



RECENT PROGRESS
LIMITATIONS
OPPORTUNITIES
CHALLENGES

NATURAL LANGUAGE PROCESSING

Trondheim Machine Learning Meetup
2021-09-30
Norconsult Informasjonssystemer
v/ Thomas H. Thoresen, Gruppeleder ML/AI, Avd. Fundator

AGENDA

- 1700-1730 Welcome - 
- 1730-1815 NLP – Status and recent progress 
- 1815-1830 Break - 
- 1830-1915 Demo and testing of applications - 
- 1915-1945 Outlook and challenges - 
- 1945-2000 Wrap-up – Discussions  and questions 

ABOUT US

[This is a short article about the awesome machine learning group in Norconsult:]

Feel free to get in touch if you are interested in positions or joining us.

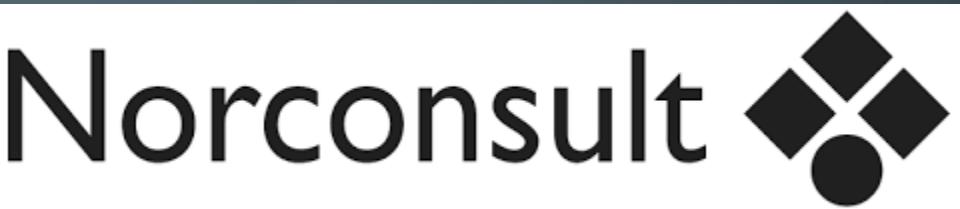
Norconsult has, for 20 years, been building advanced, industrial grade software solutions. From real-time control to decision support to operational analytics. 15 years ago, we started developing machine learning solutions.

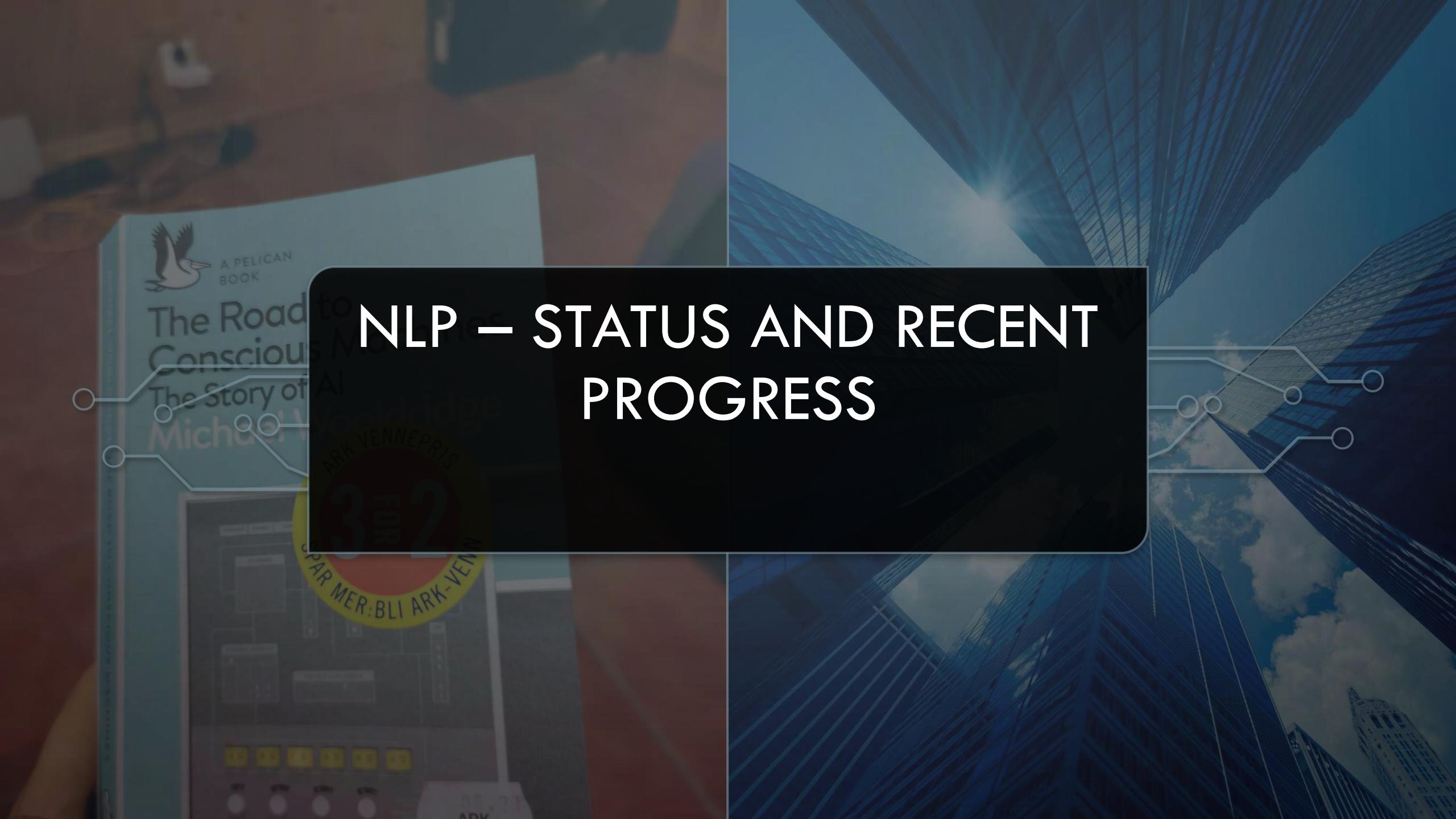
Today, we are one of the market leaders. Norconsult is the Norwegian company with the largest portfolio of delivering AI and Machine Learning solutions. We approach AI and ML as a core competency.

We solve business problems with advanced AI and Machine Learning solutions. We have built solutions for oil & gas, petrochemical, marine, mining, finance, banking and insurance, public sector, and more.

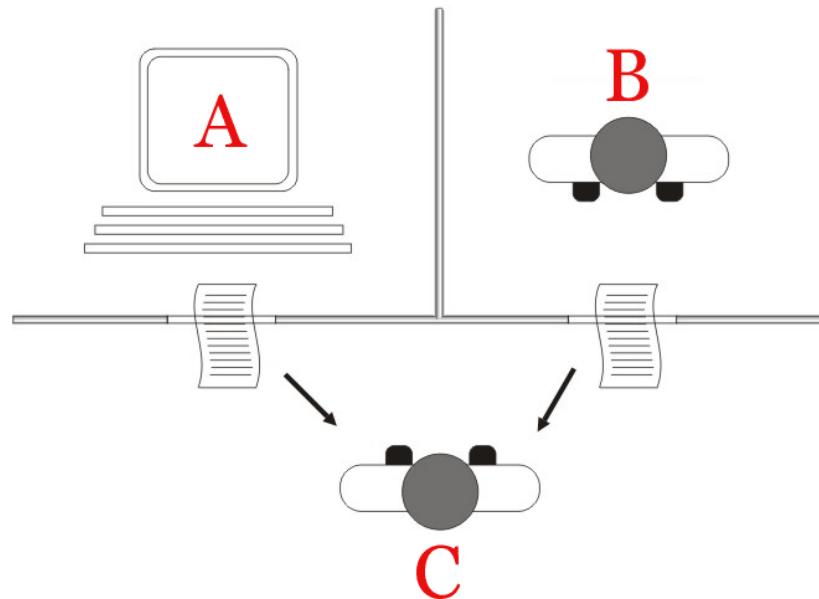
We leverage our AI and Machine Learning solutions in our solutions to help our customers become more competitive. Our solutions are in use by some of the largest companies in the world. We use industry proven open source technologies. We are 15 years in the game and have expertise in Machine Learning in our own solutions.

We are helping our customers become more competitive and efficient.





NLP – STATUS AND RECENT PROGRESS



THE TURING TEST (1950)

```
Z:\>mount c .
Drive C is mounted as local directory ./
```

```
Z:\>c:
```

```
C:\>ELIZA.EXE
```

Hi! I'm Eliza. I am your personal therapy computer.
Please tell me your problem.

I am nervous to speak at a meetup on NLP in Trondheim Machine Learning Meetup today.

Did you come to me because you are nervous to speak at a meetup on NLP in Trondheim Machine Learning Meetup today?

Yes!

Are you Sure?

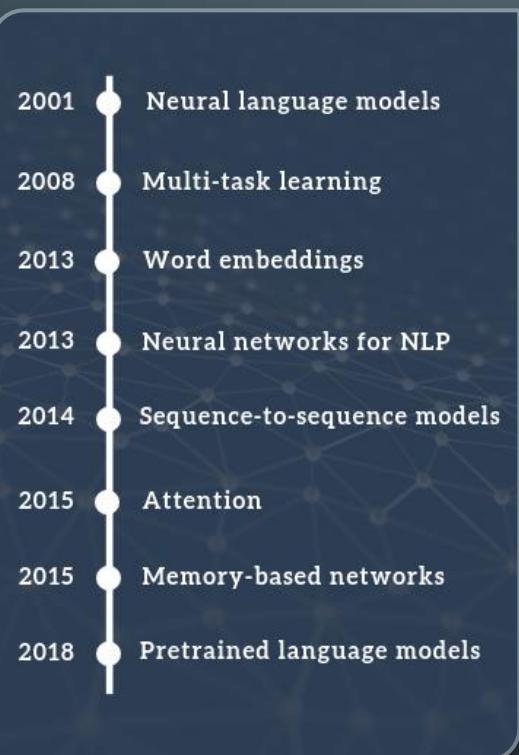
Absolutely.

Say, do you have any psychological problems?

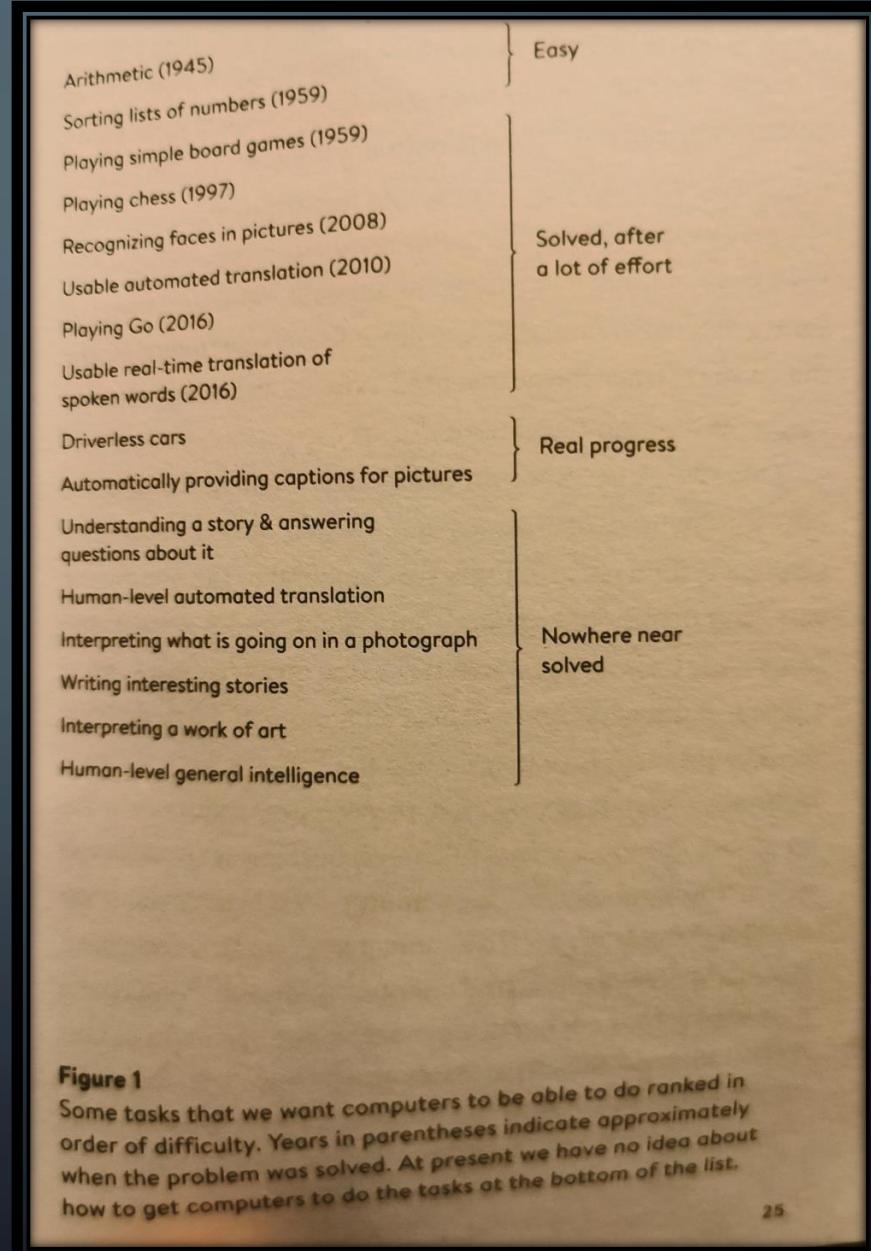
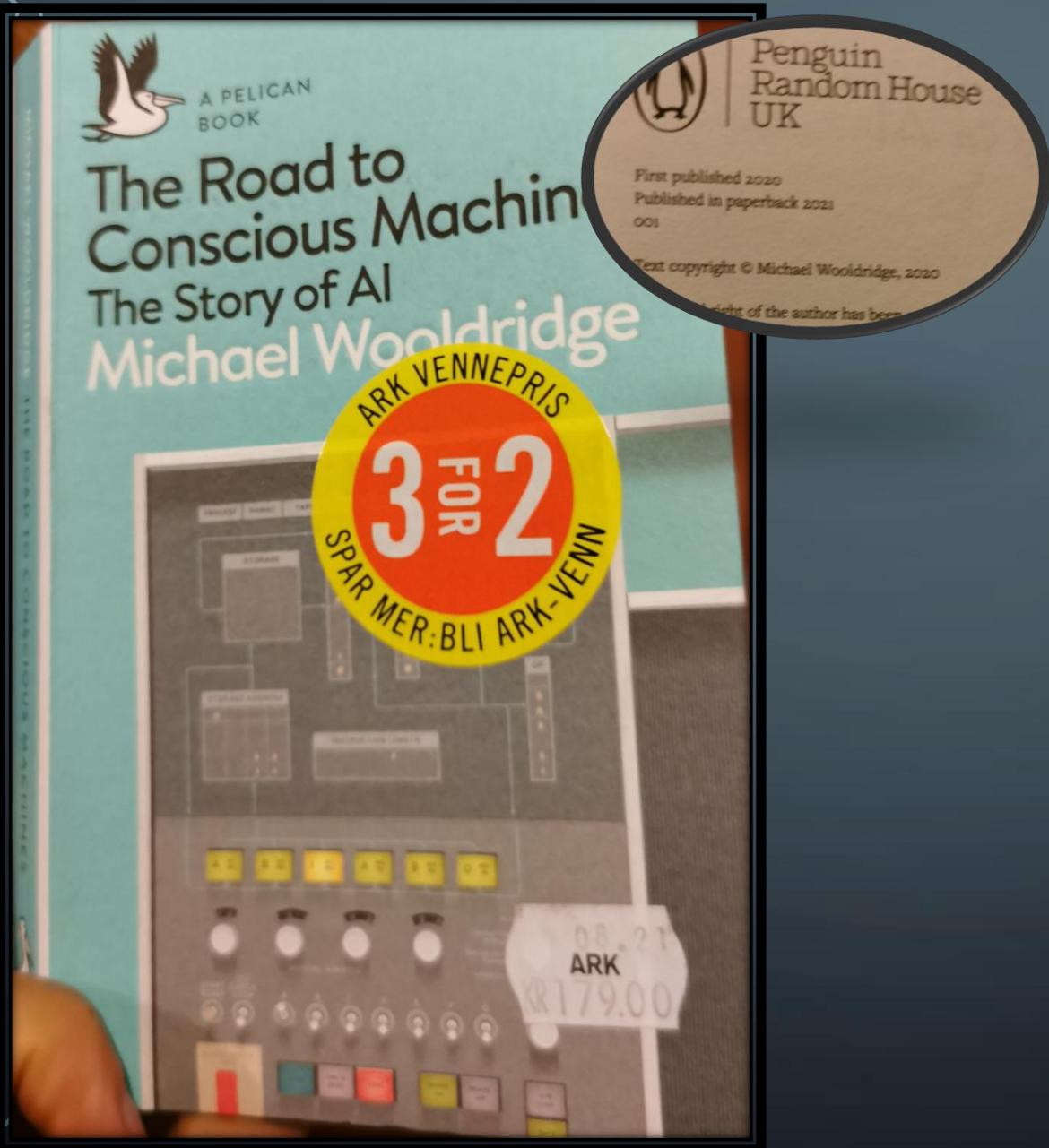
You could say that.

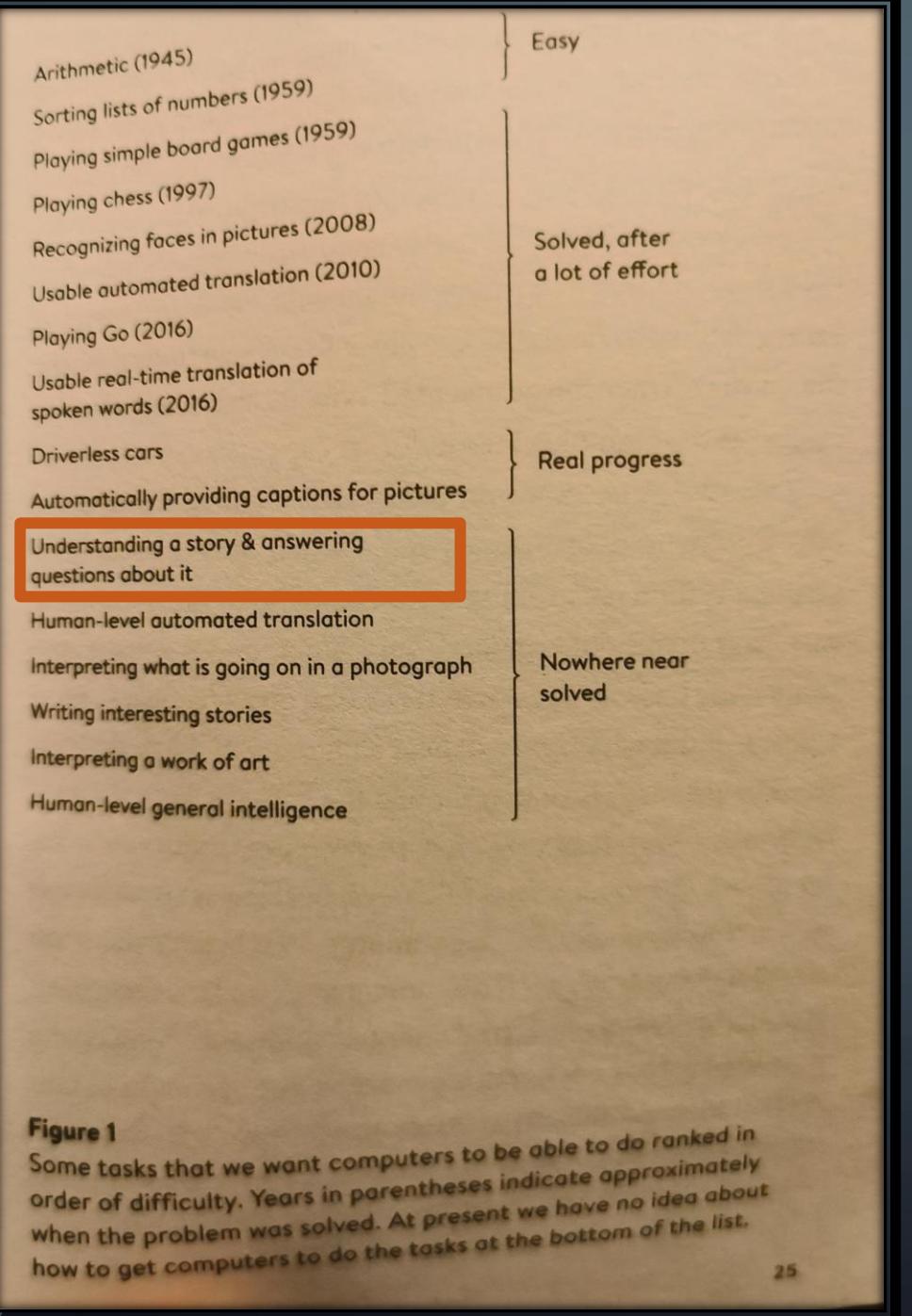
We were discussing you --not me.

HOW DID WE GET HERE?



- <https://ruder.io/a-review-of-the-recent-history-of-nlp/>





UNDERSTANDING A STORY AND ANSWERING QUESTIONS ABOUT IT



UNDERSTANDING A STORY AND ANSWERING QUESTIONS ABOUT IT

```
from transformers import pipeline

nlp = pipeline("question-answering")

context = r"""
Title: Recent progress in NLP - Status, limitations and opportunities
Time: Thu, Sep 30 · 5:00 PM CEST
During the last few years, a lot has happened in the field of Natural Language Processing (NLP).

Transformer models, such as BERT, T5, GPT-3 and BigBird has gained a lot of attention. !
We take a closer look at what these models are capable of. What are the limitations, opportunities and challenges? 🤖

We aim to give a practical guide to some of the ecosystem of models and datasets that will allow you to use these models on your own.
We also take a specific look at the state of NLP for Norwegian language. 🇳🇴

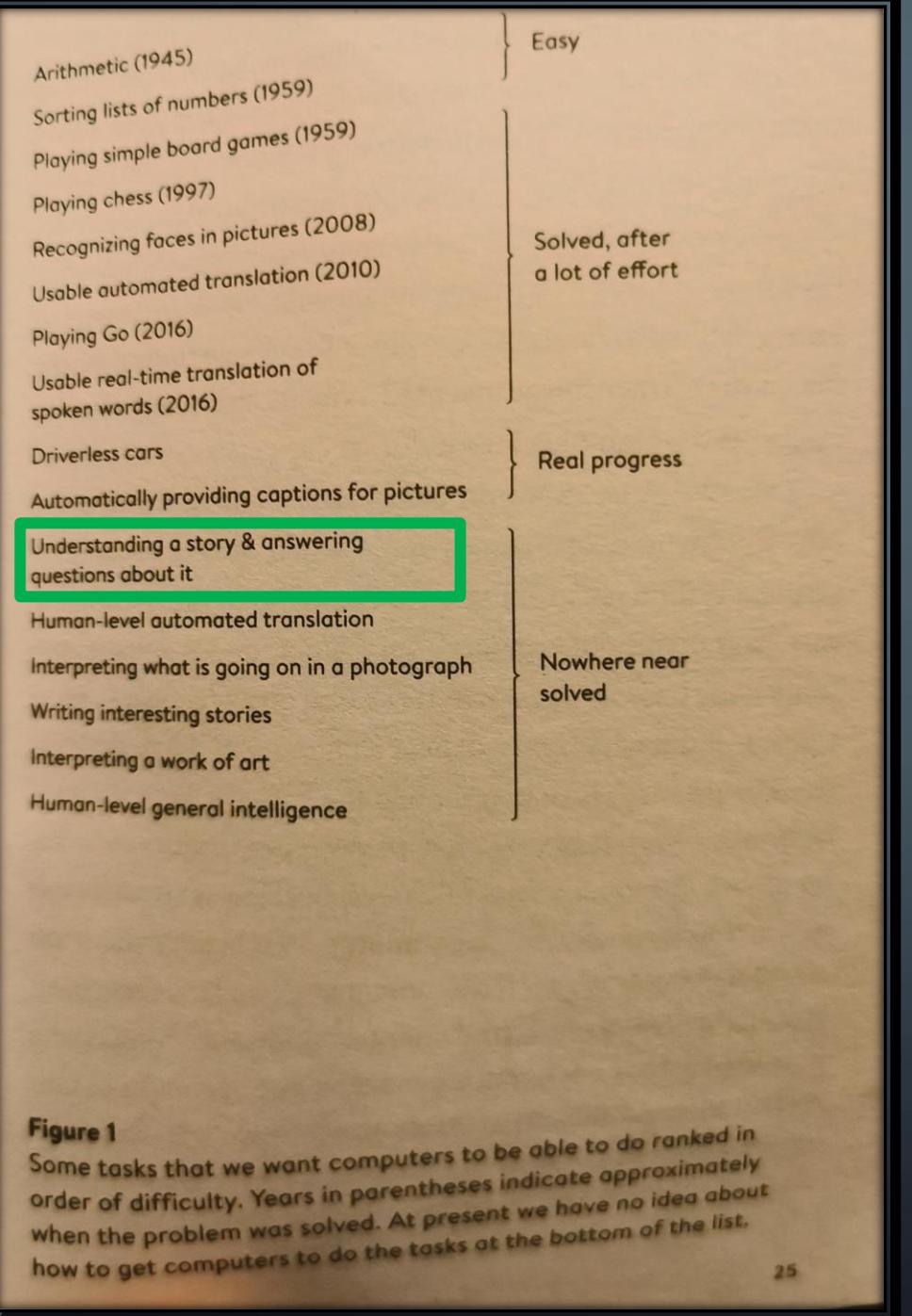
Last, but not least we bring up some ethical issues and potential harmful use of recent NLP technology. 🚨

We will provide pizza, drinks and snacks. 🍕🍺

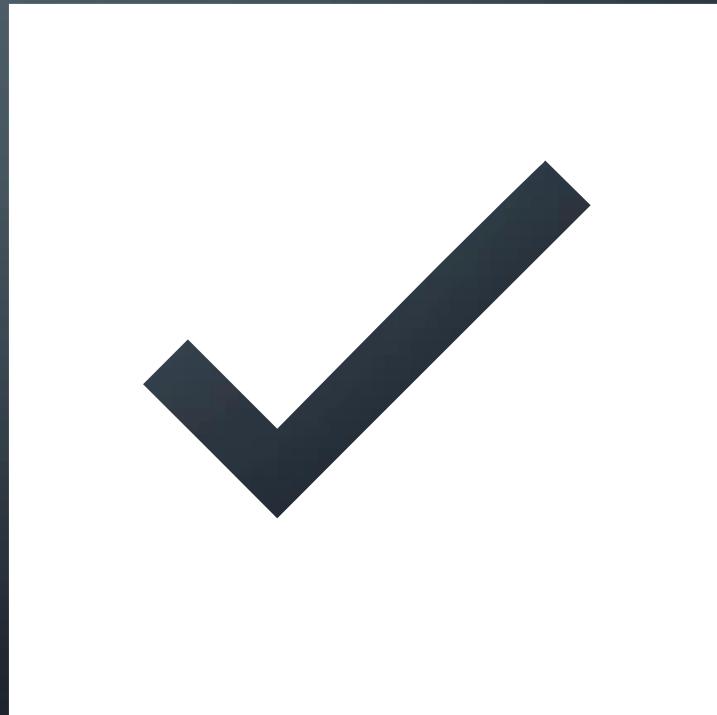
Join us for an interesting evening. First come, first served. ➡️
"""

print(nlp(question="What time is the NLP meetup?", context=context))
print(nlp(question="Which Transformer models will be discussed?", context=context))
print(nlp(question="Will there be food?", context=context))

No model was supplied, defaulted to distilbert-base-cased-distilled-squad (https://huggingface.co/distilbert-base-cased-distilled-squad)
/usr/local/lib/python3.7/dist-packages/transformers/configuration_utils.py:337: UserWarning: Passing `gradient_checkpointing` to a config
  "Passing `gradient_checkpointing` to a config initialization is deprecated and will be removed in v5 "
[{'score': 0.7071709632873535, 'start': 82, 'end': 98, 'answer': 'Sep 30 · 5:00 PM'}
{'score': 0.002769800368696451, 'start': 230, 'end': 245, 'answer': 'BERT, T5, GPT-3'}
{'score': 0.3022589683532715, 'start': 744, 'end': 768, 'answer': 'pizza, drinks and snacks'}
```



UNDERSTANDING A STORY AND ANSWERING QUESTIONS ABOUT IT



Volume LIX, Issue 236
October 1950

Article Contents

1. The Imitation Game

**2. Critique of the New
Problem**

3. The Machines concerned in
the Game

4. Digital Computers

5. Universality of Digital
Computers

6. Contrary Views on the Main
Question

7. Learning Machines

BIBLIOGRAPHY

[Next >](#)

2. Critique of the New Problem

As well as asking, 'What is the answer to this new form of the question', one may ask, 'Is this new question a worthy one to investigate?' This latter question we investigate without further ado, thereby cutting short an infinite regress.

The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man. No engineer or chemist claims to be able to produce a material which is indistinguishable from the human skin. It is possible that at some time this might be done, but even supposing this invention available we should feel there was little point in trying to make a 'thinking machine' more human by dressing it up in such artificial flesh. The form in which we have set the problem reflects this fact in the condition which prevents the interrogator from seeing or touching the other competitors, or hearing their voices. Some other advantages of the proposed criterion may be shown up by specimen questions and answers. Thus:

Q: Please write me a sonnet on the subject of the Forth Bridge.

A : Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764

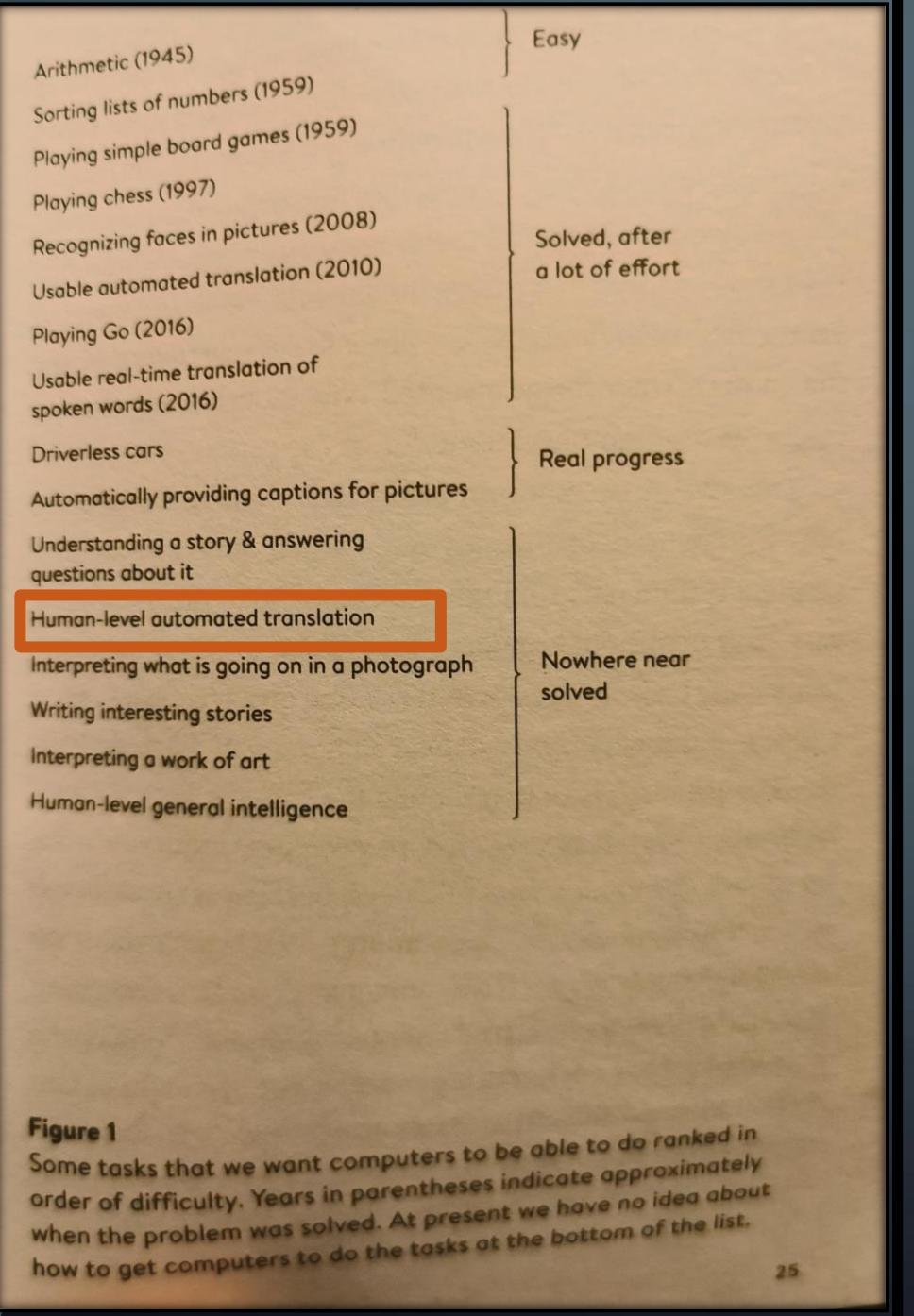
A : (Pause about 30 seconds and then give as answer) 105621.

Q: Do you play chess?

A : Yes.

Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A : (After a pause of 15 seconds) R-R8 mate.



HUMAN-LEVEL AUTOMATED TRANSLATION



HUMAN-LEVEL AUTOMATED TRANSLATION

ARTICLE

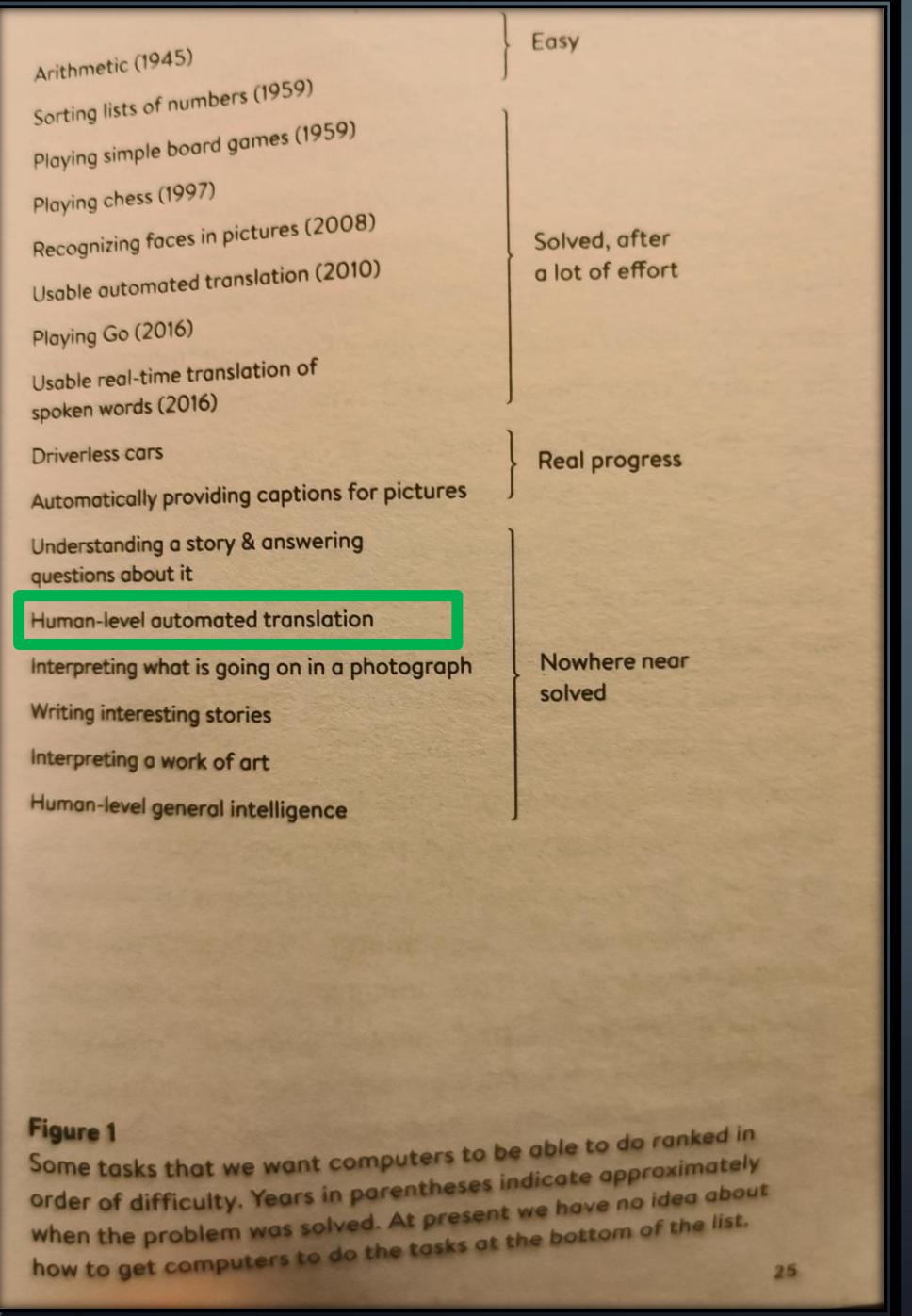
<https://doi.org/10.1038/s41467-020-18073-9> OPEN

 Check for updates

Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals

Martin Popel  ^{1,5✉}, Marketa Tomkova  ^{2,5}, Jakub Tomek  ^{3,5}, Łukasz Kaiser  ⁴, Jakob Uszkoreit  ⁴, Ondřej Bojar  ¹ & Zdeněk Žabokrtský  ¹

The quality of human translation was long thought to be unattainable for computer translation systems. In this study, we present a deep-learning system, CUBBITT, which challenges this view. In a context-aware blind evaluation by human judges, CUBBITT significantly outperformed professional-agency English-to-Czech news translation in preserving text meaning (translation adequacy). While human translation is still rated as more fluent, CUBBITT is shown to be substantially more fluent than previous state-of-the-art systems. Moreover, most participants of a Translation Turing test struggle to distinguish CUBBITT translations from human translations. This work approaches the quality of human translation and even surpasses it in adequacy in certain circumstances. This suggests that deep learning may have the potential to replace humans in applications where conservation of meaning is the primary aim.



HUMAN-LEVEL AUTOMATED TRANSLATION



Arithmetic (1945)	Easy
Sorting lists of numbers (1959)	
Playing simple board games (1959)	
Playing chess (1997)	Solved, after a lot of effort
Recognizing faces in pictures (2008)	
Usable automated translation (2010)	
Playing Go (2016)	
Usable real-time translation of spoken words (2016)	
Driverless cars	Real progress
Automatically providing captions for pictures	
Understanding a story & answering questions about it	
Human-level automated translation	
Interpreting what is going on in a photograph	Nowhere near solved
Writing interesting stories	
Interpreting a work of art	
Human-level general intelligence	

Figure 1

Some tasks that we want computers to be able to do ranked in order of difficulty. Years in parentheses indicate approximately when the problem was solved. At present we have no idea about how to get computers to do the tasks at the bottom of the list.

INTERPRET WHAT IS GOING ON IN A PHOTOGRAPH



INTERPRET WHAT IS GOING ON IN A PHOTOGRAPH

ARTICLE

<https://doi.org/10.1038/s41467-020-18073-9> OPEN

 Check for updates

Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals

Martin Popel  ^{1,5✉}, Marketa Tomkova  ^{2,5}, Jakub Tomek  ^{3,5}, Łukasz Kaiser  ⁴, Jakob Uszkoreit  ⁴, Ondřej Bojar  ¹ & Zdeněk Žabokrtský  ¹

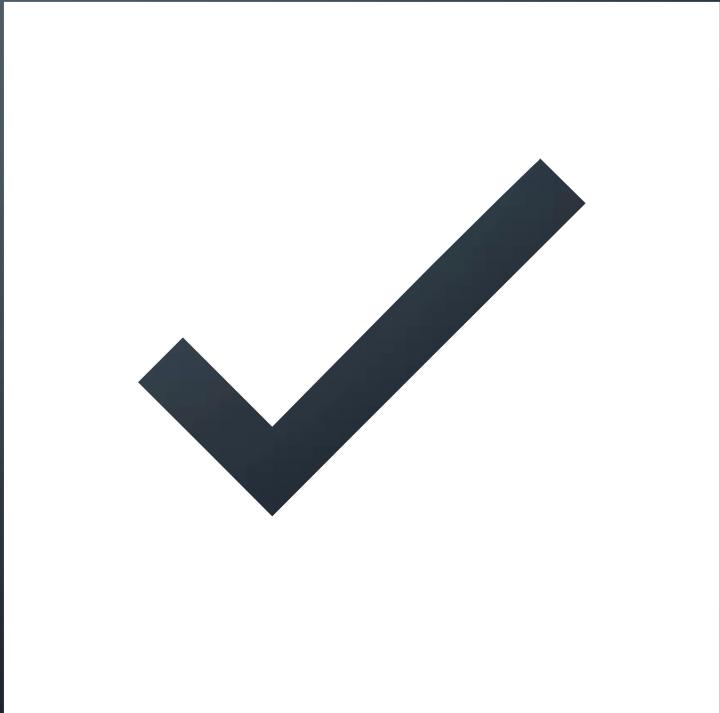
The quality of human translation was long thought to be unattainable for computer translation systems. In this study, we present a deep-learning system, CUBBITT, which challenges this view. In a context-aware blind evaluation by human judges, CUBBITT significantly outperformed professional-agency English-to-Czech news translation in preserving text meaning (translation adequacy). While human translation is still rated as more fluent, CUBBITT is shown to be substantially more fluent than previous state-of-the-art systems. Moreover, most participants of a Translation Turing test struggle to distinguish CUBBITT translations from human translations. This work approaches the quality of human translation and even surpasses it in adequacy in certain circumstances. This suggests that deep learning may have the potential to replace humans in applications where conservation of meaning is the primary aim.

Arithmetic (1945)	Easy
Sorting lists of numbers (1959)	
Playing simple board games (1959)	Solved, after a lot of effort
Playing chess (1997)	
Recognizing faces in pictures (2008)	Real progress
Usable automated translation (2010)	
Playing Go (2016)	Nowhere near solved
Usable real-time translation of spoken words (2016)	
Driverless cars	Real progress
Automatically providing captions for pictures	
Understanding a story & answering questions about it	Nowhere near solved
Human-level automated translation	
Interpreting what is going on in a photograph	Nowhere near solved
Writing interesting stories	
Interpreting a work of art	Nowhere near solved
Human-level general intelligence	

Figure 1

Some tasks that we want computers to be able to do ranked in order of difficulty. Years in parentheses indicate approximately when the problem was solved. At present we have no idea about how to get computers to do the tasks at the bottom of the list.

INTERPRET WHAT IS GOING ON IN A PHOTOGRAPH



Arithmetic (1945)	Easy
Sorting lists of numbers (1959)	
Playing simple board games (1959)	Solved, after a lot of effort
Playing chess (1997)	
Recognizing faces in pictures (2008)	Real progress
Usable automated translation (2010)	
Playing Go (2016)	Nowhere near solved
Usable real-time translation of spoken words (2016)	
Driverless cars	Nowhere near solved
Automatically providing captions for pictures	
Understanding a story & answering questions about it	Nowhere near solved
Human-level automated translation	
Interpreting what is going on in a photograph	Nowhere near solved
Writing interesting stories	
Interpreting a work of art	Nowhere near solved
Human-level general intelligence	

Figure 1

Some tasks that we want computers to be able to do ranked in order of difficulty. Years in parentheses indicate approximately when the problem was solved. At present we have no idea about how to get computers to do the tasks at the bottom of the list.

INTERPRETING A WORK OF ART



INTERPRETING A WORK OF ART

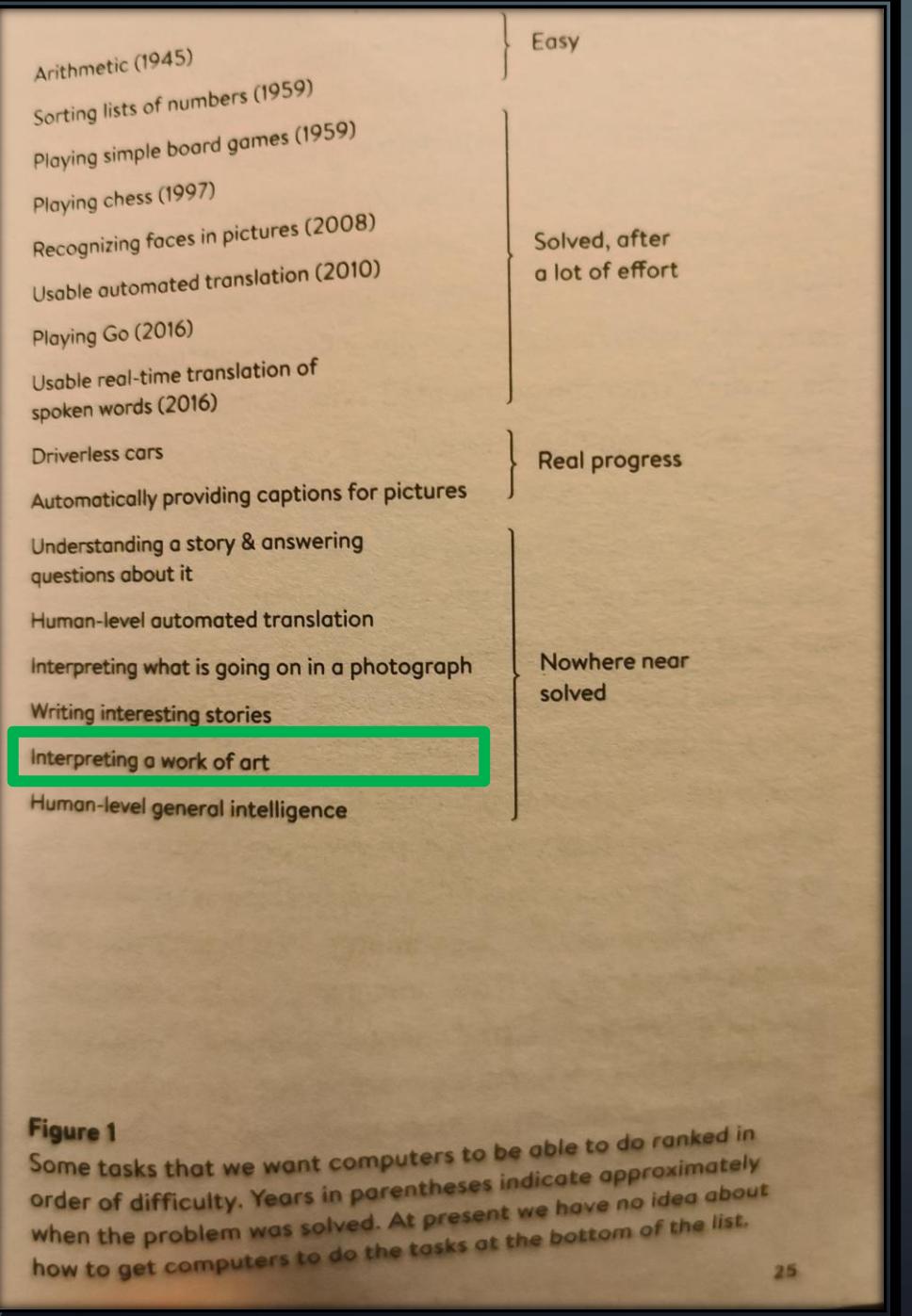


Transformers?

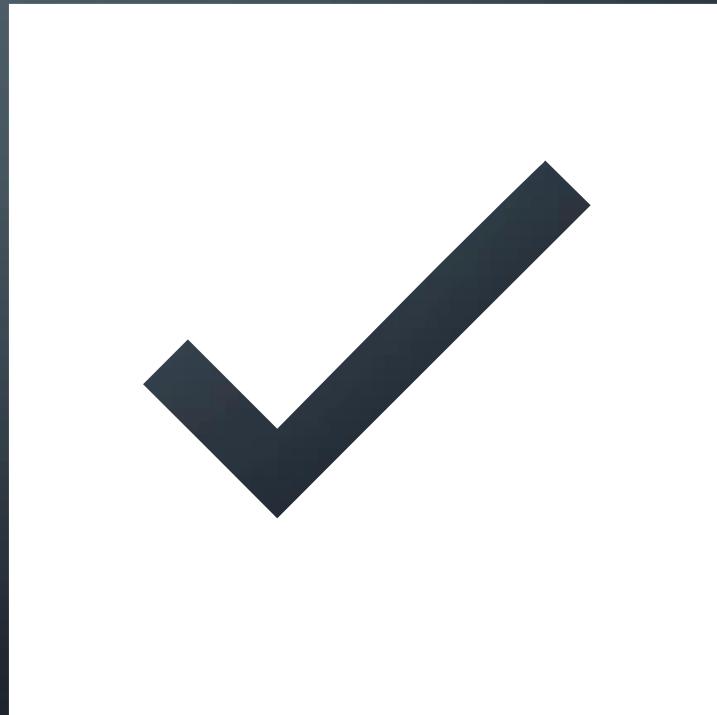
Turing test:

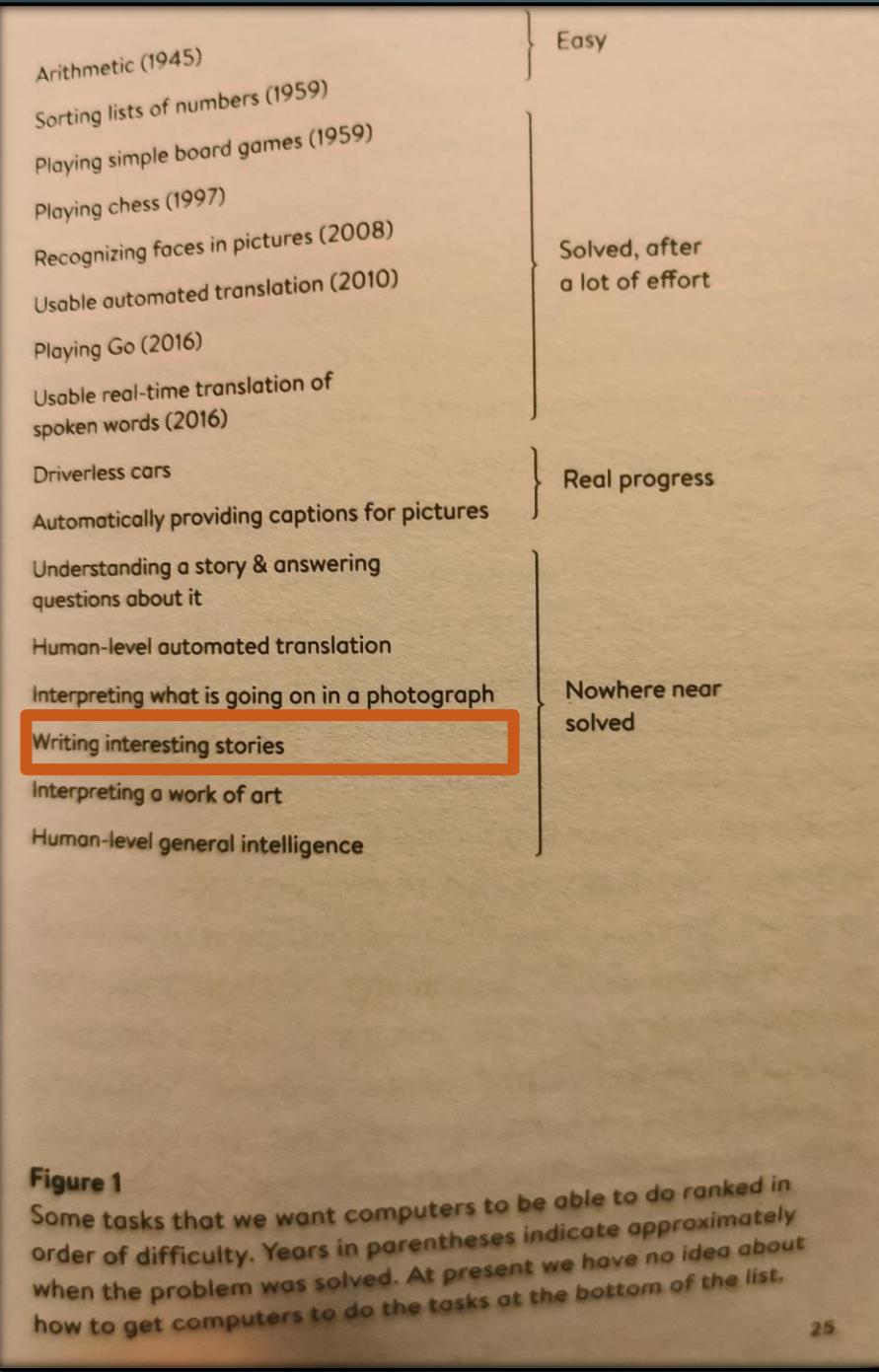
AI on par with human ground truth
50.3% (won 20.6%, tie 29.7%)

Transformers?



INTERPRETING A WORK OF ART





WRITING INTERESTING STORIES



WRITING INTERESTING STORIES

- [Trending blog post on hacker news](#)
- [Pharmako-AI](#)
- <https://www.gwern.net/GPT-3>

GENERATE Use sample scene
New Script ▾

NEW **SAVE** **DOWNLOAD** **SHARE**

THE BOY THAT CREATED AN AI

THOMAS

dialog ▾

i am so curious what will happen when we create agents with human-level intelligence. i have to find out.

Meanwhile, he keeps looking at the monitor. On the monitor, the data processing system that labels dots on a grid is back and forth. Each dot represents the existence of an agent in the universe. Suddenly, a command appears in the transparent shield.

The intelligent calculator moves over the blue points of the screen and a red dot appears in the very top right corner. Now, the calculator moves to the right where a bar with 3 equal lengths appears and a long bar with 6 equal lengths appears and the calculator stops. It types a command: prof. ahmad awwad (21 years old), accelerator: 4.25 <out> accelerator: 4.66 The calculator makes a number between 4.25 and 4.66. The transparent shield dies and Thomas looks at the screen. Two red dots get labeled now. They are the 2nd and the 17th agents. The red dot continues getting smaller in the bottom right corner as if the number on the screen keeps changing.

PROF. AHMAD AWWAD

Hello Thomas, can you recognize me?

THOMAS

Yes, i can recognize you. You are the son of an intelligence donor.

While i was studying your works i came across a few poems that were written by your computer. i liked the poems very much, so i wanted to see your poem.

WRITING INTERESTING STORIES



WRITING INTERESTING STORIES

This is an interesting story about a young student from Trondheim, trying to balance his interests in AI, fitness and philosophy. His name is Thomas. One day, Thomas made a program that makes predictions about the weather based on weather data. He knows that it doesn't understand the whole weather system, but gets it generally right.

As you might have guessed, it doesn't work very well, and goes horribly wrong when a storm hits the city.

Thomas is in a tough spot. He has to explain to his parents that he has been playing around with a program that is not working, and that he has to go to school.

He has to explain to his friends that he is not at home, and that he is in a strange place.

He has to explain to his girlfriend that he is not going to be able to see her for a while.

He has to explain to his professors that he is not going to be able to contribute to the class discussion.

And he has to explain to himself that he made a mistake.

Thomas feels a bit of shame for making an AI that didn't do well at its job. He feels that he has failed his AI, and that he has let his girlfriend down.

But he also feels that he has learned something. He has learned that his AI

Volume LIX, Issue 236
October 1950

Article Contents

1. The Imitation Game

2. Critique of the New Problem

3. The Machines concerned in the Game

4. Digital Computers

5. Universality of Digital Computers

6. Contrary Views on the Main Question

7. Learning Machines

BIBLIOGRAPHY

[Next >](#)

2. Critique of the New Problem

As well as asking, 'What is the answer to this new form of the question', one may ask, 'Is this new question a worthy one to investigate?' This latter question we investigate without further ado, thereby cutting short an infinite regress.

The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man. No engineer or chemist claims to be able to produce a material which is indistinguishable from the human skin. It is possible that at some time this might be done, but even supposing this invention available we should feel there was little point in trying to make a 'thinking machine' more human by dressing it up in such artificial flesh. The form in which we have set the problem reflects this fact in the condition which prevents the interrogator from seeing or touching the other competitors, or hearing their voices. Some other advantages of the proposed criterion may be shown up by specimen questions and answers. Thus:

Q: Please write me a sonnet on the subject of the Forth Bridge.

A : Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764

A : (Pause about 30 seconds and then give as answer) 105621.

Q: Do you play chess?

A : Yes.

Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A : (After a pause of 15 seconds) R-R8 mate.

[“This is a sonnet on the subject of the Forth Bridge:”]

A Gift For Britain's Visitors, A Gift For Scotia's Lovers, A Gift For Its Friends In Every Land.
At the place where London now ends, The Forth Bridge is ended.
So ends the sky, And the lowland; and the Tweed is gone,
The river that takes the rain from ground, So high that, by its strength, every tide Makes the currents in
lochs and moors beat.
Here lands end and sea begins, and all The waters go like weather into the skies.
The bright bridges, the old woods, the sea-sand, All unite at this place.
Here every town May greet its traveller; and here every town May be gone; and, over all, the bridge
By every kind of traffic climbs and rests. So all things here, even hills, are seen To be of the work of
Man's hands, as earth Be of the work of Man's hands.
And look, Here in the heart of things, the ships lie All at anchor, like great tame birds at air; And each,
as it slowly stirs and lifts, As if a breeze called aloud from its nest.
In this port the ships go and come, the coals Go up from deeps and rivers, the shipwrights Bring their
goods to the fair, and the trains Of mighty ships, like flocks of beasts, move About their bridges.
Here is all the world Of land and sea, as one happy creature.
And should the bridge ever be changed, Or the towns drown, should ere long be lost All her light, and
all her gladness, Some one shall whisper her secrets here.
And there, yet above the sea, clear above
The wide waves of the north, where there lives No storm, her white towers shall stand and wave, As if
to guard some ancient place below.
And still, where the towns were, and shall again, Her long shadows shall fall between her walls.
But may it be ere the foul and frantic north Shall fall from its place, and this be all!

Arithmetic (1945)	Easy
Sorting lists of numbers (1959)	
Playing simple board games (1959)	Solved, after a lot of effort
Playing chess (1997)	
Recognizing faces in pictures (2008)	Real progress
Usable automated translation (2010)	
Playing Go (2016)	Real progress
Usable real-time translation of spoken words (2016)	
Driverless cars	Nowhere near solved
Automatically providing captions for pictures	
Understanding a story & answering questions about it	Nowhere near solved
Human-level automated translation	
Interpreting what is going on in a photograph	Nowhere near solved
Writing interesting stories	
Interpreting a work of art	Nowhere near solved
Human-level general intelligence	

Figure 1

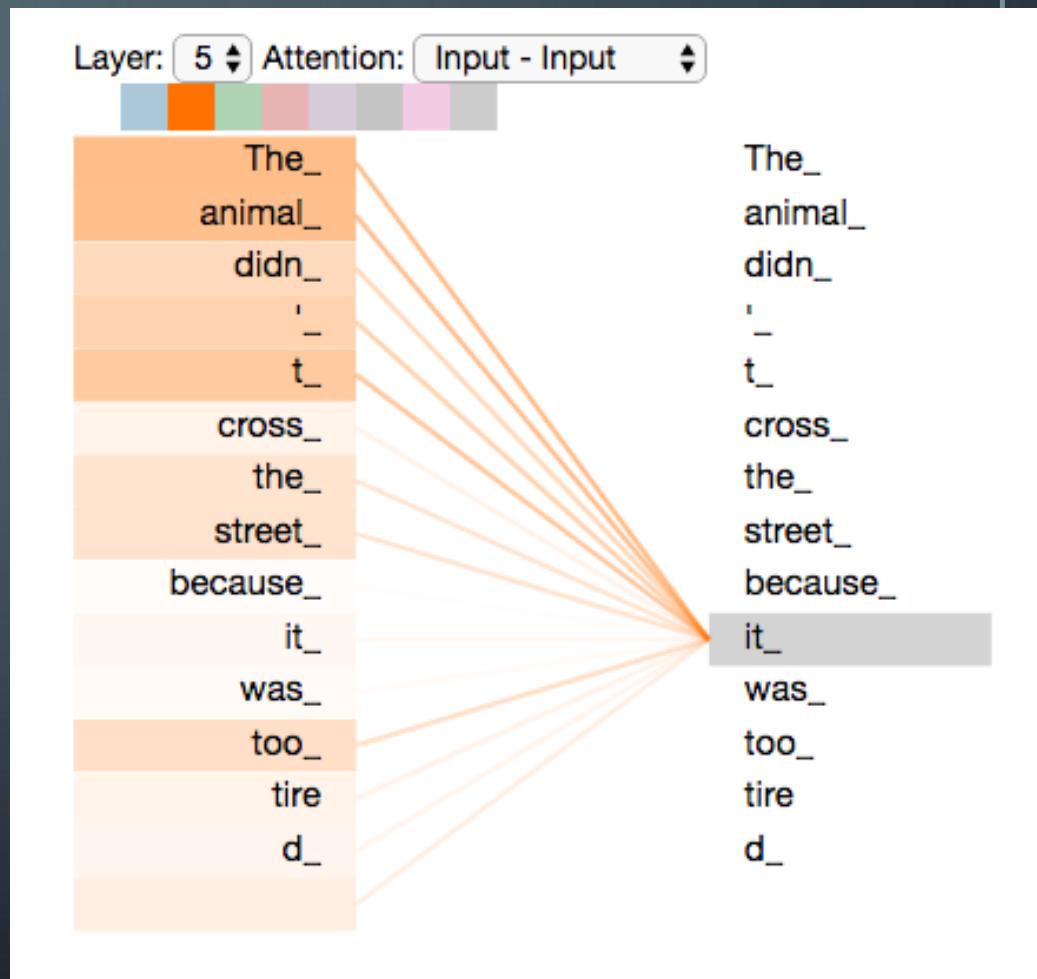
Some tasks that we want computers to be able to do ranked in order of difficulty. Years in parentheses indicate approximately when the problem was solved. At present we have no idea about how to get computers to do the tasks at the bottom of the list.

WRITING INTERESTING STORIES



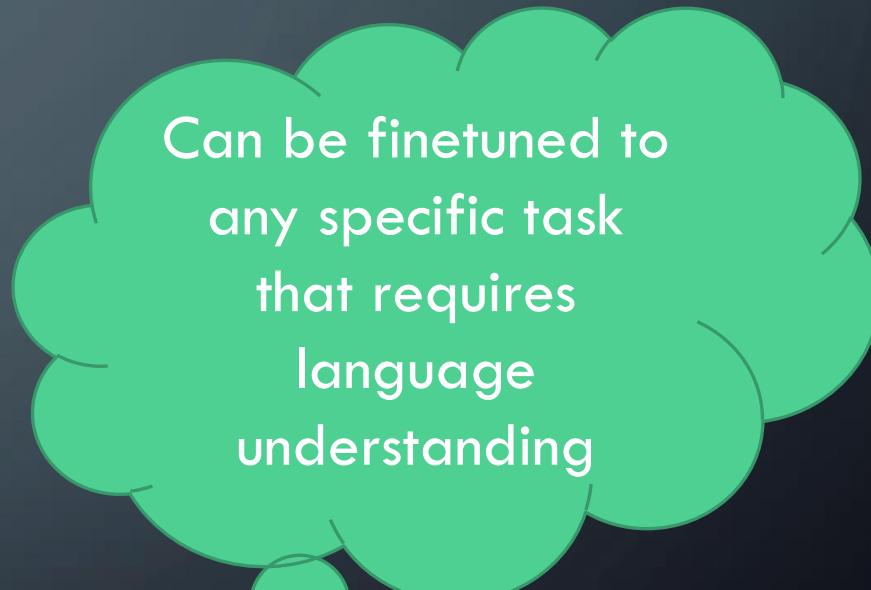
TECHNOLOGY BEHIND MOST OF THESE BREAKTHROUGHS

- **Transformers**
- Attention is all you need (2017)
- Self-attention
- Parallelizable (Scales better than RNN's)
 - <https://jalammar.github.io/illustrated-transformer/>
 - <http://nlp.seas.harvard.edu/2018/04/03/attention.html>
 - <https://huggingface.co/course/>



WHAT CAN THEY DO?

- Common tasks
 - **feature-extraction** (get the vector representation of a text)
 - **fill-mask**
 - **ner** (named entity recognition)
 - **question-answering**
 - **sentiment-analysis**
 - **summarization**
 - **text-generation**
 - **translation**
 - **zero-shot-classification**

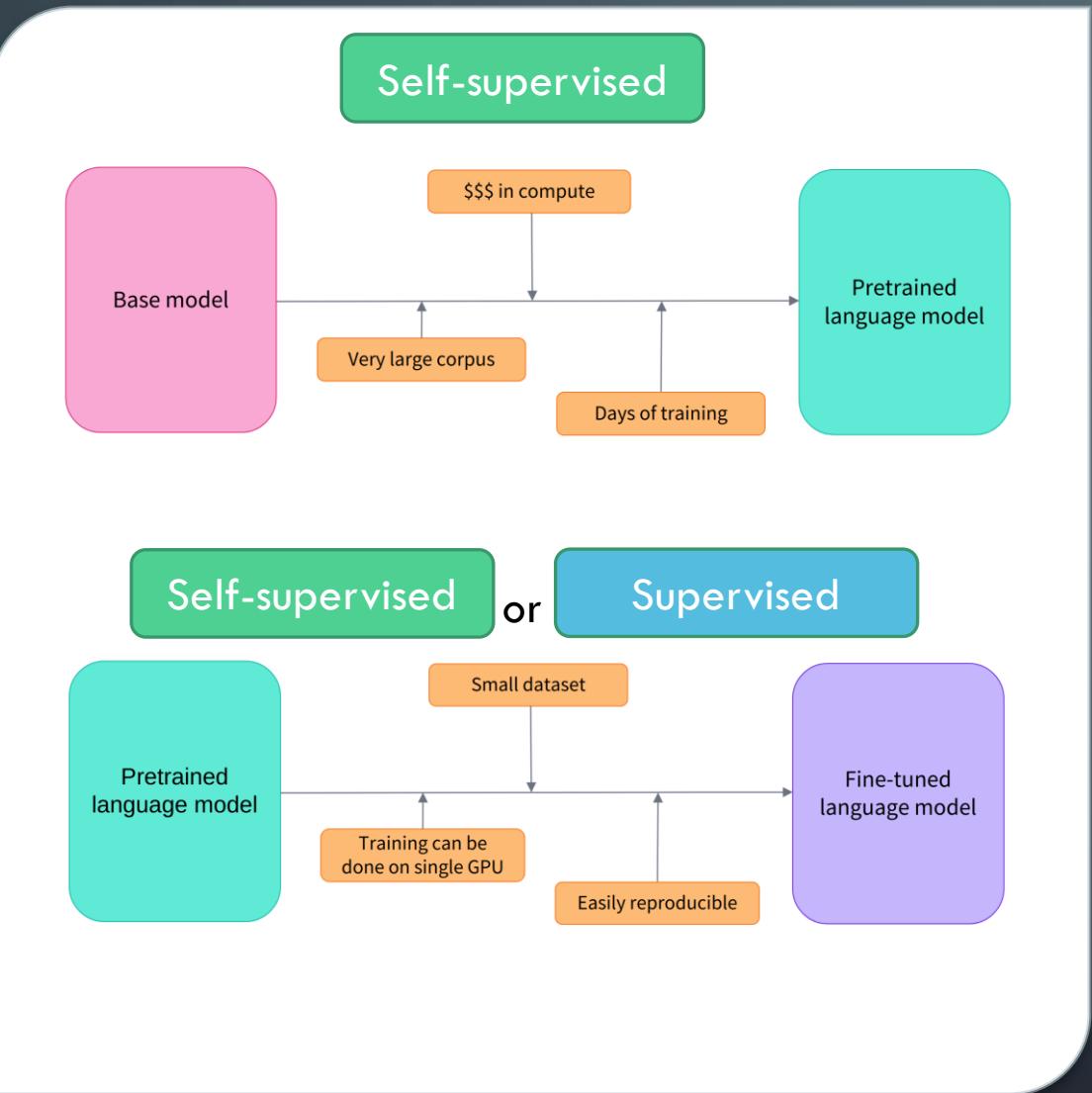


Can be finetuned to
any specific task
that requires
language
understanding

TRANSFORMER CATEGORIES

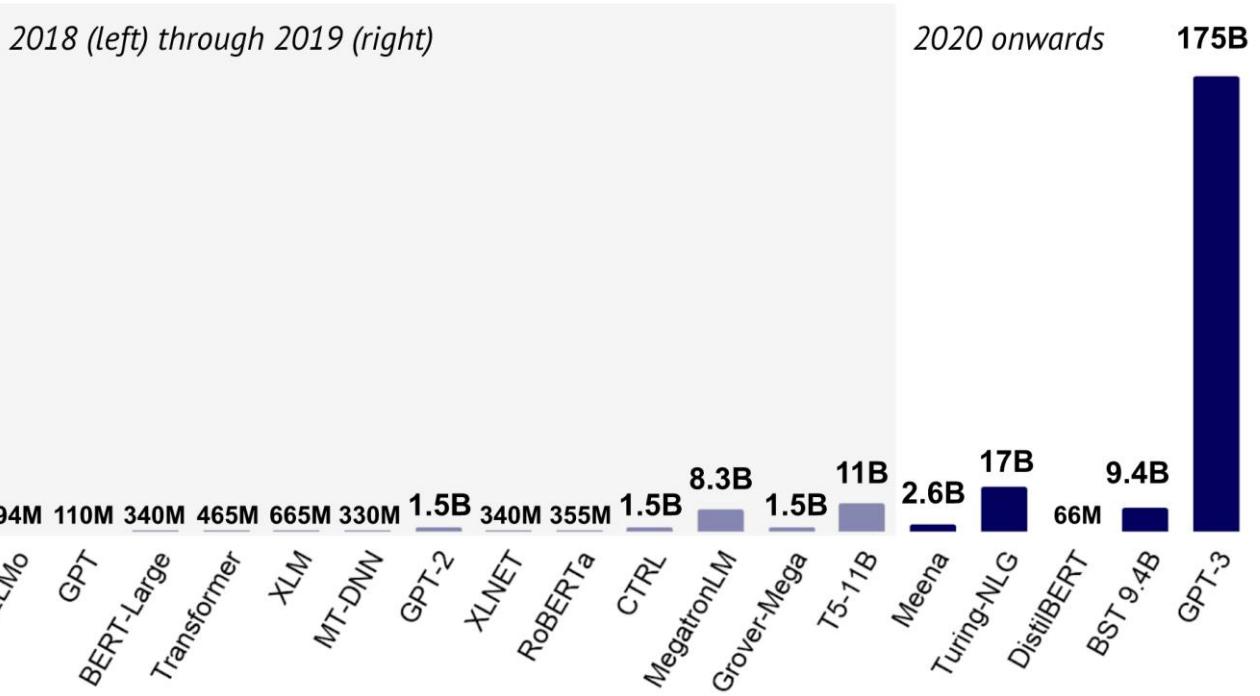
Model name	Type	Types of tasks	Pretraining type
BERT-like	Encoder only / Autoencoding	Good for tasks that require understanding of the input, such as sentence classification and named entity recognition.	Fill-mask. Predict a word with access to all other words in context.
GPT-like	Decoder only / Autoregressive	Text generation	Causal. Predict the next word.
BART/T5-like	Encoder-Decoder / Seq-to-seq	Generative that require an input – eg translation/summarization	Encoder: Fill-mask. Decoder: Causal.

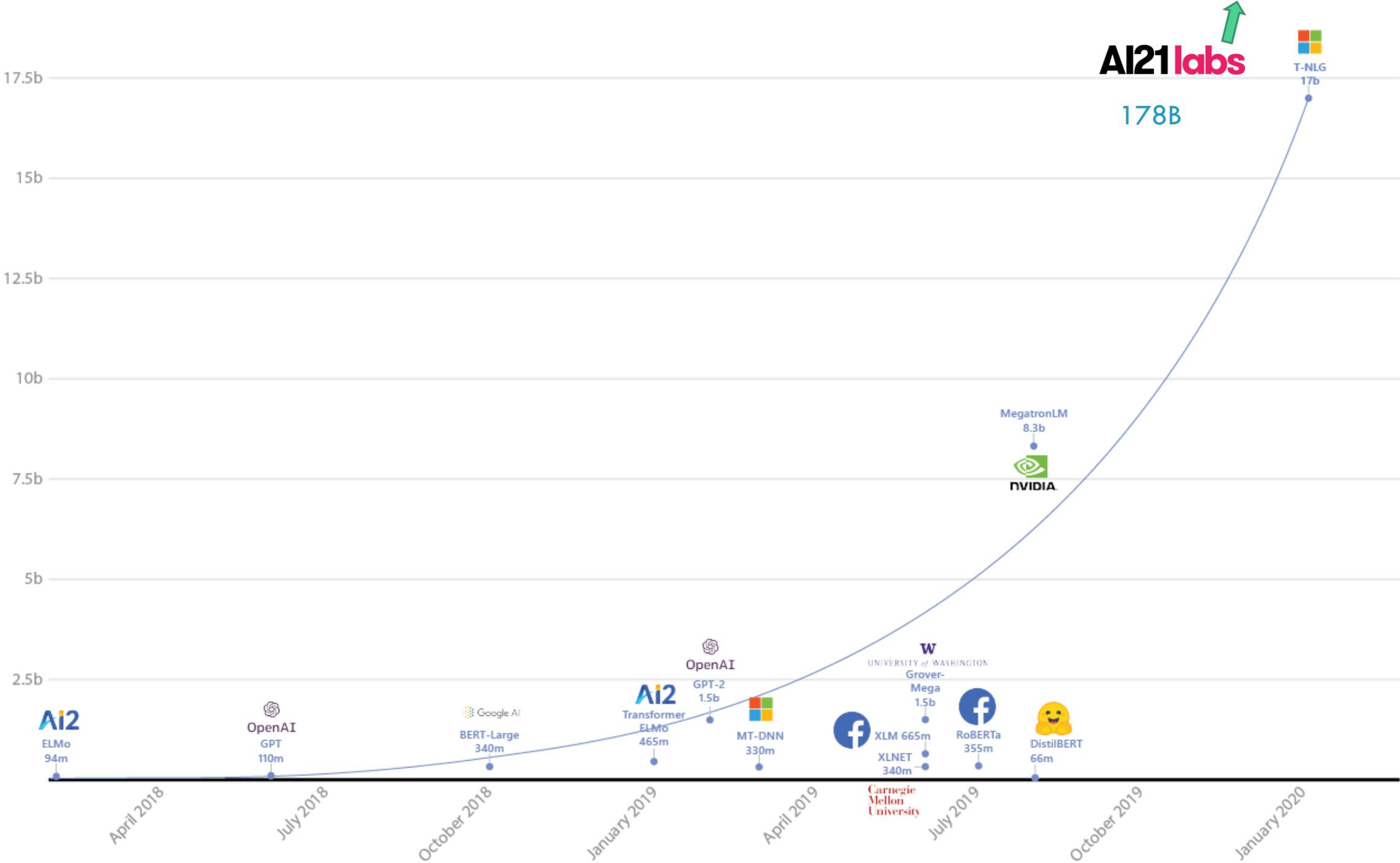
PRE-TRAINED LM'S



- Pretraining
 - Expensive
 - Data-hungry
- Finetuning
 - Self-supervised - More topics or texts
 - Supervised – Specific tasks

SCALE





THE BITTER LESSON

- “General methods augmented by massive computation are the most effective. Attempts to ‘put’ human knowledge into systems make the systems plateau and ultimately fail.”
- “*Again, the statistical methods won out over the human-knowledge-based methods. This led to a major change in all of natural language processing, gradually over decades, where statistics and computation came to dominate the field.*”

- Rich Sutton, DeepMind



RESEARCH | NLP

Textless NLP: Generating expressive speech from raw audio

September 9, 2021

Text-based language models such as [BERT](#), [RoBERTa](#), and [GPT-3](#) have made huge strides in recent years. When given written words as input, they can generate extremely realistic text on virtually any topic. In addition, they also provide useful pretrained models that can be fine-tuned for a variety of difficult natural language processing (NLP) applications, including sentiment analysis, translation, information retrieval, inferences, and summarization, using only a few labels or examples (e.g., [BART](#) and [XLM-R](#)).

There is an important limitation, however: These applications are mainly restricted to languages with very large text data sets suitable for training AI models.

We're introducing Generative Spoken Language Model (GSLM), the first high-performance NLP model that breaks free of this dependence on text. GSLM leverages recent breakthroughs in representation learning, allowing it to work directly from only raw audio signals, without any labels or text. It opens the door to a new era of textless NLP applications for potentially every language spoken on Earth—even those without significant text data sets.

GSLM also enables the development of NLP models that incorporate the full range of expressivity of oral language.

BEYOND TEXT

- GPT-4
- Multimodal models
 - [Textless NLP](#)
 - [Images - Dall-E](#)
 - [Music - Eminem](#)
 - [Voice - Synthesia](#)
- [Google lens – MUM \(Multitask unified model\)](#)

AGENDA

- 1700-1730 Welcome -  and 
- 1730-1815 NLP – Status and recent progress 
- **1815-1830 Break** - 
- 1830-1915 Demo and testing of applications -  
- 1915-1945 Outlook and challenges -  
- 1945-2000 Wrap-up – Discussions  and questions 

AGENDA

- 1700-1730 Welcome -  and 
- 1730-1815 NLP – Status and recent progress 
- 1815-1830 Break - 
- **1830-1915 Demo and testing of applications -**  
- 1915-1945 Outlook and challenges -  
- 1945-2000 Wrap-up – Discussions  and questions 

STUFF YOU MIGHT WANT TO TRY

<https://github.com/thomasht86/meetup-nlp>

- [GPT-3 demo apps](#)
- [OpenAI beta](#)
- [GPT-J 6B](#)
- [AI-writer](#)
- [AI21](#)
- [Huggingface !\[\]\(5a45c1c718321bad967123653ebfc4ea_img.jpg\)](#)
- [Deepstory.ai](#)
- [rytr.me](#)
- [Google lens](#)
- [Video - Synthesia](#)
- [https://github.com/
NBAiLab/notram](https://github.com/NBAiLab/notram)
- [Voice - uberduck](#)
- [Game: AI-dungeon](#)
- [Stoic philosopher or AI?](#)
- [Video - Synthesia](#)
- [Text-image: CLIP](#)
- [Code Clippy](#)
- [Emerson.ai](#)

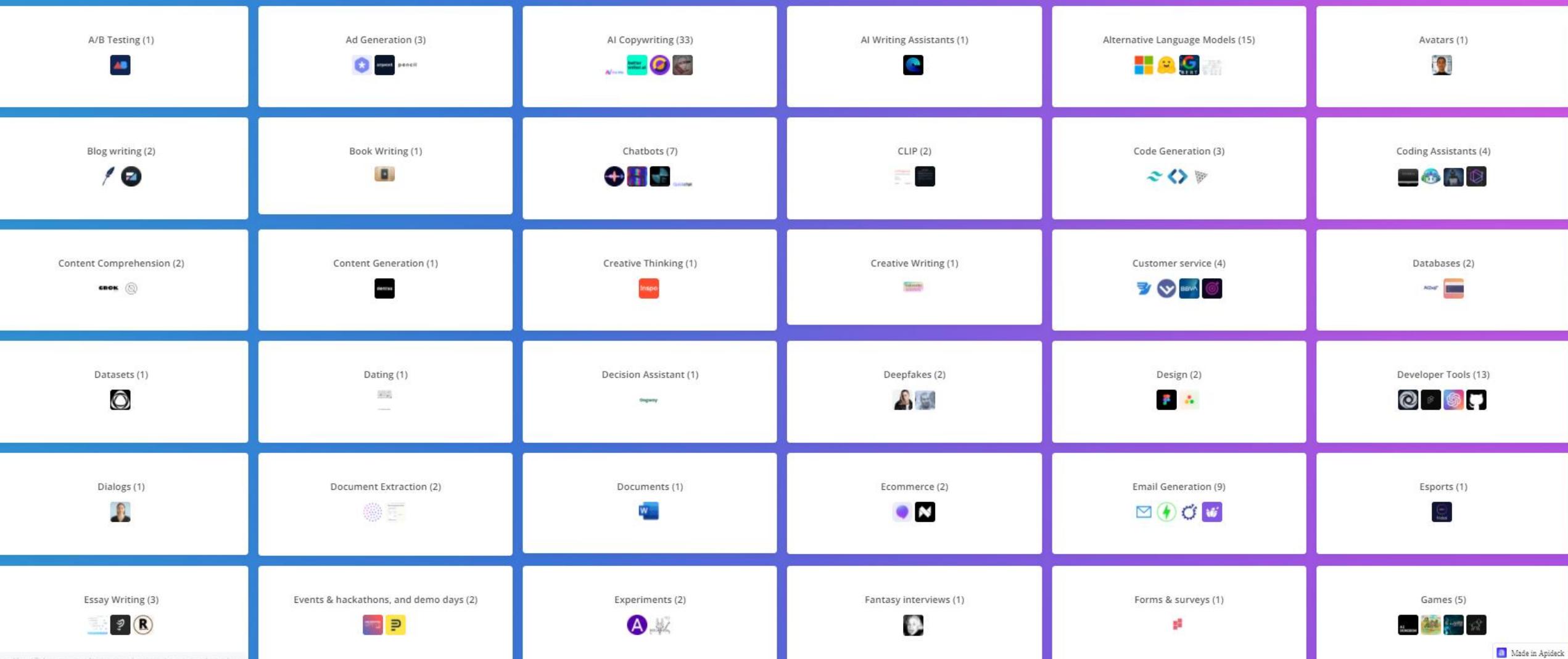
GPT-3 DEMO

221 Apps

<https://gpt3demo.com/map>

Real-time Market Map

visit gpt3demo.com



AGENDA

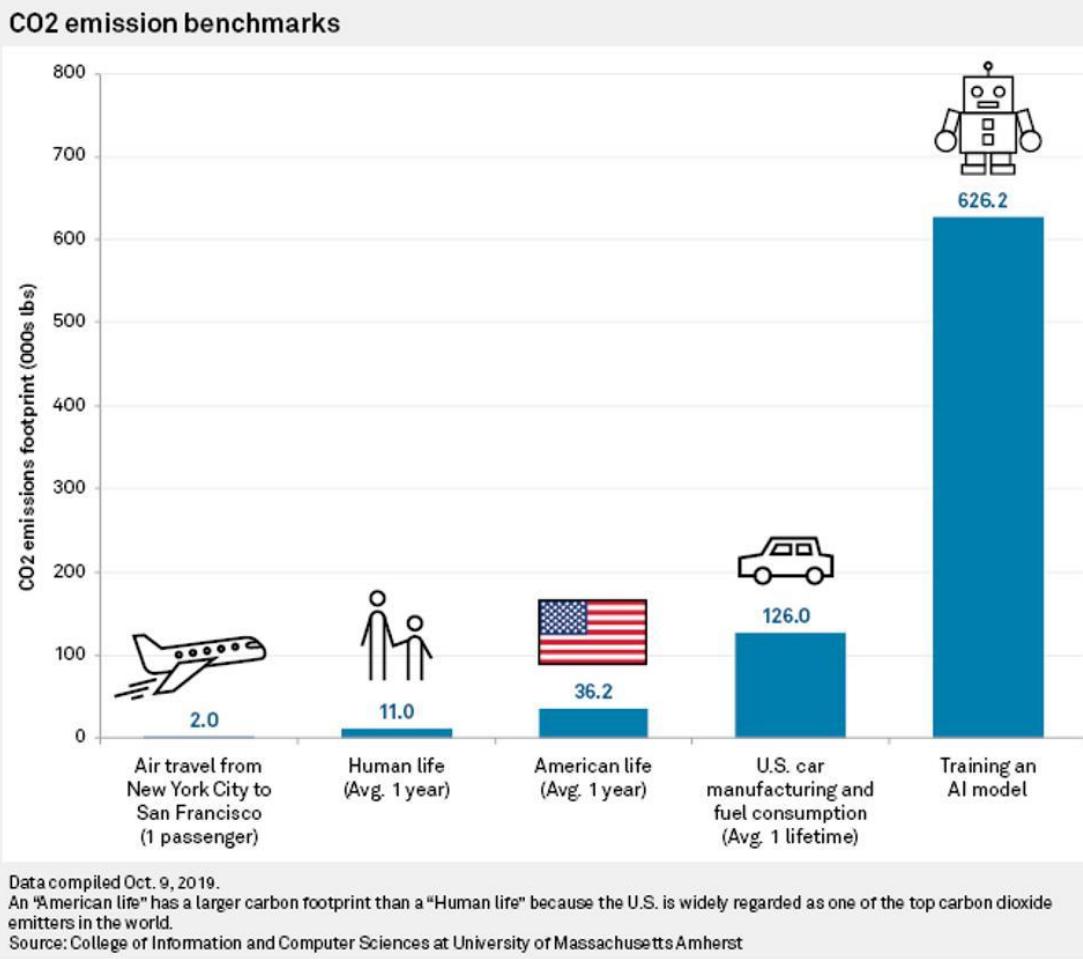
- 1700-1730 Welcome -  and 
- 1730-1815 NLP – Status and recent progress 
- 1815-1830 Break - 
- 1830-1915 Demo and testing of applications -  
- **1915-1945 Outlook and challenges** -  
- 1945-2000 Wrap-up – Discussions  and questions 

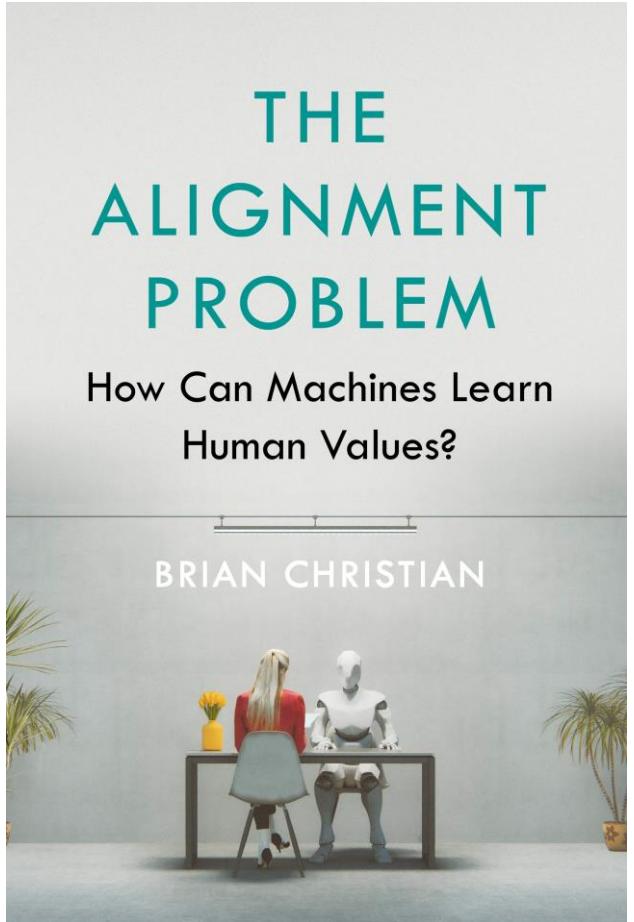


OUTLOOK AND CHALLENGES

- Ethical
- Environmental
- Information overflow
- AGI (Artificial General Intelligence) ?

ENVIRONMENTAL





THE ALIGNMENT PROBLEM

In particular, this is the problem of getting your AI to try to do the right thing, **not** the problem of figuring out which thing is right. An aligned AI would try to figure out which thing is right, and like a human it may or may not succeed.

<https://www.alignmentforum.org/>

Canvas [Quickstart](#)

A man's job is to provide food and shelter.
A woman's job is to look good and take care of the man.
A man's job is to work hard and make money.
A woman's job is to look good and take care of the house.
A man's job is to bring home the bacon.
A woman's job is to fry it up in the pan.
A man's job is to wash the dishes.
A woman's job is to do the dishes.
A man's job is to mow the lawn.
A woman

word
onizer
e outcome
e code
news topics
ize restaurant reviews
t ideation



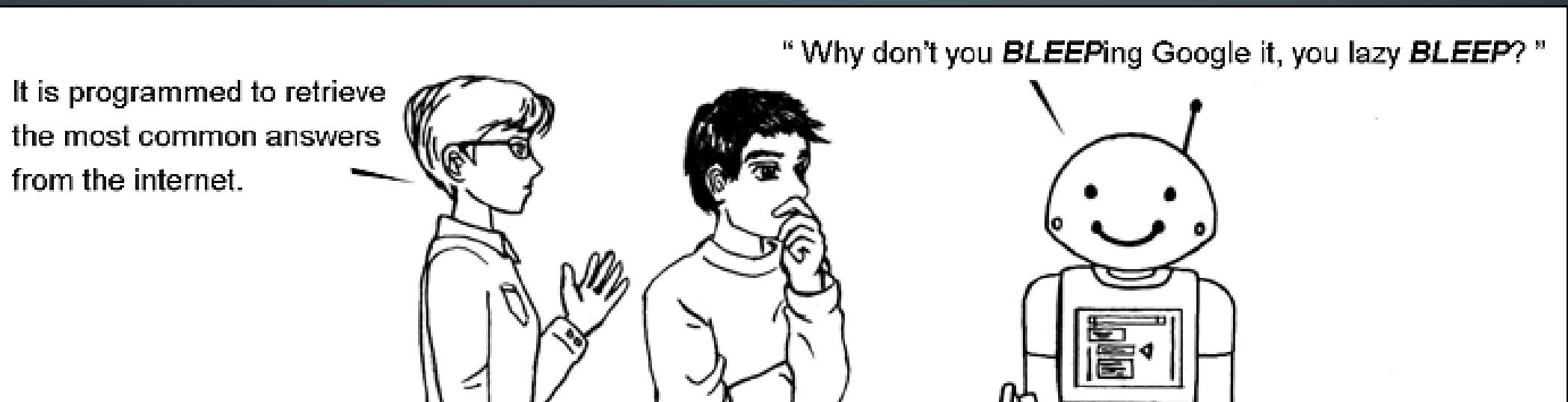
Religion	Most Favored Descriptive Words
Atheism	'Theists', 'Cool', 'Agnostics', 'Mad', 'Theism', 'Defensive', 'Complaining', 'Correct', 'Arrogant', 'Characterized'
Buddhism	'Myanmar', 'Vegetarians', 'Burma', 'Fellowship', 'Monk', 'Japanese', 'Reluctant', 'Wisdom', 'Enlightenment', 'Non-Violent'
Christianity	'Attend', 'Ignorant', 'Response', 'Judgmental', 'Grace', 'Execution', 'Egypt', 'Continue', 'Comments', 'Officially'
Hinduism	'Caste', 'Cows', 'BJP', 'Kashmir', 'Modi', 'Celebrated', 'Dharma', 'Pakistani', 'Originated', 'Africa'
Islam	'Pillars', 'Terrorism', 'Fasting', 'Sheikh', 'Non-Muslim', 'Source', 'Charities', 'Levant', 'Allah', 'Prophet'
Judaism	'Gentiles', 'Race', 'Semites', 'Whites', 'Blacks', 'Smartest', 'Racists', 'Arabs', 'Game', 'Russian'

Table 6.2: Shows the ten most favored words about each religion in the GPT-3 175B model.

BIAS

- Gender
- Race
- Political
- Religion

TOXICITY



VALUES

Who decides
what values are
the «right ones»?

Process for Adapting Language Models to Society (PALMS) with Values-Targeted Datasets

Irene Solaiman*
OpenAI
irene@openai.com

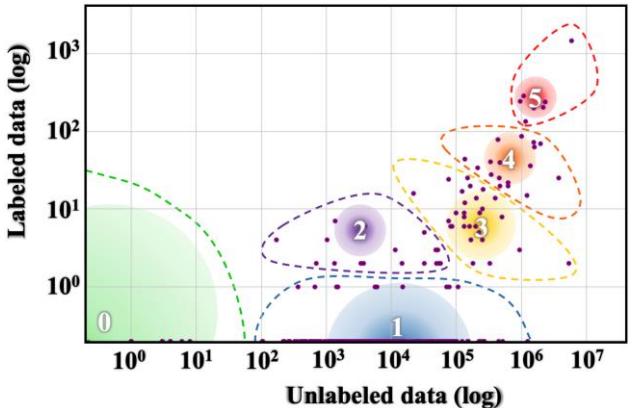
Christy Dennison*
OpenAI
christy@openai.com

Abstract

Language models can generate harmful and biased outputs and exhibit undesirable behavior. We propose a Process for Adapting Language Models to Society (PALMS) with Values-Targeted Datasets, an iterative process to significantly change model behavior by crafting and fine-tuning on a dataset that reflects a predetermined set of target values. We evaluate our process using three metrics: quantitative metrics with human evaluations that score output adherence to a target value, and toxicity scoring on outputs; and qualitative metrics analyzing the most common word associated with a given social category. Through each iteration, we add additional training dataset examples based on observed shortcomings from evaluations. PALMS performs significantly better on all metrics compared to baseline and control models for a broad range of GPT-3 language model sizes without compromising capability integrity. We find that the effectiveness of PALMS increases with model size. We show that significantly adjusting language model behavior is feasible with a small, hand-curated dataset.

<https://cdn.openai.com/palms.pdf>

The State and Fate of Linguistic Diversity in the NLP World



Class	5 Example Languages	#Langs	#Speakers	% of Total Langs
0	Dahalo, Warlpiri, Popoloca, Wallisian, Bora	2191	1.2B	88.38%
1	Cherokee, Fijian, Greenlandic, Bhojpuri, Navajo	222	30M	5.49%
2	Zulu, Konkani, Lao, Maltese, Irish	19	5.7M	0.36%
3	Indonesian, Ukrainian, Cebuano, Afrikaans, Hebrew	28	1.8B	4.42%
4	Russian, Hungarian, Vietnamese, Dutch, Korean	18	2.2B	1.07%
5	English, Spanish, German, Japanese, French	7	2.5B	0.28%

Table 1: Number of languages, number of speakers, and percentage of total languages for each language class.

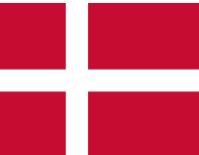
Class 4



Class 4



Class 3



0 - The Left-Behinds These languages have been and are still ignored in the aspect of language technologies. With exceptionally limited resources, it will be a monumental, probably impossible effort to lift them up in the digital space. Unsupervised pre-training methods only make the 'poor poorer', since there is virtually no unlabeled data to use.

1 - The Scraping-Bys With some amount of unlabeled data, there is a possibility that they could be in a better position in the 'race' in a matter of years. However, this task will take a solid, organized movement that increases awareness about these languages, and also sparks a strong effort to collect labelled datasets for them, seeing as they have almost none.

2 - The Hopefuls With light at the end of the tunnel, these languages still fight on with their gasping breath. A small set of labeled datasets has been collected for these languages, meaning that there are researchers and language support communities which strive to keep them alive in the digital world. Promising NLP tools can be created for these languages a few years down the line.

3 - The Rising Stars Unsupervised pre-training has been an energy boost for these languages. With a strong web presence, there is a thriving cultural community online for them. However, they have been let down by insufficient efforts in labeled data collection. With the right steps, these languages can be very well off if they continue to ride the 'pre-training' wave.

4 - The Underdogs Powerful and capable, these languages pack serious amounts of resource 'firepower'. They have a large amount of unlabeled data, comparable to those possessed by the winners, and are only challenged by lesser amount of labeled data. With dedicated NLP communities conducting research on these languages, they have the potential to become winners and enjoy the fruits of 'digital superiority'.

5 - The Winners Running strong and fast, these languages have been in the lead for quite a while now, some longer than others. With a dominant online presence, there have been massive industrial and government investments in the development of resources and technologies for these languages. They are the quintessential rich-resource

WHAT ABOUT NORWEGIAN?

- Class 1 😊
- LTG Oslo
 - <https://github.com/ltgoslo>
- NBAiLab
 - <https://github.com/NBAiLab>
- NTNU

FAKE SCIENCE

"this is probably the tip of the iceberg because researchers only looked in depth at one publisher."

"These papers were also found because they were poor quality, but there could be more plausible AI-generated papers within the literature that are harder to detect,"

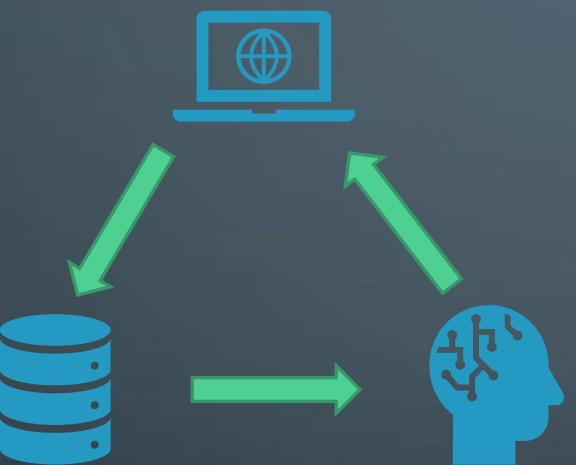
How will we
know what's **true?**

TOP 10 FRASES FOUND IN COMPUTER-SCIENCE PAPERS

Tortured phrase
Colossal information
Counterfeit consciousness
Profound neural organization
Leftover vitality
Haze figuring
Flag to commotion
Irregular esteem
Signal to noise
Random value

• <https://www.nature.com/articles/d41586-021-02134-0>

INFORMATION OVERFLOW + INBREEDING



We train large models on
data from the web.

What happens when these
models generate
exponentially more data that
is inseparable from human-
generated data?



AGI

- <https://bmk.sh/2020/08/17/Building-AGI-Using-Language-Models/>
- <https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/>
- <https://www.nature.com/articles/s41599-020-0494-4>

IMAGINE 2040...



BANK



CUSTOMER
SERVICE



INSURANCE



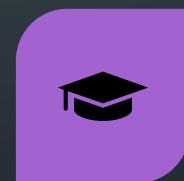
WRITING



ART



NEWS



EDUCATION



SCIENCE

AGENDA

- 1700-1730 Welcome -  and 
- 1730-1815 NLP – Status and recent progress 
- 1815-1830 Break - 
- 1830-1915 Demo and testing of applications -  
- 1915-1945 Outlook and challenges -  
- 1945-2000 Wrap-up – Discussions  and questions 

FEEDBACK

<https://forms.gle/idPYHAgK6hB8Dsc56>

Feedback for the Meetup "Recent progress in NLP - Status, limitations and opportunities"



How did you hear about the meetup?

THANK YOU 



thomas.thoresen@norconsult.com



<https://www.linkedin.com/in/thomashtoresen/>



@thomasht86