- 3.1 Named Entity Recognition
- 3.2 Named Entity Recognition Methodology
- 3.3 Named Entity Recognition Using BERT
- 3.3 openAI (GPT-3.5 prompt Engineering)

## 4. On the Opportunities and Risks of Foundation Models

## 5. Discussion



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and gray. The diagram is partially cut off by the top and left edges of the slide.

# 1. Introduction

# 1. Introduction

---

## 1.1 Overview

In recent years, Natural Language Processing (NLP) has made significant strides, largely due to the advent of transfer learning models based on transformers. Models such as BERT and GPT-3.5 have proven highly effective in performing a wide range of linguistic tasks, owing to their ability to leverage vast amounts of pre-trained data

## 1.2 Objective and Problem Statement

Despite the progress, tasks like Named Entity Recognition (NER) and keyphrase extraction remain challenging. Traditional methods, including fine-tuned BERT models, deliver strong results but often require extensive computational resources and time-consuming fine-tuning. The key challenge is to develop more efficient methods that can maintain or improve performance while reducing the need for such intensive customization.

## 1.3 Proposed Solution

To address this, we propose a prompt-based approach utilizing large language models (LLMs) such as GPT-3.5 and T5. By designing tailored prompts, we aim to guide these models to perform NER and keyphrase extraction effectively. This approach not only offers a reduction in fine-tuning complexity but also provides flexibility to apply across various linguistic contexts.

Through this project, we will compare the effectiveness of this method with fine-tuned models, particularly using the WikiAnn dataset for NER.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The diagram is partially cut off by the top and left edges of the slide.

## 2. Literature Review

# 2.1

## Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer

### Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer

Colin Raffel\*

Noam Shazeer\*

Adam Roberts\*

Katherine Lee\*

Sharan Narang

Michael Matena

Yang Zhou

Wei Li

Peter J. Liu

Google, Mountain View, CA 94043, USA

CRAFFEL@GMAIL.COM

NOAM@GOOGLE.COM

ADAROB@GOOGLE.COM

KATHERINELEE@GOOGLE.COM

SHARANANRANG@GOOGLE.COM

MMATENA@GOOGLE.COM

YANGZ@GOOGLE.COM

WVLI@GOOGLE.COM

PETERLIU@GOOGLE.COM

Editor: Ilya Sutskever

#### Abstract

Transfer learning, where a model is first pre-trained on a data-rich task before being fine-tuned on a downstream task, has emerged as a powerful technique in natural language processing (NLP). The effectiveness of transfer learning has given rise to a diversity of approaches, methodology, and practice. In this paper, we explore the landscape of transfer learning techniques for NLP by introducing a unified framework that converts all text-based language problems into a text-to-text format. Our systematic study compares pre-training objectives, architectures, unlabeled data sets, transfer approaches, and other factors on dozens of language understanding tasks. By combining the insights from our exploration with scale and our new “Colossal Clean Crawled Corpus”, we achieve state-of-the-art results on many benchmarks covering summarization, question answering, text classification, and more. To facilitate future work on transfer learning for NLP, we release our data sets, pre-trained models, and code.

**Keywords:** transfer learning, natural language processing, multi-task learning, attention-based models, deep learning

#### 1. Introduction

Training a machine learning model to perform natural language processing (NLP) tasks often requires that the model can process text in a way that is amenable to downstream learning. This can be loosely viewed as developing general-purpose knowledge that allows the model to “understand” text. This knowledge can range from low-level (e.g. the spelling

\* Equal contribution. A description of each author’s contribution is available in Appendix A. Correspondence to: craffel@gmail.com.

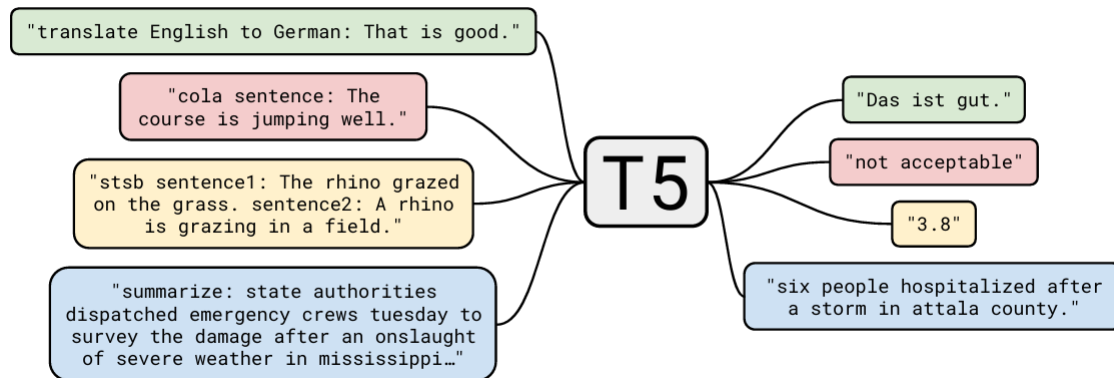
1. <https://github.com/google-research/text-to-text-transfer-transformer>

## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

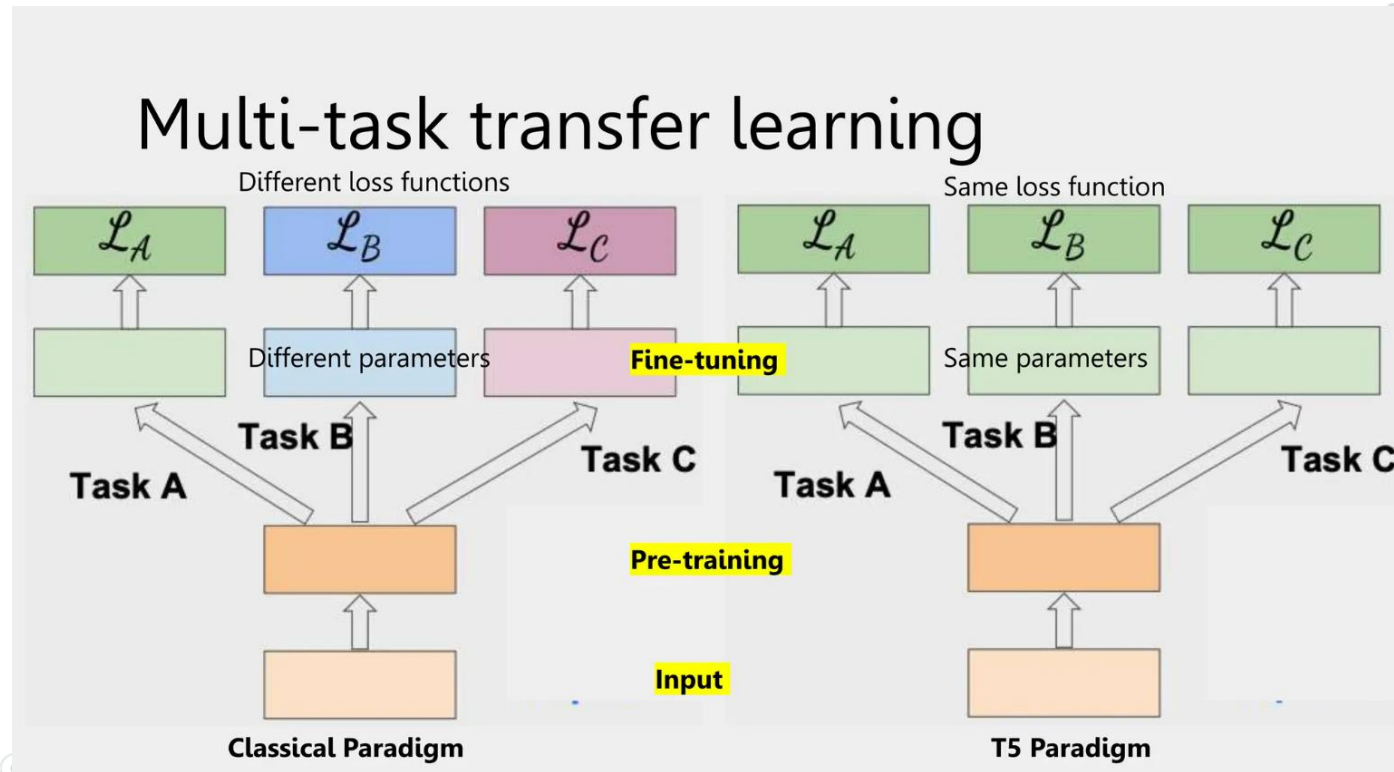
### Text-to-Text Transfer Transformer:

Use the complete encoder – decoder.

Pretrained with Clean dataset: “Colossal Clean Crawled Corpus” (C4).



## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :





## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

### Setup: Fine-tuning tasks

- Text classification: GLUE and SuperGLUE.
- Abstractive summarization: CNN/Daily Mail
- Question Answering: SQuAD
- Translation: WMT English to German, French, and Romanian

### Setup: Input & Output

- "text-to-text" format: Preprocessed Examples in Appendix D in T5  
(Figure 1: Preprocessed Examples)
- consistent training objective: maximum likelihood
- task-specific (text) prefix
- Mismatch label Issue
  - e.g. given a premise and hypothesis, classify into one of 3 categories - 'entailment', 'contradiction' and 'neutral'
  - Potentially possible for decoder to output 'hamburger'
  - This issue never observed with their trained models

#### D.1. CoLA

Original input:

Sentence: John made Bill master of himself.

Processed input: cola sentence: John made Bill master of himself.

Original target: 1

Processed target: acceptable

#### D.2. RTE

Original input:

Sentence 1: A smaller proportion of Yugoslavia's Italians were settled in Slovenia (at the 1991 national census, some 3000 inhabitants of Slovenia declared themselves as ethnic Italians).

Sentence 2: Slovenia has 3,000 inhabitants.

Processed input: rte sentence1: A smaller proportion of Yugoslavia's Italians were settled in Slovenia (at the 1991 national census, some 3000 inhabitants of Slovenia declared themselves as ethnic Italians). sentence2: Slovenia has 3,000 inhabitants.

Original target: 1

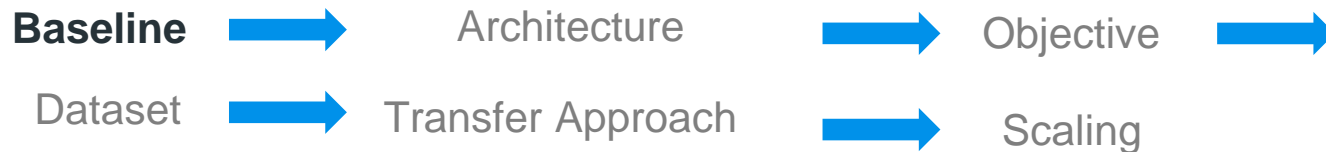
Processed target: not\_entailment

(Figure 1: Preprocessed Examples)

## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

T5: Exploring the Limits of Transfer Learning with a unified Text-to-Text Transformer

### Empirical Survey



## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

---

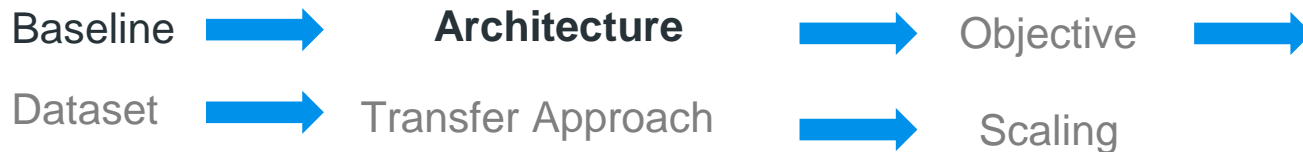
### Baseline (Pre-training details)

- Max Sequence length: 512 tokens
- Batch size: 128 sequences
- Training size = 219 steps
- Constant learning rate = 0.001
- 5,000 Steps/Checkpoint

## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

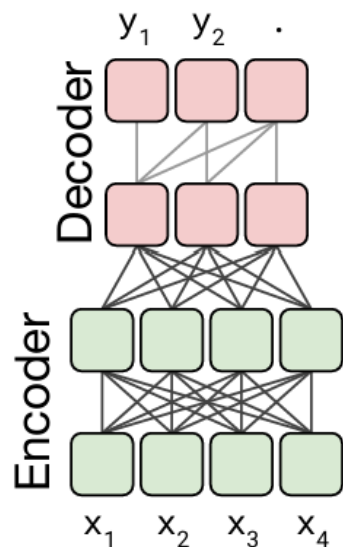
T5: Exploring the Limits of Transfer Learning with a unified Text-to-Text Transformer

### Empirical Survey

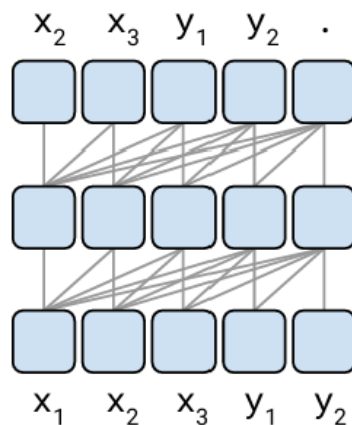


## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

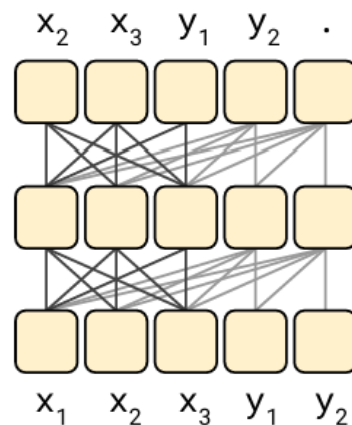
### Architectural Variants



Language model



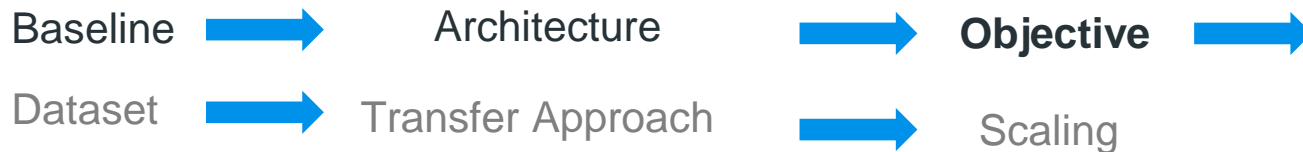
Prefix LM



## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

T5: Exploring the Limits of Transfer Learning with a unified Text-to-Text Transformer

### Empirical Survey



## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

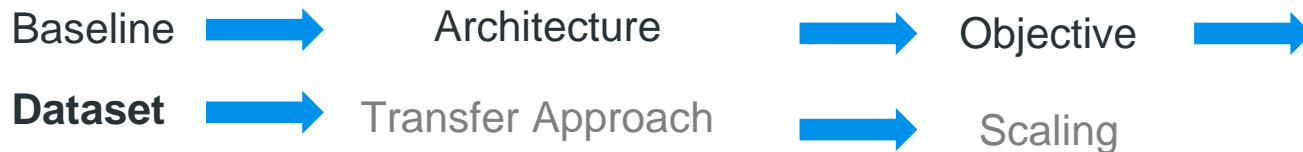
### Pre-training: BERT VS Non-BERT Style

Objective	GLUE	CNNDM	SQuAD	SGLUE	EnDe	EnFr	EnRo
Prefix language modeling	80.69	18.94	77.99	65.27	<b>26.86</b>	39.73	<b>27.49</b>
BERT-style (Devlin et al., 2018)	<b>82.96</b>	<b>19.17</b>	<b>80.65</b>	<b>69.85</b>	<b>26.78</b>	<b>40.03</b>	<b>27.41</b>
Deshuffling	73.17	18.59	67.61	58.47	26.11	39.30	25.62

## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

T5: Exploring the Limits of Transfer Learning with a unified Text-to-Text Transformer

### Empirical Survey





## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

### **Setup: Pre-training Dataset (Data Cleaning for C4 Dataset):**

- Retaining Valid Sentences: Only lines ending with terminal punctuation (period, exclamation mark, question mark, or quotation mark) were kept.
- Sentence and Word Filtering: Discarded pages with fewer than 5 sentences and kept lines containing at least 3 words.
- Profanity and Inappropriate Content: Removed pages with words from a “dirty” word list.
- JavaScript Warnings: Eliminated lines mentioning "Javascript".
- Placeholder Text: Removed pages with “lorem ipsum” text.
- Code Content: Removed pages containing curly brackets “{”, common in code but not natural text.
- Deduplication: Kept only one instance of any three-sentence span appearing multiple times.
- Language Filtering: Used langdetect to retain only English pages with a probability of at least 0.99.

This resulted in the C4 (Colossal Clean Crawled Corpus) dataset, a large (750 GB) and clean English text resource for NLP tasks.

This dataset is more refined than previous datasets, which had limited filtering and were often not publicly available.

### **Source:**

Data collected from Common Crawl, a large publicly-available web archive (20/TB/month).

[” https://commoncrawl.org/”](https://commoncrawl.org/)

## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

---

### Pre-training Datasets

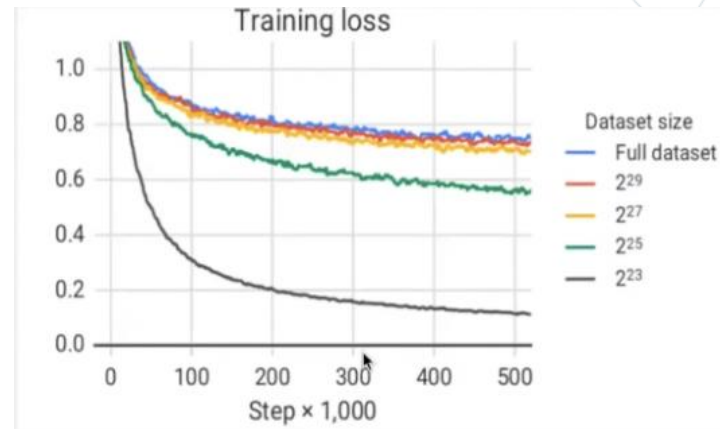
- C4: Common Crawl with heuristic filterin.
- Unfiltered C4: Common Crawl only use langdetect to extract English text.
- RealNews-like (GPT2-like): high Reddit score webpages in C4.
- WebText-like (GBT2-like): high Reddit score webpages in C4.
- Wikipedia.
- Wikipedia + Toronto Books Corpus (BERT).

## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

### Pre-training Datasets

#### Varying No. of epochs

- Keeping total number of Training steps = constant

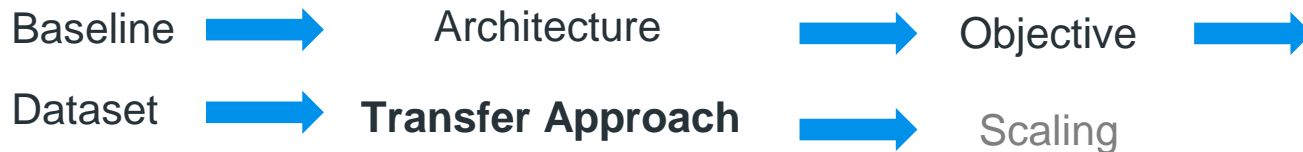


Number of tokens	Repeats	GLUE	CNNDM	SQuAD	SGLUE	EnDe	EnFr	EnRo
★ Full dataset	0	<b>83.28</b>	<b>19.24</b>	<b>80.88</b>	<b>71.36</b>	<b>26.98</b>	<b>39.82</b>	<b>27.65</b>
$2^{29}$	64	<b>82.87</b>	<b>19.19</b>	<b>80.97</b>	<b>72.03</b>	<b>26.83</b>	<b>39.74</b>	<b>27.63</b>
$2^{27}$	256	82.62	<b>19.20</b>	79.78	69.97	<b>27.02</b>	<b>39.71</b>	27.33
$2^{25}$	1,024	79.55	18.57	76.27	64.76	26.38	39.56	26.80
$2^{23}$	4,096	76.34	18.33	70.92	59.29	26.37	38.84	25.81

## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

### Empirical Survey

Methodology “coordinate descent”



## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

---

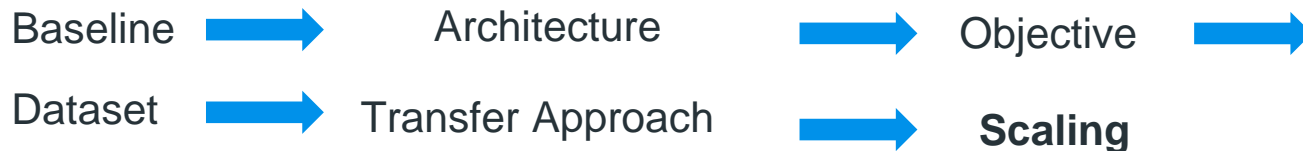
### Transfer Approach

- The Transfer Approach involves utilizing a pre-trained model on an initial dataset and then fine-tuning or retraining it for a new task.
- The goal of this method is to leverage the knowledge acquired from a prior task to reduce the need for extensive new training data.
- The approach is discussed in the context of "Scaling," indicating that expanding or adapting to different datasets or environments may be a key challenge or focus in the analysis.

## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

### Empirical Survey

Methodology “coordinate descent”



## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

### Scaling

- Scaling refers to expanding the model's ability to handle larger, more complex tasks or datasets. It involves:
  1. Increasing Model Depth: Adding more layers to the model to learn complex patterns and details.
  2. Increasing Computational Resources: Using more CPUs or GPUs to improve training and efficiency.
  3. Improving Training Algorithms: Utilizing advanced algorithms to maximize the benefits of increased resources and data.
  4. Handling Larger Datasets: Enabling the model to process and learn from larger datasets, improving accuracy.
  5. Scaling to More Complex Tasks: Allowing the model to handle more challenging tasks, improving its versatility.

#### Impact of Scaling:

- Improved Performance: The model handles bigger problems.
- Increased Efficiency: Faster and more effective training.
- Better Accuracy: Learning from a larger variety of data.

## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

### Model

- Objective: span-corruption (SpanBERT) with span length 3
- Longer training: 1M steps with batch size 2048  $\rightarrow$  1T tokens
  - 8x BERT, 2x XLNet, 1/2 x ROBERTa
- Model sizes:
  - Small: 60M Base: 220M Large: 770M XLarge: 3B  
XXLarge: 11B
- Multi-task pre-training (MT-DNN):
  - Monitor downstream task performance while pre-training
- Finetune on GLUE and SuperGLUE: 8 batch size



## 2.1 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer :

Model	GLUE Average	CoLA Matthew's	SST-2 Accuracy	MRPC F1	MRPC Accuracy	STS-B Pearson	STS-B Spearman
Previous best	89.4 <sup>a</sup>	69.2 <sup>b</sup>	<b>97.1<sup>a</sup></b>	<b>93.6<sup>b</sup></b>	<b>91.5<sup>b</sup></b>	<b>92.7<sup>b</sup></b>	<b>92.3<sup>b</sup></b>
T5-Small	77.4	41.0	91.8	89.7	86.6	85.6	85.0
T5-Base	82.7	51.1	95.2	90.7	87.5	89.4	88.6
T5-Large	86.4	61.2	96.3	92.4	89.9	89.9	89.2
T5-3B	88.5	67.1	97.4	92.5	90.0	90.6	89.8
T5-11B	<b>89.7</b>	<b>70.8</b>	<b>97.1</b>	91.9	89.2	92.5	92.1

Model	QQP F1	QQP Accuracy	MNLI-m Accuracy	MNLI-mm Accuracy	QNLI Accuracy	RTE Accuracy	WNLI Accuracy
Previous best	<b>74.8<sup>c</sup></b>	<b>90.7<sup>b</sup></b>	91.3 <sup>a</sup>	91.0 <sup>a</sup>	<b>99.2<sup>a</sup></b>	89.2 <sup>a</sup>	91.8 <sup>a</sup>
T5-Small	70.0	88.0	82.4	82.3	90.3	69.9	69.2
T5-Base	72.6	89.4	87.1	86.2	93.7	80.1	78.8
T5-Large	73.9	89.9	89.9	89.6	94.8	87.2	85.6
T5-3B	74.4	89.7	91.4	91.2	96.3	91.1	89.7
T5-11B	74.6	90.4	<b>92.0</b>	<b>91.7</b>	96.7	<b>92.5</b>	<b>93.2</b>

Model	SQuAD EM	SQuAD F1	SuperGLUE Average	BoolQ Accuracy	CB F1	CB Accuracy	COPA Accuracy
Previous best	88.95 <sup>d</sup>	94.52 <sup>d</sup>	84.6 <sup>e</sup>	87.1 <sup>e</sup>	90.5 <sup>e</sup>	95.2 <sup>e</sup>	90.6 <sup>e</sup>
T5-Small	79.10	87.24	63.3	76.4	56.9	81.6	46.0
T5-Base	85.44	92.08	76.2	81.4	86.2	94.0	71.2
T5-Large	86.66	93.79	82.3	85.4	91.6	94.8	83.4
T5-3B	88.53	94.95	86.4	89.9	90.3	94.4	92.0
T5-11B	<b>90.06</b>	<b>95.64</b>	<b>88.9</b>	<b>91.0</b>	<b>93.0</b>	<b>96.4</b>	<b>94.8</b>

Model	MultiRC F1a	MultiRC EM	ReCoRD F1	ReCoRD Accuracy	RTE Accuracy	WiC Accuracy	WSC Accuracy
Previous best	84.4 <sup>e</sup>	52.5 <sup>e</sup>	90.6 <sup>e</sup>	90.0 <sup>e</sup>	88.2 <sup>e</sup>	69.9 <sup>e</sup>	89.0 <sup>e</sup>
T5-Small	69.3	26.3	56.3	55.4	73.3	66.9	70.5
T5-Base	79.7	43.1	75.0	74.2	81.5	68.3	80.8
T5-Large	83.3	50.7	86.8	85.9	87.8	69.3	86.3
T5-3B	86.8	58.3	91.2	90.4	90.7	72.1	90.4
T5-11B	<b>88.2</b>	<b>62.3</b>	<b>93.3</b>	<b>92.5</b>	<b>92.5</b>	<b>76.1</b>	<b>93.8</b>

Model	WMT EnDe BLEU	WMT EnFr BLEU	WMT EnRo BLEU	CNN/DM ROUGE-1	CNN/DM ROUGE-2	CNN/DM ROUGE-L
Previous best	<b>33.8<sup>f</sup></b>	<b>43.8<sup>f</sup></b>	<b>38.5<sup>g</sup></b>	43.47 <sup>h</sup>	20.30 <sup>h</sup>	40.63 <sup>h</sup>
T5-Small	26.7	36.0	26.8	41.12	19.56	38.35
T5-Base	30.9	41.2	28.0	42.05	20.34	39.40
T5-Large	32.0	41.5	28.1	42.50	20.68	39.75
T5-3B	31.8	42.6	28.2	42.72	21.02	39.94
T5-11B	32.1	43.4	28.1	<b>43.52</b>	<b>21.55</b>	<b>40.69</b>

## 2.2

# Transformers are Meta-Reinforcemem Learners

### Abstract

The transformer architecture and variants presented a remarkable success across many machine learning tasks in recent years. This success is intrinsically related to the capability of handling long sequences and the presence of context-dependent weights from the attention mechanism. We argue that these capabilities suit the central role of a Meta-Reinforcement Learning algorithm. Indeed, a meta-RL agent needs to infer the task from a sequence of trajectories. Furthermore, it requires a fast adaptation strategy to adapt its policy for a new task - which can be achieved using the self-attention mechanism. In this work, we present TMRL (Transformers for Meta-Reinforcement Learning), a meta-RL agent that mimics the memory reinstatement mechanism using the transformer architecture. It associates the recent past of working memories to build an episodic memory recurrently through the transformer layers. We show that the self-attention computes a consensus representation that minimizes the Hayes Risk at each layer and provides meaningful features to compute the best actions. We conducted experiments in high-dimensional continuous control environments for locomotion and discrete manipulation. Results show that TMRL presents comparable or superior performance, sample efficiency, and out-of-distribution generalization compared to the baselines in these environments.

### 1. Introduction

In recent years, the Transformer architecture (Vaswani et al., 2017) achieved exceptional performance on many machine learning applications, especially for text (Devlin et al., 2019; Raffel et al., 2020) and image processing (Dosovitskiy et al., 2020).

<sup>1</sup>Microsoft, USA <sup>2</sup>Center of Excellence in Artificial Intelligence (CEAI) Learning Branch, Brazil. Correspondence to: Luchiano C. Melo - luchiano@protonmail.com.

2020b; Caron et al., 2021; Yuan et al., 2021). This intrinsically relates to its few-shot learning nature (Brown et al., 2020b): the attention weights work as context-dependent parameters, inducing better generalization. Furthermore, this architecture parallelizes token processing by design. This property avoids backpropagation through time, making it less prone to vanishing/exploding gradients, a very common problem for recurrent models. As a result, they can handle longer sequences more efficiently.

This work argues that these two capabilities are essential for a Meta-Reinforcement Learning (meta-RL) agent. We propose TMRL (Transformers for Meta-Reinforcement Learning), a memory-based meta-Reinforcement Learner which uses the transformer architecture to formulate the learning process. It works as a memory reinstatement mechanism (Rovee-Collier, 2012) during learning, associating recent working memories to create an episodic memory which is used to contextualize the policy.

Figure 1 illustrates the process. We formulated each task as a distribution over working memories. TMRL associates these memories using self-attention blocks to create a task representation in each head. These task representations are combined in the position-wise MLP to create an episodic output (which we identify as episodic memory). We recurrently apply this procedure through layers to refine the episodic memory. In the end, we select the memory associated with the current timestep and feed it into the policy head.

Nonetheless, transformer optimization is often unstable, especially in the RL setting. Past attempts either fail to stabilize (Mishra et al., 2019) or require architectural additions (Pantano et al., 2019) or restrictions on the observations space (Loyal et al., 2020). We hypothesize that this challenge is because the instability of early stages of transformer optimization hampers initial exploration, which is crucial for environments where the learned behaviors must guide exploration to prevent poor policies. We argue that this challenge can be mitigated through a proper weight initialization scheme. For this matter, we applied T4-mp initialization (Huang et al., 2020).

We conducted a series of experiments to evaluate meta-

# Introduction

---

## Description:

WikiANN (sometimes called PAN-X) is a multilingual named entity recognition dataset consisting of Wikipedia articles annotated with LOC (location), PER (person), and ORG (organization) tags in the IOB2 format. This version corresponds to the balanced train, dev, and test splits of Rahimi et al. (2019), which supports 176 of the 282 languages from the original WikiANN corpus.

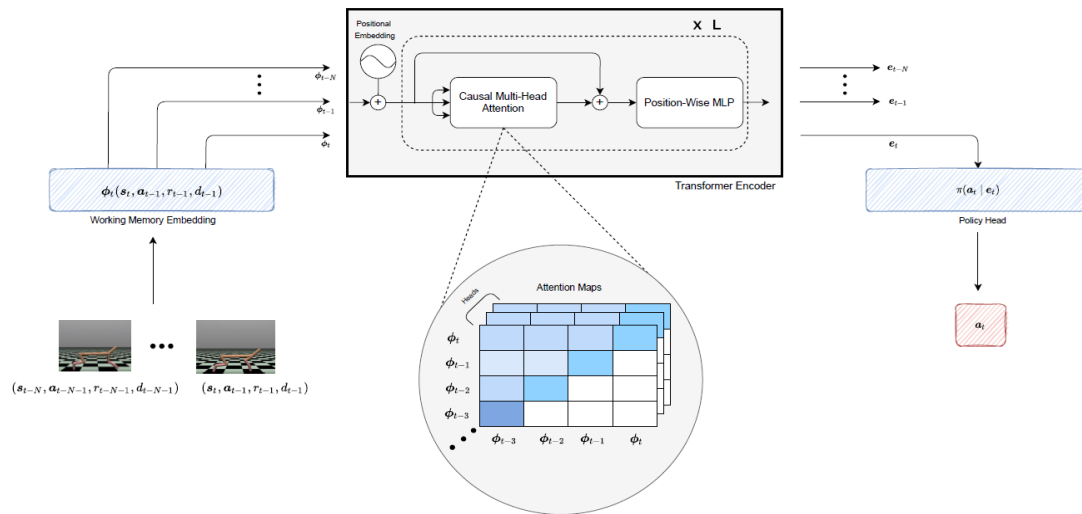
## Paper contents:

- Introduction and Study Objectives
- Related Work
- TrMRL Design
- Experiment Setup
- Experiment Results
- Conclusion and Future Directions
- Practical Implications
- Benefits:
- Challenges

# Transformers are Meta-Reinforcement Learners

## Introduction and Study Objectives

- The paper addresses the use of Transformers as a tool for Meta-Reinforcement Learning (Meta-RL).
- The characteristics of Transformers, such as handling long sequences and the self-attention mechanism, make them suitable for rapid adaptation to new tasks by building episodic memory based on past experiences.



# Transformers are Meta-Reinforcement Learners

---

## Related Work

- **RNN Limitations:** Struggle with long-term dependencies, making it difficult to represent information across extended sequences.
- **Latent Variable Models:** Require careful design and may not generalize well to complex tasks.
- **Transformer-Based Architectures:** Designed to handle long-range dependencies and better represent tasks without the drawbacks of RNNs or latent variable models.
- **Prior Research in Meta-Learning and Reinforcement Learning:** Integrates memory mechanisms to enhance task generalization and adaptation.
- **Key Methods:**
  - **RL<sup>2</sup>:** Utilizes RNNs to encode task information.
  - **PEARL:** Uses probabilistic latent variables for dynamic task representation.
  - **MAML:** Focuses on fast adaptation by optimizing initial model parameters to minimize updates across tasks.

# Transformers are Meta-Reinforcement Learners

## TrMRL Design

- **TrMRL (Transformers for Meta-Reinforcement Learning)** is designed to leverage the strengths of Transformers for meta-learning, focusing on their ability to model sequential dependencies and extract contextual task representations.
- The key innovation lies in how TrMRL constructs **episodic memory**:
  - **Episodic Memory**: TrMRL uses the Transformer to store and retrieve relevant task-specific experiences from a rolling memory buffer, enabling the model to maintain a coherent understanding of the task context over time.
  - **Working Memory**: This component dynamically integrates new observations with existing episodic memory, refining the task representation with each interaction.
- The self-attention mechanism plays a central role in TrMRL by:
  - Prioritizing the most relevant portions of the task history for decision-making.
  - Allowing the model to adapt flexibly to new task scenarios by dynamically adjusting the weight assigned to past experiences.
- To address the known training instability of Transformers in reinforcement learning contexts, the study employs the **T-Fixup** initialization strategy, which eliminates the need for complex learning rate scheduling and improves training convergence.
- The model is modular, making it adaptable to a wide range of meta-learning scenarios, including tasks with sparse rewards, high variability, or ambiguous structures.

# Transformers are Meta-Reinforcement Learners

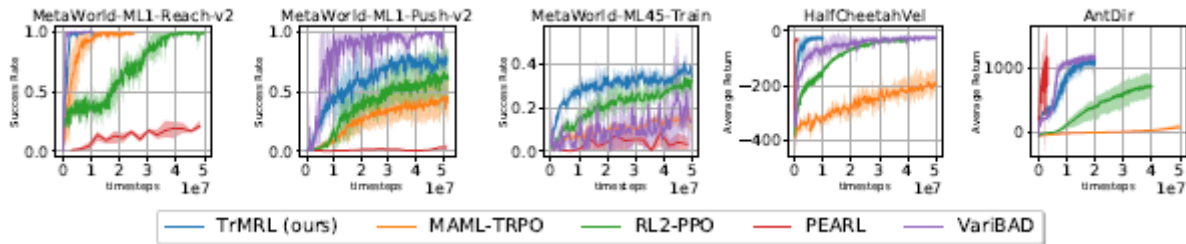
## Experiment Setup

TrMRL was tested in high-dimensional continuous control environments, including **MuJoCo** and **MetaWorld**:

- **MuJoCo**: Evaluated on tasks like **HalfCheetahVel** to measure continuous control performance.
- **MetaWorld**: Assessed for quick adaptation to new tasks.

## Performance Metrics:

- Learning efficiency.
- Adaptation speed.
- Generalization to unseen tasks.



## Key Results:

- Outperformed other methods in low-ambiguity environments (e.g., HalfCheetahVel).
- Showed exceptional efficiency in adapting to unfamiliar tasks.
- Excelled in **Out-of-Distribution (OOD)** tasks through dynamic, context-dependent weighting.

# Transformers are Meta-Reinforcement Learners

---

## Conclusion and Future Directions

- **TrMRL** proved to be an effective solution for meta-reinforcement learning, offering significant improvements in rapid adaptation and generalization across tasks.
- The framework successfully leveraged Transformers to address challenges in meta-RL, such as task ambiguity and the need for dynamic context representation.
- Future work could explore integrating **self-supervised learning tasks**, which would allow the model to learn auxiliary representations and improve sample efficiency further.
- Another potential direction involves testing **TrMRL** in more complex environments with diverse task distributions to push its generalization capabilities further.

## Practical Implications

- **TrMRL's** ability to generalize and adapt quickly makes it a promising tool for real-world applications where tasks may vary significantly or lack prior training data.
- Examples include robotics, where an agent must adapt to new environments or tasks without extensive retraining.
- The episodic memory mechanism provides an advantage in environments with sparse rewards or high uncertainty, enabling better decision-making.
- The model's performance in Out-of-Distribution tasks highlights its potential in scenarios where task data is incomplete or poorly defined during training.



# Transformers are Meta-Reinforcement Learners

---

## Benefits:

### 1-Rapid Adaptation to New Tasks:

Transformers enable task representation based on prior data, making them effective in understanding rapidly changing contexts.

The self-attention mechanism highlights the most relevant parts of the data, improving model performance in reinforcement learning environments.

### 2-Enhanced Generalization Ability:

The dynamic focus on episodic memory allows the model to generalize to new or out-of-distribution (OOD) tasks effectively..

### 3-A Modular Approach to Task Adaptation:

Prompt engineering with Transformers facilitates creating task-specific instructions, reducing the need for retraining the model.

### 4-Reduced Complexity in Setup:

With the stability provided by T-Fixup, models can be trained without complex learning rate schedules or additional configurations.

# Transformers are Meta-Reinforcement Learners

---

## Challenges

### 1-Training Instability:

- Training Transformers in reinforcement learning environments can be unstable, especially when dealing with noisy data or sparse rewards.

### 2-Effectively Designing Prompts:

- Crafting effective prompts requires a deep understanding of the task and the data, making it time-consuming and cognitively demanding.

### 3-Increased Computational Complexity:

- Compared to traditional models like RNN or CRF, prompt-based Transformers require higher computational resources due to the self-attention mechanism and working with long-term memories.

### 4-Difficulty Generalizing in Highly Ambiguous Environments:

- While generalization is a strength, the model may struggle with complex tasks involving unseen or ambiguous contexts.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The overall structure is organic and sprawling, resembling a neural network or a social graph.

# 3.

## Named Entity Recognition with BERT and GPT-3.5

## 3.1 Named Entity Recognition

Named entity recognition (NER) is a field of natural language processing (NLP) that involves the identification and extraction of a variety of named entities from text. These entities are specific objects, people, places, organizations, and other entities that are referred to by proper nouns.

Why is this important?

NER has many real-world applications. For example, in the field of information extraction, NER can be used to extract important information from large amounts of unstructured text, such as identifying the names of people, organizations, and locations mentioned in news articles or social media posts.

NER can help with better:

- Summary.
- Classification.
- Search.
- Recommendations.
- business intelligence.
- question answering.
- sentiment analysis.
- machine translation.

In a letter to **Harvard University** dated **Tuesday** and posted on **the Education Department** website, officials cited the recent **Justice Department** case and asked the school to disclose records of gifts or contacts involving the governments of **China**, **Qatar**, **Russia**, **Saudi Arabia**, and **Iran**. It also requested records regarding telecommunications giants **Huawei Technologies** and **ZTE Corp.** of **China**, the Kaspersky Lab and **Skolkovo Foundation** of **Russia**, and **the Alawi Foundation** or **Iran**, among others.

**The Education Department** said **Yale University** had failed to disclose **at least \$2/2 million** in foreign funding after filing no reports from **2014-17**, according to a document viewed by the **Journal**. The department, also in a letter **Tuesday** to the university, sought records regarding contributions from **Saudi Arabia**, **China**, and its telecom giants **Huawei Technologies**, **ZTE Corp.**, **the National University of Singapore**, **Qatar**, and others. It also asked

the university to detail foreign funding of **Yale Law School's** **Paul Tsai**, **China Center**, and the new **Yale Jackson School of Global Affairs**. If the schools refuse to disclose the information, **the Education Department** can refer the matter to **the Justice Department**, which could pursue civil or criminal actions.

The push for funding disclosure is being driven by concerns about foreign efforts to exploit **American** academia.

**Trump** administration officials and a bipartisan group of allies in **Congress** fear **China** and other foreign rivals are seeking to use donations or collaborative research to gain access to scientific knowledge that would allow them to achieve national strategic goals and narrow their economic or

## 3.2 Named Entity Recognition Methodology

### Tokenization:

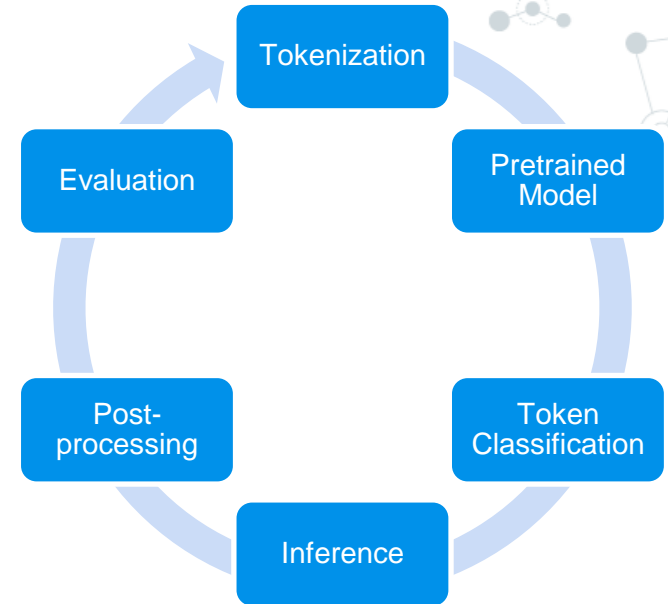
In the initial step of named entity recognition, the input text undergoes tokenization. It basically refers to the process of breaking down a sentence into its constituent parts, we could use [sent\\_tokenize](#) from [nltk](#).

### Pretrained Model:

NER models often rely on powerful deep learning architectures like [BERT \(Bidirectional Encoder Representations from Transformers\)](#). These models are pre-trained on extensive datasets to grasp the contextualized meanings and relationships between words. By leveraging this pre-training, the models gain a deep understanding of language and its nuances.

### Token Classification:

During the training phase, the pre-trained model is [fine-tuned](#) using labeled data containing text sequences and their corresponding entity labels. The model learns to classify each token in the input sequence into specific entity categories, or it may assign a special label for tokens that do not represent entities.



## 3.2 Named Entity Recognition Methodology

### Inference:

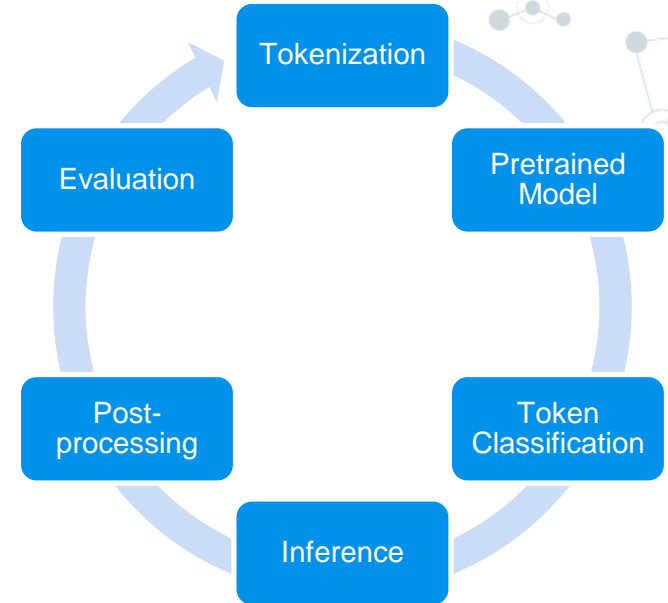
Once the NER model is trained, it can make predictions on new, unseen text. To make predictions, the input text is tokenized and the model processes the tokenized sequence.

### Post-processing:

To obtain the final named entities, post-processing steps are applied to the predicted labels. These steps involve refining the predictions by filtering out unwanted labels, resolving overlapping entities, addressing ambiguities, and applying language-specific rules.

### Evaluation:

Metrics like **precision**, **recall**, and **F1 score** are commonly used to measure the accuracy and completeness of the NER system in identifying the correct entities.



## 3.3 Named Entity Recognition Using BERT

To accomplish this, we'll leverage the power of [BERT \(Bidirectional Encoder Representations from Transformers\)](#), one of the first transformer language models developed by Google AI. BERT has been pretrained on a massive amount of text data and has demonstrated impressive performance in various NLP tasks, including NER.

By fine-tuning the pretrained BERT model on the [wikiann](#) dataset, which contains labeled examples of named entities, we can train our model to recognize and classify different types of named entities, such as persons, locations, organizations, and more.

[WikiANN](#) (sometimes called PAN-X) is a multilingual named entity recognition dataset consisting of Wikipedia articles annotated with LOC (location), PER (person), and ORG (organisation) tags in the IOB2 format. This version corresponds to the balanced train, dev, and test splits of Rahimi et al. (2019), which supports 176 of the 282 languages from the original WikiANN corpus.

Other datasets

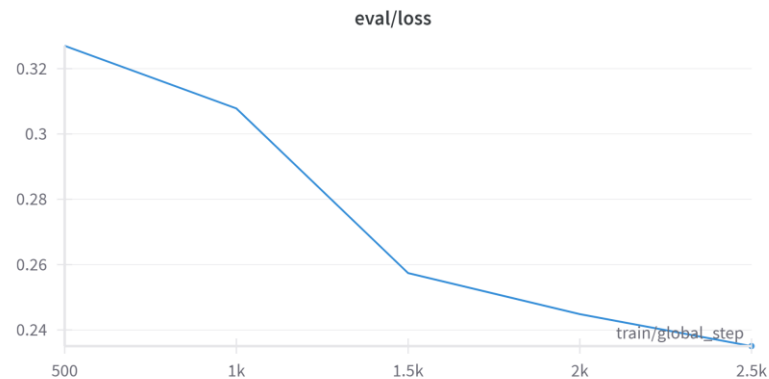
[CONLL](#) NER Datasets (CoNLL-2002, CoNLL-2003), used for benchmark datasets for NER in specific languages. Collecting from news articles and manually entities annotated to ensuring high-quality labels for Person, Organization, Location.

While the CoNLL datasets are high-quality benchmarks for a small set of languages, WikiANN provides a much broader multilingual scope with some quality trade-offs. WikiANN builds upon the CoNLL format and expands its usability to multilingual and low-resource contexts, making it a complementary resource in the NER field.

1	Albert	Albert	NNP	B-PER
2	Einstein	Einstein	NNP	I-PER
3	was	be	VBD	O
4	born	bear	VBN	O
5	in	in	IN	O
6	Ulm	Ulm	NNP	B-LOC

## 3.3 Named Entity Recognition Using BERT

### Model Training



### Input:

"Albert Einstein was born in Ulm, Germany."

### output:

[CLS] -> I-ORG , Albert -> I-PER , Einstein -> B-ORG , was -> I-ORG , born -> I-ORG , in -> I-ORG  
U -> O , ##lm -> B-LOC , -> I-ORG , Germany -> O , . -> I-ORG , [SEP] -> I-ORG



## 3.3 openAI (GPT-3.5 prompt Engineering)

GPT-3.5 is a large language model designed to perform a variety of tasks, including NER, without specific training for a task. Unlike traditional models such as BERT, GPT-3.5 uses few-shot or zero-shot learning to achieve NER through prompts.

**Input Text:** Provide GPT-3.5 with unstructured text from which entities need to be extracted.

Example: "Albert Einstein was born in Ulm, Germany."

**Prompt Engineering:** Design a natural language prompt that explicitly instructs the model to extract named entities and classify them. For example: **prompt** (Extract the named entities from the text below. For each entity, specify its type (e.g., PERSON, LOCATION, ORGANIZATION).)

**Text** ("Albert Einstein was born in Ulm, Germany.")

**API Interaction:** Use the OpenAI GPT API to process the prompt. The API generates responses based on the input text and the task described in the prompt.

**Output:** GPT-3.5 generates a structured output containing the identified entities and their types. Example Response:

```
[
  {"entity": "Albert Einstein", "type": "PERSON"},
  {"entity": "Ulm", "type": "LOCATION"},
  {"entity": "Germany", "type": "LOCATION"}
]
```

**Post-Processing:** The output can be parsed and used for downstream applications such as data analysis, visualization, or integration into a database.

## 3.4 BERT and openAI (GPT-3.5 prompt Engineering)

### Advantages of Using GPT-3.5 for NER:

- **Few-Shot Learning:** Requires minimal training data compared to traditional models like BERT.
- **Flexibility:** GPT-3.5 can adapt to various NER tasks with tailored prompts.
- **Ease of Use:** Eliminates the need for complex pipeline setup or training phases.

### Comparison with BERT for NER:

- **Training:** BERT is a transformer model that requires pre-training and fine-tuning on specific NER datasets. GPT-3.5 does not need this fine-tuning due to its prompt-driven architecture.
- **Performance:** BERT often performs better in domain-specific NER tasks after fine-tuning. GPT-3.5 excels in general-purpose tasks and scenarios requiring minimal effort to implement.
- **Cost:** GPT-3.5 can be costlier due to API calls, while BERT requires computational resources for training and inference.

### BERT:

- Often used for token-level classification tasks like NER.
- Fine-tuning BERT on datasets like CoNLL-2003 achieves state-of-the-art (SOTA) results.
- **F1-score for NER** : ~92-93% with fine-tuning.
- Its bidirectional nature allows it to effectively identify entities by understanding both left and right contexts.

### OpenAI GPT:

- Not primarily designed for token-level tasks. While it can be prompted for NER tasks, its performance is typically inferior to BERT for structured tasks.
- Few-shot learning can be used to approximate NER results, but these may lack consistency and precision compared to BERT's fine-tuning.
- **F1-score:** Significantly lower (~70-80%) compared to fine-tuned models like BERT.

**Conclusion:** BERT is superior for NER due to its bidirectional contextual understanding and ability to fine-tune for specific tasks.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The diagram is partially cut off by the top and left edges of the slide.

4.

## **On the Opportunities and Risks of Foundation Models**



# Exploring and Summarizing the Capabilities of Foundational Models

## On the Opportunities and Risks of Foundation Models

Rishi Bommasani<sup>1</sup> Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora  
Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill  
Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji  
Annie Chen Kathleen Creel Jared Quincy Davis Dorothy Demisley Chris Donahue  
Moussa Doumbouya Ekin Dumus Stefano Ermon John Etchemendy Kewin Eshwaran  
Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman  
Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt  
Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Sashil Jain  
Dan Jurafsky Pratyusha Kalluri Siddharth Karancheti Geoff Keeling Fereschte Khani  
Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi  
Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent  
Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning  
Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avnika Narayan  
Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Niforoshan  
Julian Nyarko Ginty Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech  
Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren  
Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dora Sadigh  
Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin  
Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu  
Jiajun Wu Yuhua Wu Sang Michael Xie Michihito Yasunaga Jiaxuan You Matei Zaharia  
Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou  
Percy Liang<sup>1\*</sup>

Center for Research on Foundation Models (CRFM)  
Stanford Institute for Human-Centered Artificial Intelligence (HAI)  
Stanford University

*AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical research on foundation models will require deep interdisciplinary collaboration commensurate with their fundamentally sociotechnical nature.*


<sup>1</sup>Corresponding author: [pliang@cs.stanford.edu](mailto:pliang@cs.stanford.edu)

<sup>\*</sup>Equal contribution.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The overall structure is organic and sprawling, resembling a neural network or a social graph.

# Exploring and Summarizing the Capabilities of Foundational Models

- **Paper contents:**

- Language
  - Vision
  - Robotics
  - Reasoning and search
  - Interaction
  - Philosophy of understanding
- 
- A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It consists of a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The overall structure is organic and sprawling, resembling a neural network or a social graph.

# CAPABILITIES

## Language:

- This section emphasizes the linguistic capabilities of foundational models. Examples include their ability to perform tasks such as translation and summarization. These tasks demonstrate the versatility of the models in processing and generating coherent text across various contexts and domains, making them valuable for applications in communication, documentation, and global connectivity.

### World Languages



# CAPABILITIES

---

## Language:

### Nature of Human Language

- **Language as the Core of Human Communication:** It is used for interaction, thinking, forming social and emotional relationships, defining identity, and preserving knowledge.
- **Characteristics of Language:**
  - Remarkably diverse yet interconnected in its structure and richness.
  - Complex and efficient; children learn it quickly, and it adapts to the needs of linguistic communities.
- **The Importance of Language in Artificial Intelligence:**
  - Understanding and generating language are fundamental in the field of Natural Language Processing (NLP).
  - Language processing has evolved to become a central focus of "Foundation Models".
- **The Impact of Foundation Models:**
  - They have shown flexibility and high capacity in handling a variety of linguistic situations.
  - They have moved the NLP field towards the goal of learning general language.

# CAPABILITIES

---

## Language:

### The Impact of Foundation Models on NLP

- **Their General and Adaptive Capabilities:**
  - Used to perform a variety of linguistic tasks with high flexibility.
- **Shift in Methodologies:**
  - Transition from specialized architectural systems to customizable foundation models.
- **Outstanding Performance:**
  - Outperformed older systems in most tasks like classification, translation, and text summarization.
- **Role of Models in Language Generation:**
  - Generative models have become a key tool for analyzing and understanding language, leading to improvements in tasks like summarization and dialogue generation.



# CAPABILITIES

---

## Language:

### Language Diversity and Multilingualism

- **Limitations of Foundation Models:**
  - Struggles with handling large linguistic diversity and variations within a single language.
- **Multilingualism:**
  - Multilingual models like mBERT and XLM-R have been trained on about 100 languages, enabling knowledge transfer between high-resource and low-resource languages.
  - Performance is better for languages similar to those with high data resources in training.
- **Future Challenges:**
  - Developing models that can fairly and effectively represent linguistic differences.

# CAPABILITIES

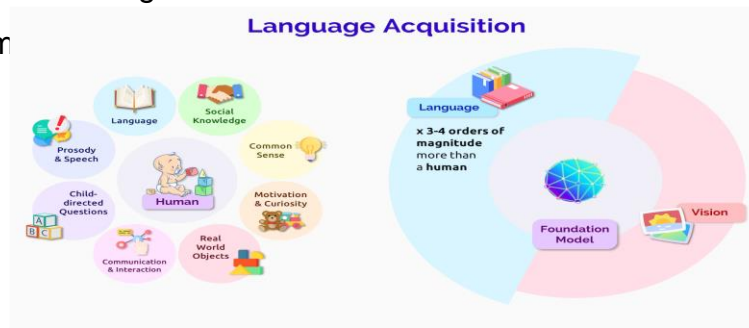
## Language:

### Inspiration from Human Language Acquisition

- **Gaps Between Humans and Models:**
- Humans learn language more efficiently and in ways integrated with the real world, while models rely on text disconnected from the world.
- **Adaptability:**
- Humans develop a general linguistic system, while models remain largely static after training.
- **Future Research Opportunities:**
- Simulating human language acquisition, developing more interactive and adaptive models.

### General Notes

- Foundation models have radically changed research in Natural Language Processing.
- Challenges include fair representation of linguistic diversity, simulating human adaptation, and achieving a deeper understanding of these models.



# CAPABILITIES

---

## **Vision:**

- The vision-related capabilities of foundational models are highlighted with examples such as image recognition and caption generation. These tasks showcase the models' ability to interpret and describe visual content.
- The integration of vision with other modalities enhances their applicability in areas like multimedia analysis, accessibility technologies, and augmented reality.

## **The Goal of Core Models in Vision:**

- Transform raw sensory information into visual knowledge using self-supervised learning, which supports traditional tasks in perception.
- Enhance skills like common temporal and logical reasoning.

## **Importance of Vision:**

- Vision is a fundamental way for living organisms to understand their surroundings.
- Significant challenges exist in enabling machines to acquire human-like vision skills.

## **Potential Applications of Computer Vision:**

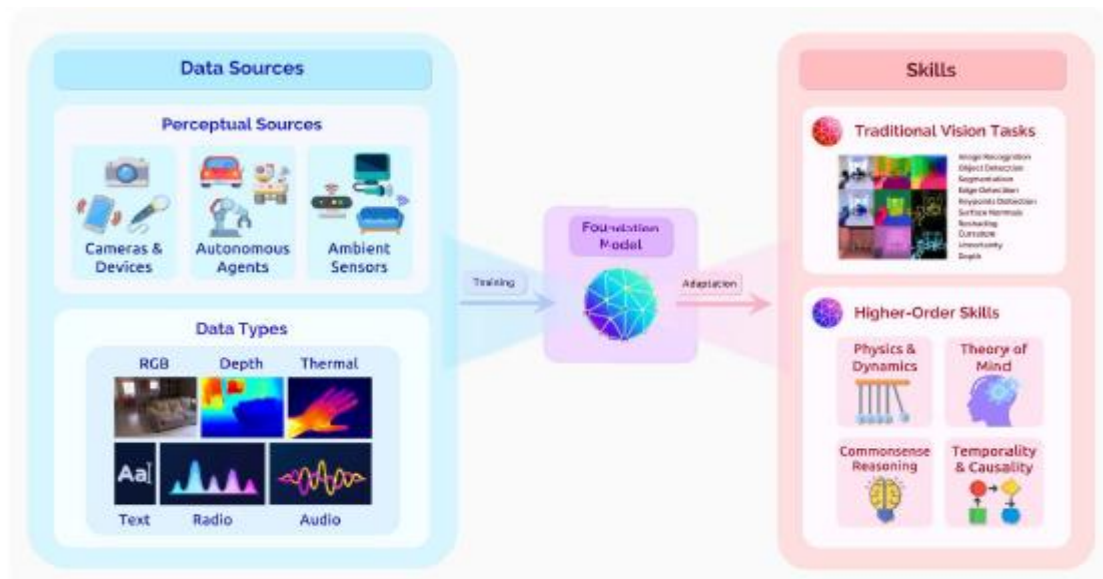
- Autonomous vehicles.
- Life-saving medical tools.
- Next-generation tools for multimedia editing

# CAPABILITIES

## Vision:

### Recent Developments:

- Transition from traditional models that rely on full supervision to self-supervised learning.
- Improved ability to generalize and reduced reliance on manually labeled data.



# CAPABILITIES

---

## **Vision:**

### **Core Potential and Approaches**

#### **Core Computer Vision Tasks:**

- Understanding visual scenes (image classification, object detection, image segmentation).
- 3D tasks like depth estimation and motion representation.
- Multimodal integration with language (e.g., visual question answering).

#### **Current Models:**

- Rely on pre-training using large datasets (such as ImageNet).
- Core models leverage self-supervised learning, reducing the need for manual annotations.

#### **Achievements:**

- Techniques like GANs and self-supervised representations have delivered competitive performance compared to traditional learning.
- Examples like DALL-E and CLIP highlight the potential of multimodal models.

#### **Challenges:**

- Difficulty in achieving logical and social understanding in current models.
- The need to develop architectures and approaches that incorporate more multimodal inputs.

# CAPABILITIES

---

## **Vision:**

### **Key Research Challenges**

### **Potential Impact Areas:**

**Environmental Intelligence:** Enhancing smart interaction in medical and home environments.

**Consumer and Mobile Applications:** Improving interaction between text and images.

**Interactive Agents:** Supporting interactive perception in robots.

### **Open Issues:**

- Achieving strong perception requires external context beyond the current data.
- There is a need to combine efficient designs, large-scale training, and adaptive techniques

# CAPABILITIES

---

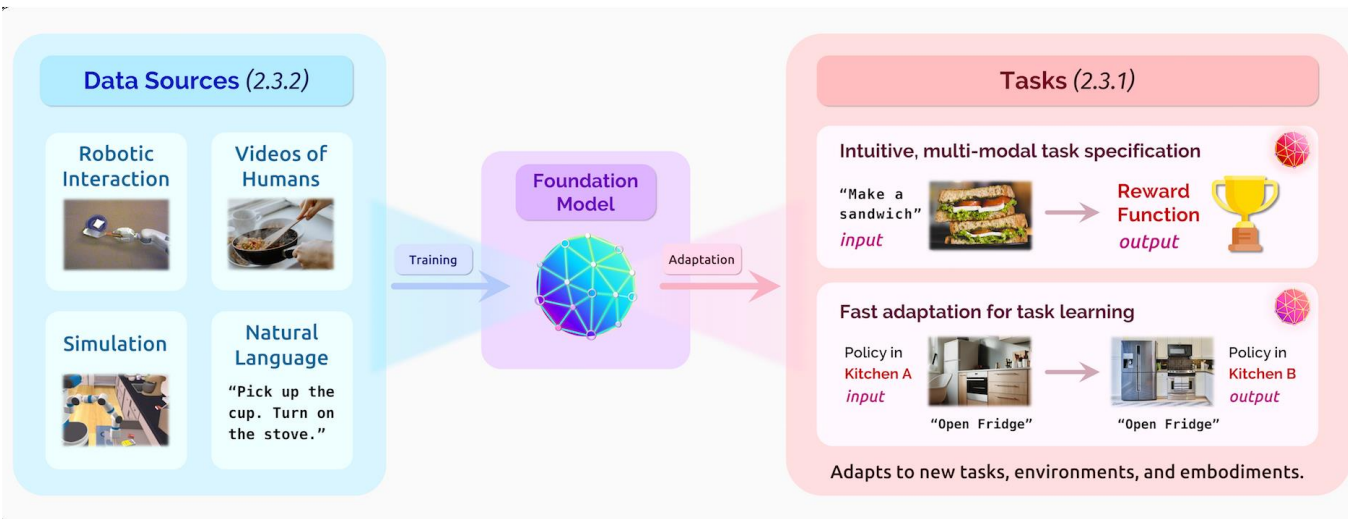
## Robotics:

- In this section, the authors explore the potential of foundation models in robotics, focusing on building robots capable of performing tasks in real-world environments, such as cooking in a new kitchen. The idea is to develop new models that leverage diverse datasets, including robotic interaction, videos of human tasks, and natural language descriptions. The aim is to create robots that can adapt to different settings, handle complex decision-making, and complete tasks like humans do. Foundation models can help tackle high-dimensional, closed-loop decision-making problems, where robot actions directly influence its perception, creating a continuous feedback loop that affects future actions..

# CAPABILITIES

## Robotics:

- Key challenges include acquiring diverse data sets for training, especially as robotic data is harder to collect and more varied than visual or language data. Safety and robustness in new environments are also critical concerns, as deploying robots in the real world could lead to damage if not properly trained. Despite these challenges, the development of robotic foundation models could greatly enhance the capabilities of robots across fields like manufacturing, autonomous driving, and personal assistance.





# CAPABILITIES

---

## Robotics:

### Opportunities

- The authors highlight several opportunities for foundation models in robotics, focusing on task specification and learning across various tasks, environments, and robot embodiments. Robots need to understand the tasks they are assigned, which requires effective task specification models. This involves transforming human descriptions of tasks into quantifiable metrics, such as reward functions, that guide robot learning and actions. These models need to handle different types of task descriptions (e.g., natural language, videos, physical feedback) and generalize across various environments.
- For task learning, foundation models could learn from large, diverse datasets to improve efficiency and reliability. They could model dynamics, policies, and reward functions that enable robots to handle new tasks. A promising direction is to train models on a variety of sensor data (e.g., RGB, depth cameras, microphones) to predict sensory observations or relationships between different data streams. This could lead to more sample-efficient learning, where robots can adapt to new tasks with fewer demonstrations.

# CAPABILITIES

---

## Robotics:

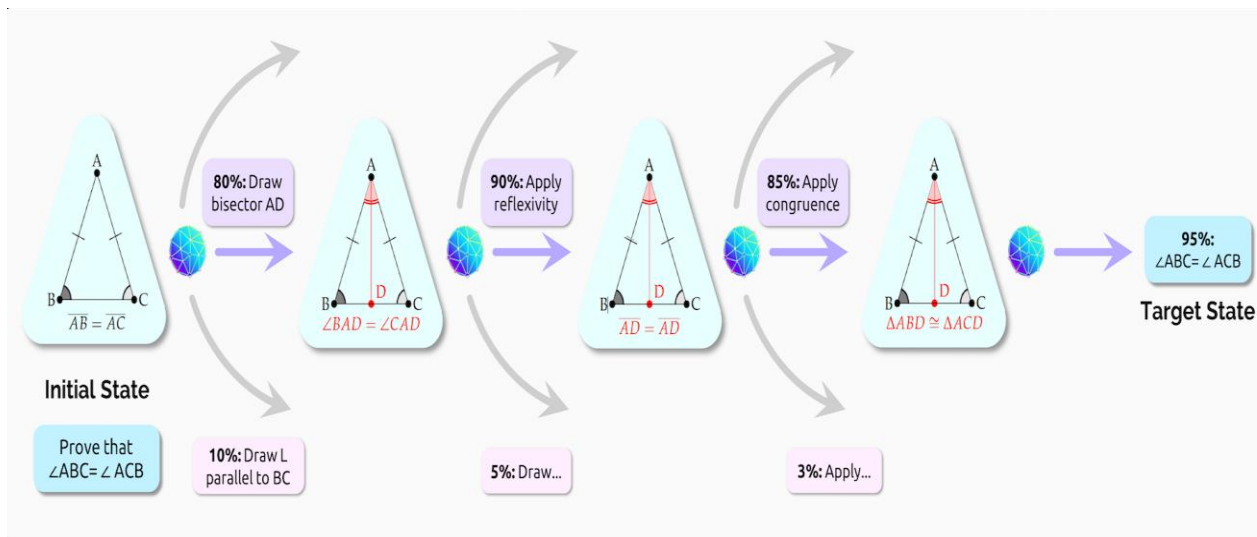
### Challenges and Risks

- One of the key challenges is data collection. Unlike language and vision, robotic datasets are not as abundant or diverse. Collecting large-scale, diverse datasets that reflect a wide range of tasks, environments, and robot embodiments is a critical hurdle. Methods like teleoperation, kinesthetic teaching, and autonomous learning show promise but are still far from achieving the scale needed for general-purpose robotic models.
- Another significant risk is ensuring that the learned models can be safely deployed in real-world environments. Since robots interact with the physical world, there's a need for mechanisms that ensure the robots don't cause harm or damage while performing tasks. The authors emphasize the importance of safety and robustness in these systems.
- In conclusion, while there is great potential for foundation models to advance robotics, overcoming these challenges in data acquisition and safety is crucial for achieving practical, general-purpose robots.

# CAPABILITIES

## Reasoning and search:

- This section discusses how reasoning and search have been central in AI development, addressing problems involving unbounded search spaces that require effective methods to find solutions. While early AI relied on symbolic methods, recent advancements with neural networks, including large language models (LLMs), have shown promising results in tackling complex reasoning tasks. The section explores various reasoning tasks, the role of foundation models, and future challenges in this area.



# CAPABILITIES

## Reasoning and search:

### What are the current tasks?

- Current tasks involve a wide range of problems that require reasoning in unbounded search spaces.

### Examples of tasks:

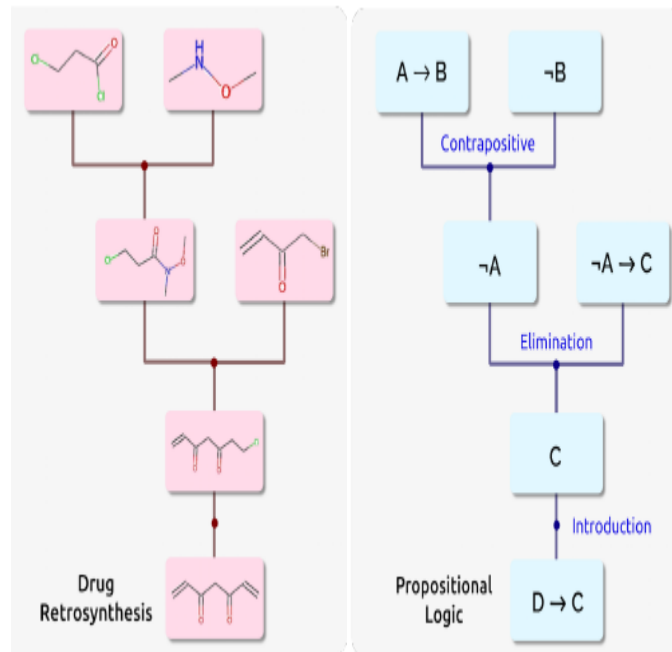
#### Mathematical theorem proving:

- Like proving the equality of angles in an isosceles triangle.

#### Real-world problems:

- Such as drug discovery, program design, and optimization challenges.

These tasks require searching among numerous alternatives, which becomes increasingly complex as the search space grows.



# CAPABILITIES

---

## Reasoning and search:

### What's the role of foundation models?

- Foundation models play a crucial role in advancing AI capabilities by providing a flexible, general-purpose framework that can be adapted to various tasks with minimal task-specific training. These models leverage large-scale pre-training on diverse datasets, enabling them to capture a broad range of patterns and knowledge from text, images, audio, and other forms of data.

### Foundation models:

- These models are trained on a vast amount of general data, which makes them capable of learning and generalizing from this information across different tasks.

### General-purpose capabilities:

- Foundation models offer a versatile foundation for various AI applications, from natural language processing (NLP) and computer vision to reinforcement learning and beyond.

# CAPABILITIES

---

## Reasoning and search:

### Efficiency:

- Because foundation models have already learned a lot of the underlying knowledge and representations from pre-training, they can perform new tasks more quickly and efficiently compared to models trained from scratch.

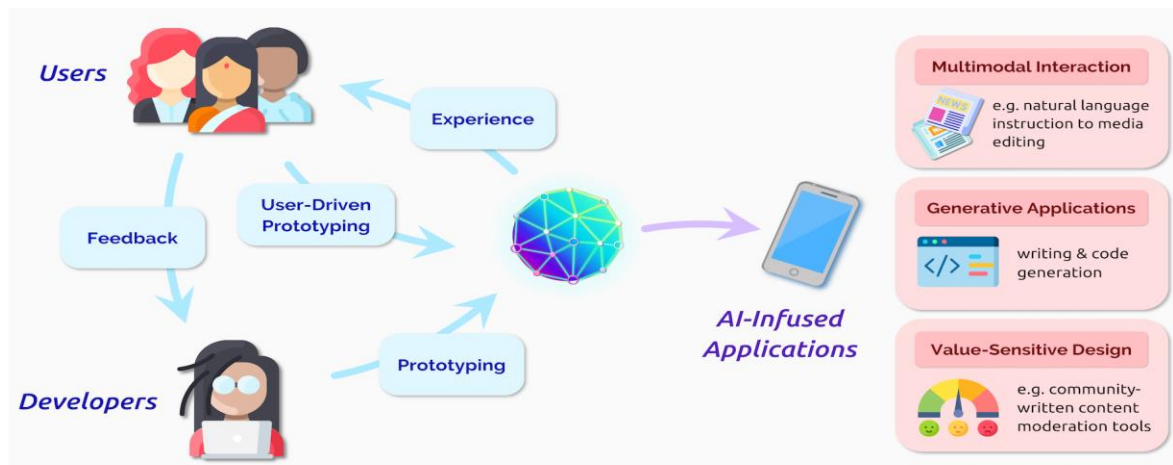
### Task adaptation:

- These models can be fine-tuned to specific tasks or domains with relatively less data and computational resources, making them highly effective for solving a wide range of problems.

# CAPABILITIES

## Interaction:

- This section discusses the impact of foundation models on both developers and end-users, as well as how the boundaries between them may blur. Foundation models, such as GPT-3.5 and DALL-E, are transforming the way AI applications are developed and interacted with, offering developers new opportunities and giving users more power to influence application creation.



# CAPABILITIES

---

## **Interaction:**

### **Impact on AI-infused Application Developers' Process:**

#### **Reduced Development Threshold:**

- Foundation models can significantly lower the complexity involved in building sophisticated AI applications. They allow non-ML experts to prototype powerful applications without large datasets or model training.

#### **Generative and Multi-modal Abilities:**

- These models enable high-quality, diverse interactions through generative and multi-modal capabilities, pushing the limits of what's possible in application development.

#### **Challenges with Unpredictability:**

- Foundation models, though powerful, are more complex and unpredictable compared to task-specific models, making them harder to manage. Ensuring consistent performance, through techniques like fine-tuning and prompt-engineering, is necessary.



# CAPABILITIES

---

## Interaction:

### Impact on End-User Interaction with AI-infused Applications:

#### Augmentation of User Abilities:

- Foundation models enhance user agency and reflect their values. They enable users to interact with AI in ways that augment their capabilities, especially in creative tasks such as writing, art, and music creation.

#### Diverse Forms of Interaction:

- These models empower users to create high-quality multimedia content intuitively, with applications like text-to-image generation and collaborative authoring. They also facilitate dynamic and personalized experiences, such as customized video game interactions or remastering legacy media.

### Impact on End-User Interaction with AI-infused Applications:

#### Potential Risks and Biases:

- Foundation models can also lead to unintended consequences, such as reflecting biases or exposing inappropriate content. Ensuring that these models align with users' values is essential to prevent harm.

#### Ethical Considerations:

- With the increased use of generative models, ethical questions arise about trust, ownership, and the impact on work, culture, and language.

# CAPABILITIES

---

## **Interaction:**

### **Blurring the Line Between Developers and End-Users:**

### **User Involvement in Development:**

Foundation models lower the threshold for users to participate in the development process. This could allow end-users to co-create AI models that reflect their values and needs, such as content moderation tools for specific communities.

### **Customization and Adaptation:**

Users could directly influence the behavior of foundation models, allowing for tailored experiences based on specific preferences or needs. However, challenges like bias mitigation and ensuring robust behavior even for non-experts remain.

### **Future Opportunities:**

The potential for end-users to play an active role in application development could fundamentally change how we create and interact with AI applications, enabling a more user-centered approach.

# CAPABILITIES

---

## Philosophy of Understanding:

- This section explores whether foundation models, such as those used for natural language processing, can truly "understand" the data they are trained on, particularly language. While these models generate language fluently, they often show incoherence, leading to the skepticism that they are merely "stochastic parrots" (echoing patterns without comprehension). The key issue here is whether these models can reach genuine understanding
- **What is a Foundation Model?**
- Foundation models are self-supervised, meaning they learn abstract co-occurrence patterns from large datasets (text, images, etc.) without explicit understanding of what the symbols mean.
- Self-supervision involves predicting or generating sequences (e.g., filling gaps in sentences) based on the patterns found in the training data, but this doesn't inherently teach the model the meanings of the words.

# CAPABILITIES

---

## Philosophy of Understanding:

- **What's at Stake?**
- **Trust:** We may need models to understand language to trust them. Language can mislead, and understanding may be necessary for trust in language processing systems.
- **Interpretability:** Understanding may be crucial for making models interpretable and predictable, which is important for control and transparency.
- **Accountability:** As AI systems become more integral to decision-making, understanding language might be required to hold models accountable for their outputs.

# CAPABILITIES

---

## Philosophy of Understanding:

- **What is Understanding?**
- **Metaphysics of Understanding:** This addresses what it means for a model to truly understand language. There are different philosophical views.
- **Internalism:** Understanding involves having internal representations or concepts that correspond to linguistic inputs.
- **Referentialism:** Understanding means being able to evaluate the truth of statements relative to the world.
- **Pragmatism:** Understanding is about using language effectively without necessarily needing internal representations.
- **Epistemology of Understanding:** This is concerned with how we determine if a model understands. Since pragmatism focuses on observable behaviors, it might offer a practical approach to testing understanding.

## challenges and Considerations:

- There are concerns about whether foundation models can ever achieve true understanding. If they rely solely on text data, they might fail to refer to the world in a meaningful way.
- However, multimodal data (e.g., combining text with images, audio, or sensor data) could help bridge this gap and enable models to learn more meaningful associations that reflect real-world understanding.

## In conclusion

- "this section of the paper has provided valuable insights into the remarkable capabilities of foundational models across a range of domains, including language, vision, robotics, and reasoning. Through examining tasks such as translation, image recognition, and autonomous decision-making, we have highlighted how these models drive advancements in automation and enhance human-machine collaboration. This exploration underscores the transformative potential of foundational models, offering exciting opportunities for future technological innovations and their wide-ranging applications across various industries."

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. Some nodes are highlighted with blue circles, and others with blue dots. The lines are thin and gray, creating a mesh-like structure.

# Discussion

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a network of nodes and lines, with some nodes highlighted by blue circles and others by blue dots.

A decorative network diagram consisting of a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are solid blue, some are solid grey, and some are hollow with a blue outline. The lines connecting the nodes are thin and grey, creating a dense, organic structure that fills the corners of the slide. The overall aesthetic is clean and modern, with a focus on connectivity and network structure.

**Thank you ..**