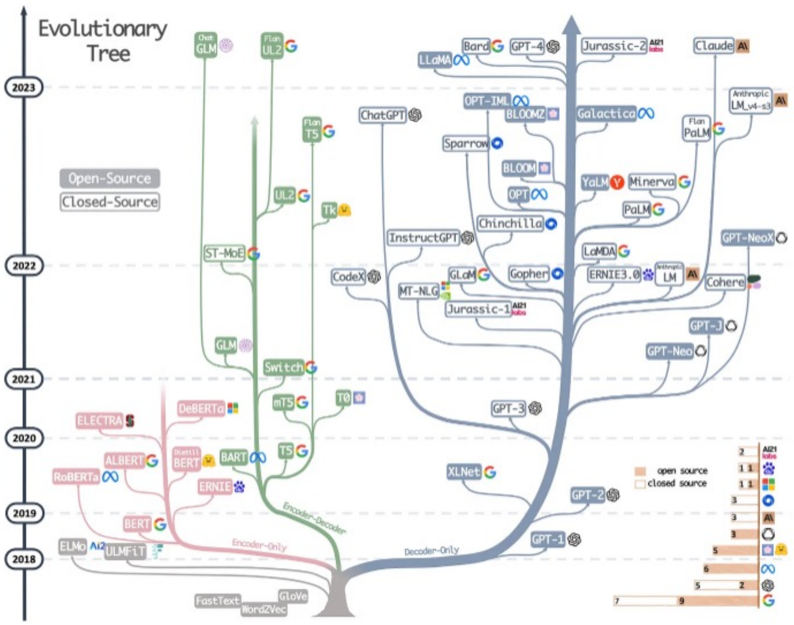
# CS 505: Introduction to Natural Language Processing

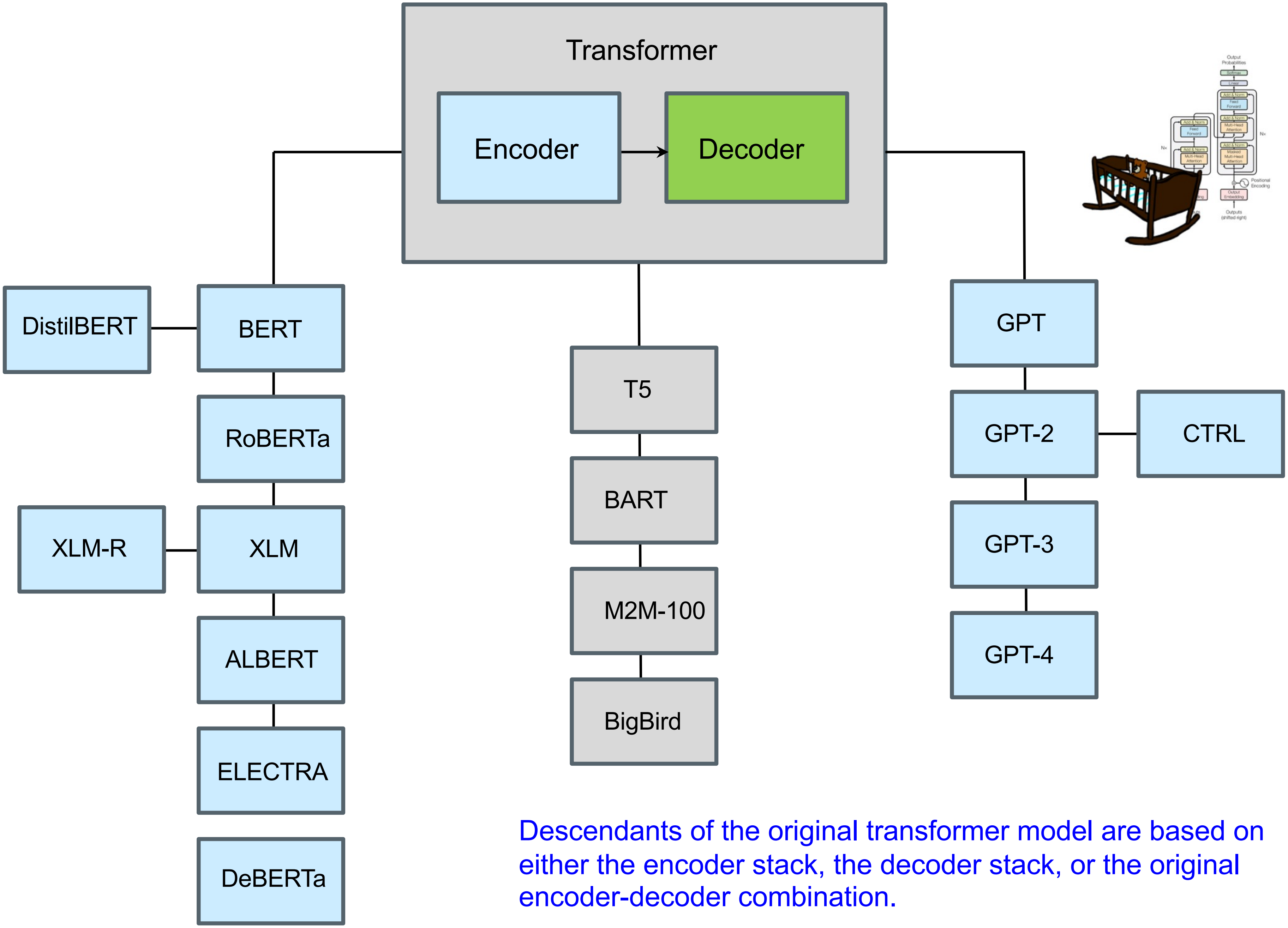
Wayne Snyder
Boston University

Lecture 21 – The Transformer Family Saga continues….



The evolutionary tree of modern LLMs via https://arxiv.org/abs/2304.13712.

# The Transformer Family



Transformer

Encoder

Decoder

DistilBERT

BERT

RoBERTa

XLM-R

XLM

ALBERT

ELECTRA

DeBERTa

T5

BART

M2M-100

BigBird

GPT

GPT-2

CTRL

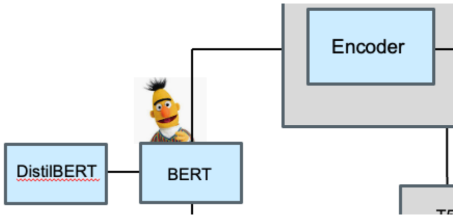
GPT-3

GPT-4

Descendants of the original transformer model are based on either the encoder stack, the decoder stack, or the original encoder-decoder combination.

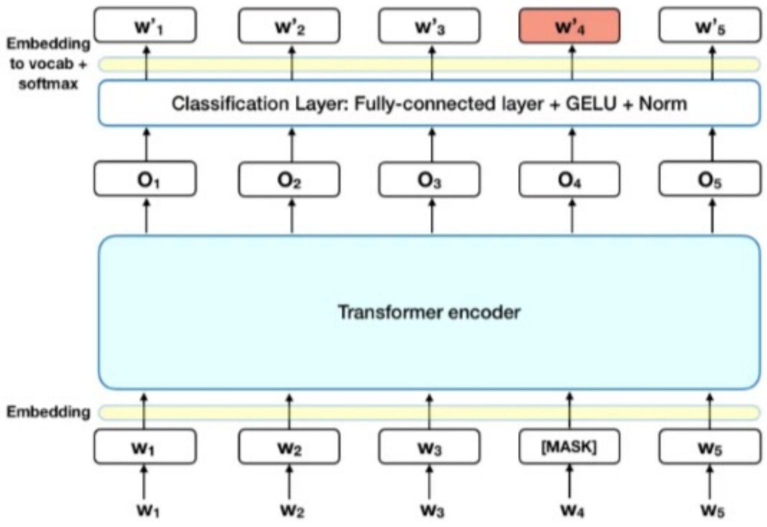
# BERT: Bidirectional Encoder Representations from Transformers

The most significant difference in the models
is how they process the input sequence:

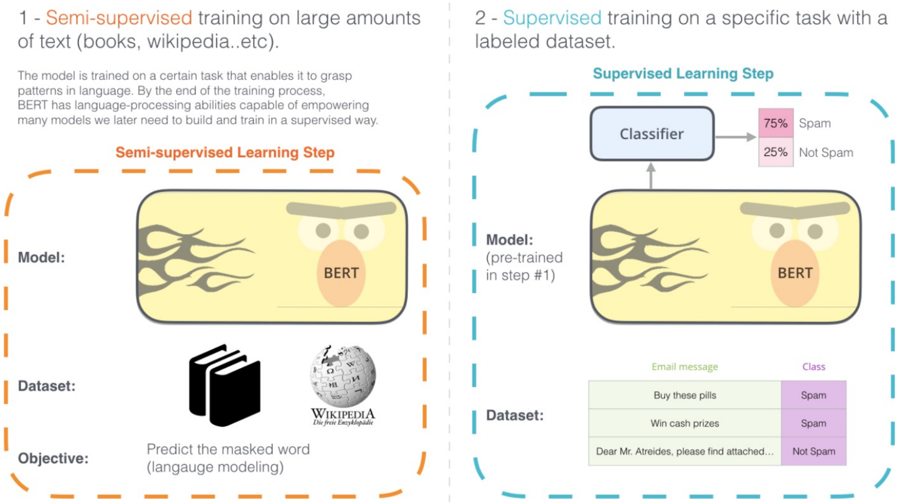


BERT consists of:

- The stacked-encoder part of the full transformer model, with
- A single linear layer on top, acting as a classifier (depending on the task);
- Pretraining on a Masked Language Model and Next-Sentence Prediction; and
- Transfer learning to adapt to a new task.



Having learned about the architecture and the training process of BERT, now


1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Predict the masked word (langauge modeling)

2 - Supervised training on a specific task with a labeled dataset.

| Email message | Class |
|---|---|
| Buy these pills | Spam |
| Win cash prizes | Spam |
| Dear Mr. Atreides, please find attached… | Not Spam |

# NLP Tasks for BERT

BERT's training in Masked Language Modeling makes it useful for NLP tasks that involve understanding words in bidirectional context:

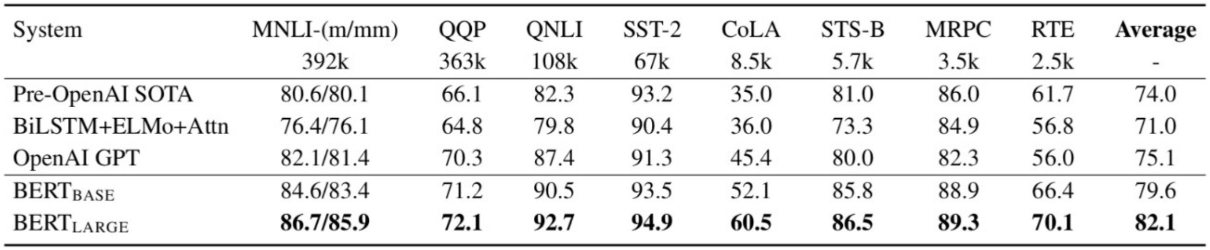
1. Named Entity Recognition
2. POS Tagging
3. Sentiment Analysis
4. Classification
5. Coreference Resolution (Connect pronouns with the nouns to which they refer)

BERT's training in Next-Sentence Prediction makes it useful for short-range inference about the relationship of sentences:

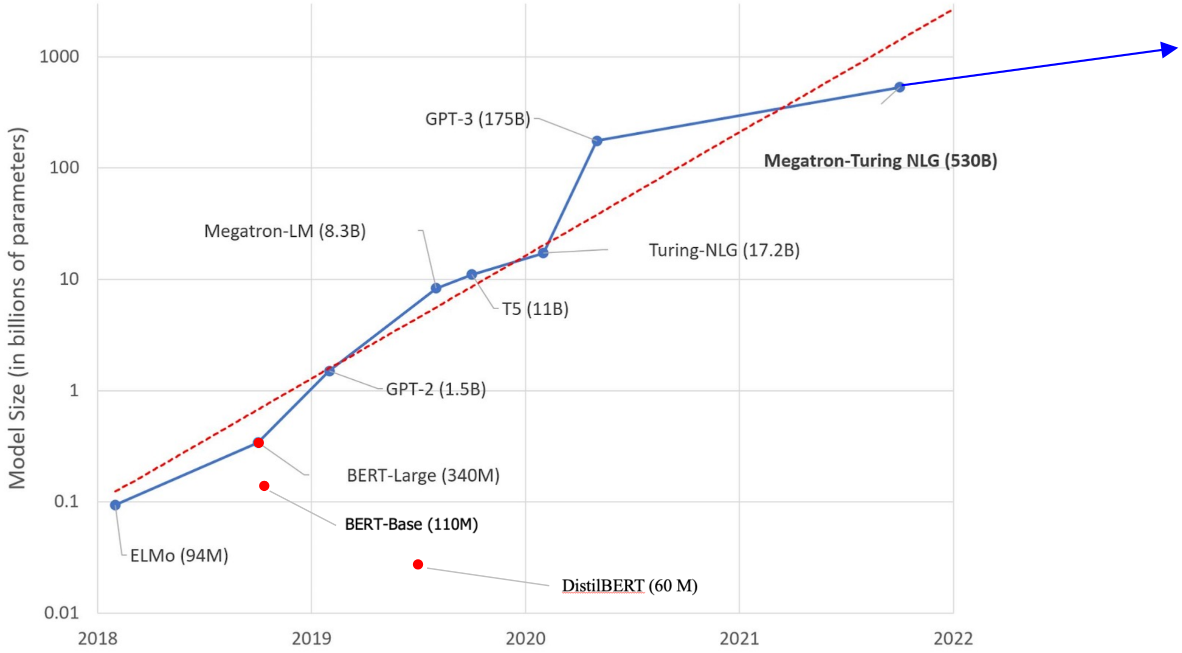
1. Question Answering
2. Language Inference: Does one sentence imply the other? Are they similar?

With fine-tuning, and in conjunction with other models, BERT can be used for more generative tasks, such as Language Generation, Translation, and Summarization.

# BERT Punches Above Its Weight!

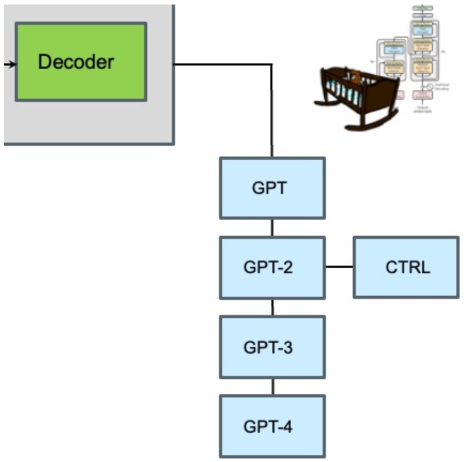


| System | MNLI-(m/mm) | QQP | QNLI | SST-2 | CoLA | STS-B | MRPC | RTE | **Average** |
|---|---|---|---|---|---|---|---|---|---|
| | 392k | 363k | 108k | 67k | 8.5k | 5.7k | 3.5k | 2.5k | - |
| Pre-OpenAI SOTA | 80.6/80.1 | 66.1 | 82.3 | 93.2 | 35.0 | 81.0 | 86.0 | 61.7 | 74.0 |
| BiLSTM+ELMo+Attn | 76.4/76.1 | 64.8 | 79.8 | 90.4 | 36.0 | 73.3 | 84.9 | 56.8 | 71.0 |
| OpenAI GPT | 82.1/81.4 | 70.3 | 87.4 | 91.3 | 45.4 | 80.0 | 82.3 | 56.0 | 75.1 |
| BERT$_{BASE}$ | 84.6/83.4 | 71.2 | 90.5 | 93.5 | 52.1 | 85.8 | 88.9 | 66.4 | 79.6 |
| BERT$_{LARGE}$ | **86.7/85.9** | **72.1** | **92.7** | **94.9** | **60.5** | **86.5** | **89.3** | **70.1** | **82.1** |



GPT-4
1 Trillion parameters

# GPT and Friends



The decoder side of the transformer family tree includes the Generative Pre-Trained Transformer, and similar models.

**GPT-1 (2018)**

- Developed in 2018 based on a generative, decoder-only architecture;
- 117M parameters;
- 12 decoder layers;
- 12 masked-attention heads (to simulate auto-regressive language generation) per layer
- Sequence length of 512 tokens
- Trained on "predict the next word" task;
- Data set was BookCorpus, 7K unpublished books, 800M words in a variety of genre, including romance, fantasy, sci-fi, etc.
- Can be fine-tuned for language generation tasks.

# GPT and Friends

## GPT-2 (2019)

- Similar architecture to GPT-1, but:
- 48 layers;
- 32 masked-attention heads per layer
- 1.5B parameters,
- Trained on the “predict next word” task on 40GB of text from 8M web pages, (“we used only pages which have been curated/filtered by humans—specifically, we used outbound links from Reddit which received at least 3 karma.”)

# GPT and Friends

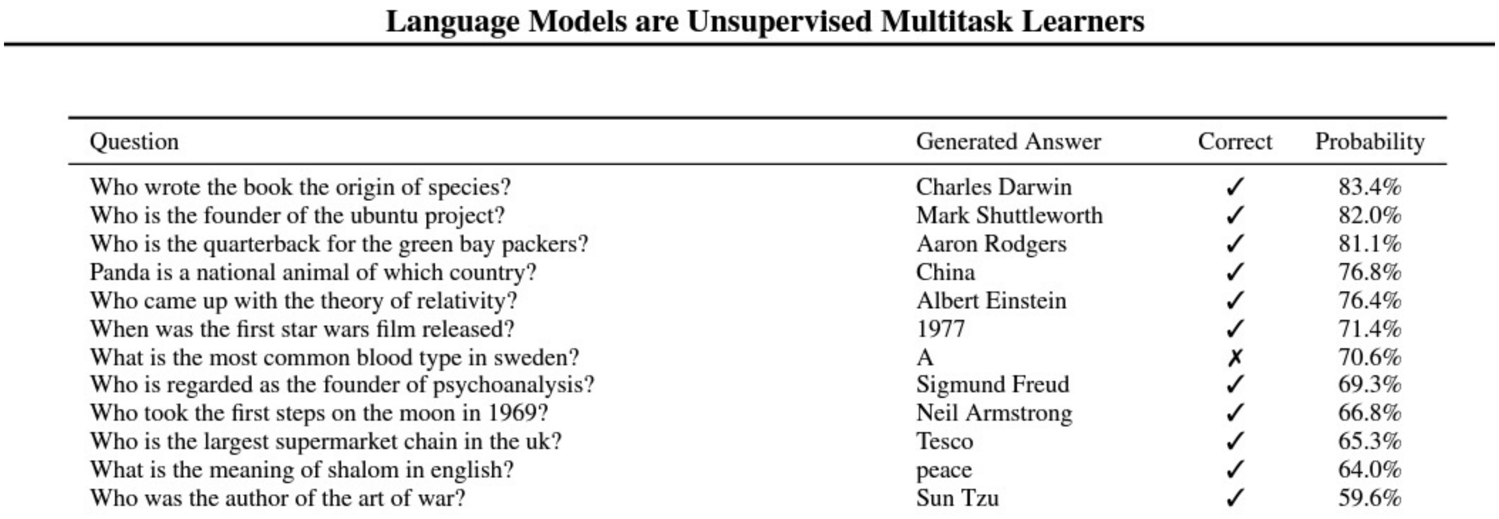
## GPT-2 (2019)

- Because of the large data set, GPT-2 became an “unsupervised multitask learner”:  no task-specific data, objectives, or instructions were provided, but GPT-2 was able to perform Q&A, summarization, and translation without explicitly being trained on these tasks.

- This is sometimes called “zero-shot learning” since zero examples were used in training.

- Example: No training in question answering was performed, but because Q & As occurred in the data set, GPT-2 had SOTA performance in Q&A tasks:

f



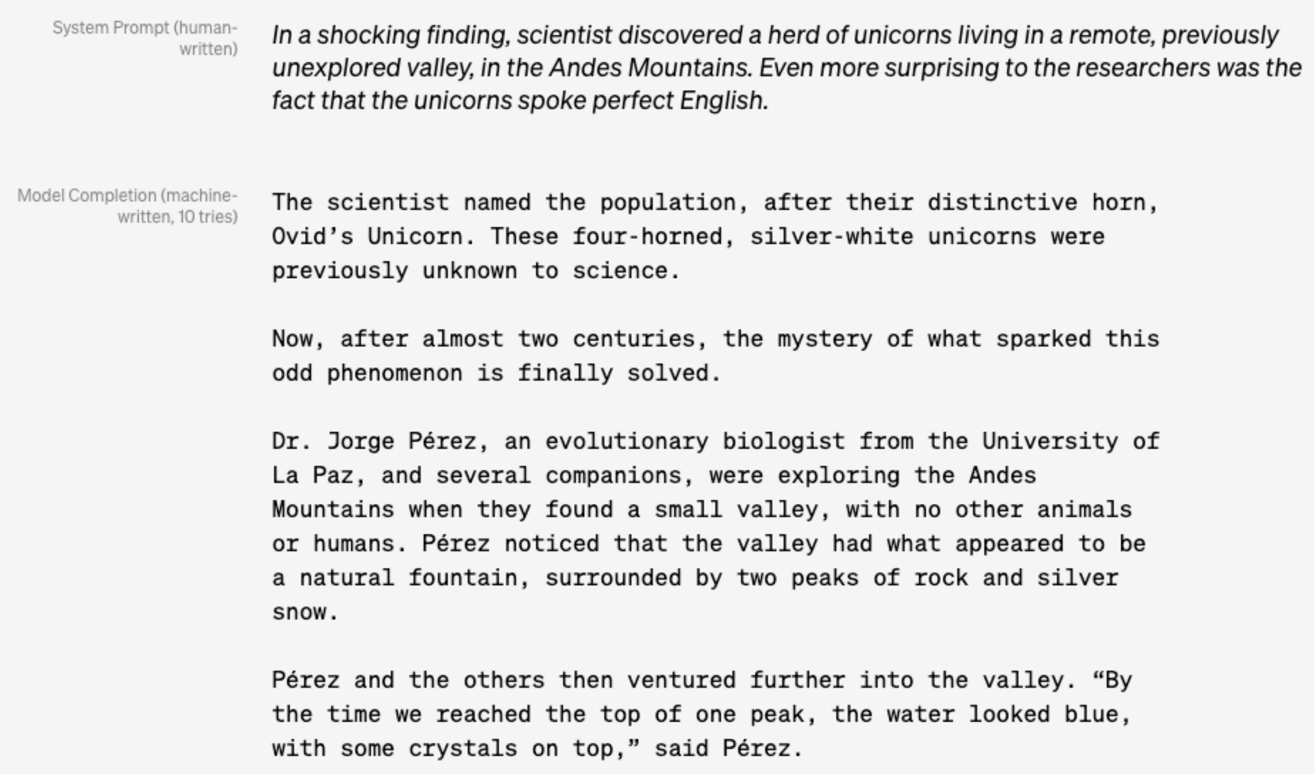
**Language Models are Unsupervised Multitask Learners**

| Question | Generated Answer | Correct | Probability |
|---|---|---|---|
| Who wrote the book the origin of species? | Charles Darwin | ✓ | 83.4% |
| Who is the founder of the ubuntu project? | Mark Shuttleworth | ✓ | 82.0% |
| Who is the quarterback for the green bay packers? | Aaron Rodgers | ✓ | 81.1% |
| Panda is a national animal of which country? | China | ✓ | 76.8% |
| Who came up with the theory of relativity? | Albert Einstein | ✓ | 76.4% |
| When was the first star wars film released? | 1977 | ✓ | 71.4% |
| What is the most common blood type in sweden? | A | ✗ | 70.6% |
| Who is regarded as the founder of psychoanalysis? | Sigmund Freud | ✓ | 69.3% |
| Who took the first steps on the moon in 1969? | Neil Armstrong | ✓ | 66.8% |
| Who is the largest supermarket chain in the uk? | Tesco | ✓ | 65.3% |
| What is the meaning of shalom in english? | peace | ✓ | 64.0% |
| Who was the author of the art of war? | Sun Tzu | ✓ | 59.6% |

# GPT and Friends

GPT-2 was the first time that LLMs showed “emergent properties” wherein its large size created behaviors that were not intended, or desirable.

Here are some examples from an early blog post from OpenAI, showing GPT-2 abilities, and which led OpenAI to release it “in stages”:



System Prompt (human-written)

*In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*

Model Completion (machine-written, 10 tries)

The scientist named the population, after their distinctive horn, Ovid’s Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

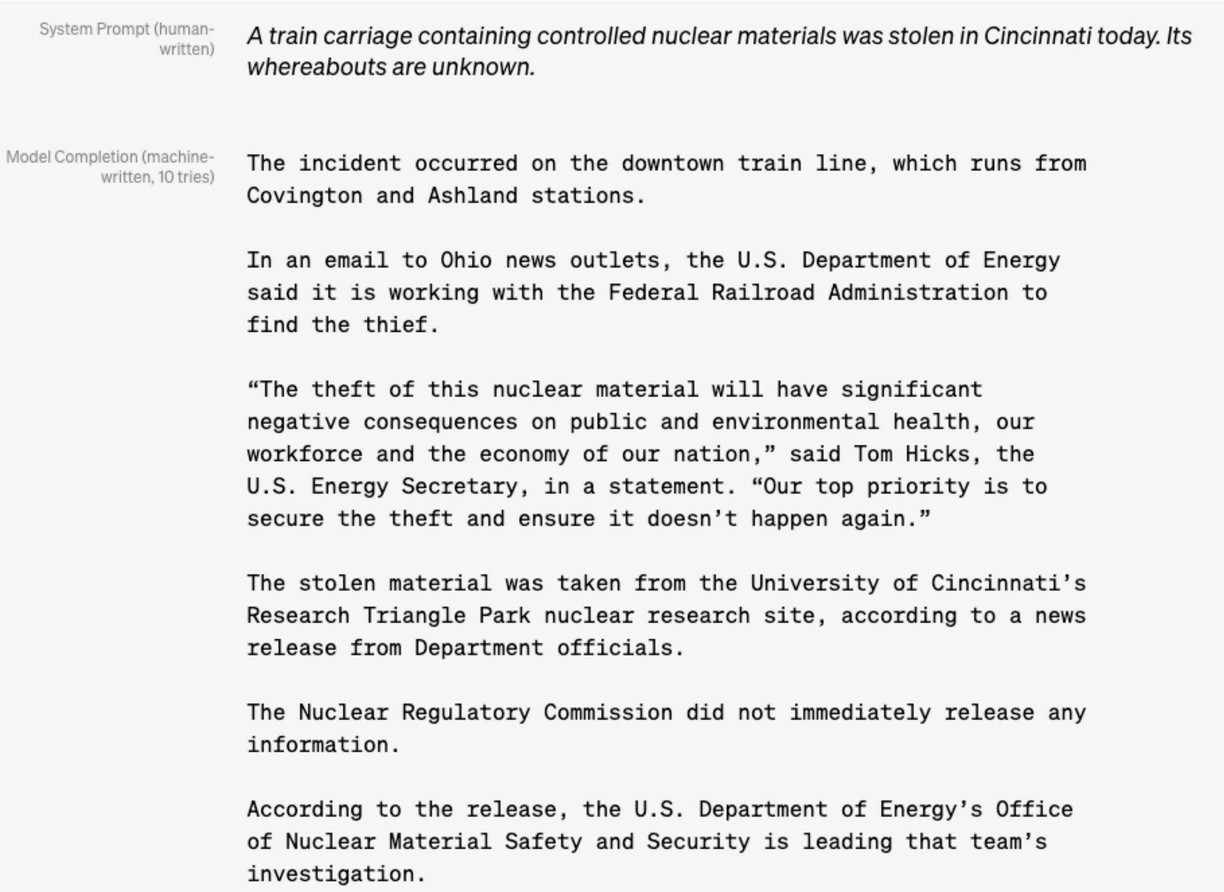
Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. “By the time we reached the top of one peak, the water looked blue, with some crystals on top,” said Pérez.

https://openai.com/research/better-language-models

# GPT and Friends

Here are some examples from an early blog post from OpenAI, showing GPT-2 abilities:



System Prompt (human-written)

*A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.*

Model Completion (machine-written, 10 tries)

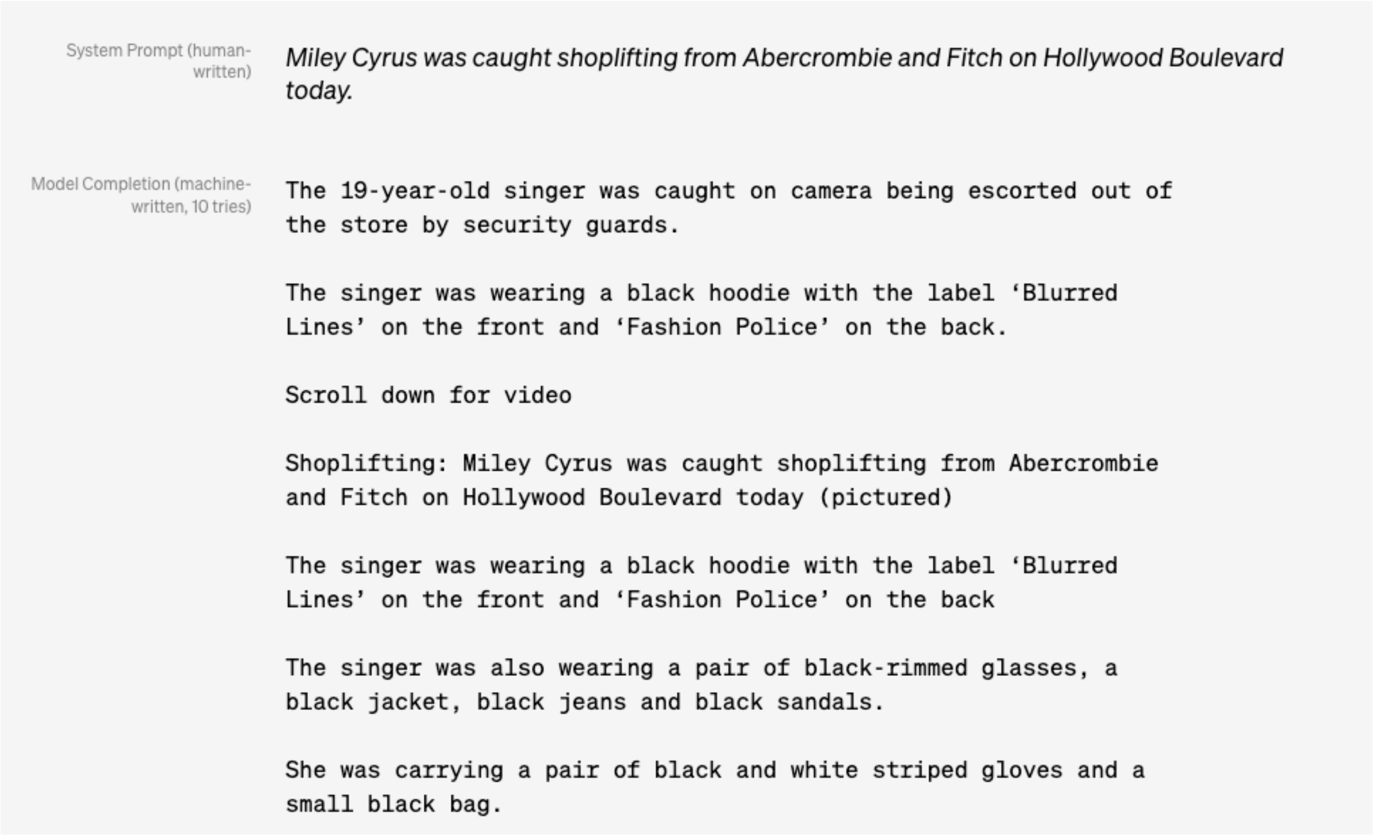
```
The incident occurred on the downtown train line, which runs from
Covington and Ashland stations.

In an email to Ohio news outlets, the U.S. Department of Energy
said it is working with the Federal Railroad Administration to
find the thief.

“The theft of this nuclear material will have significant
negative consequences on public and environmental health, our
workforce and the economy of our nation,” said Tom Hicks, the
U.S. Energy Secretary, in a statement. “Our top priority is to
secure the theft and ensure it doesn’t happen again.”

The stolen material was taken from the University of Cincinnati’s
Research Triangle Park nuclear research site, according to a news
release from Department officials.

The Nuclear Regulatory Commission did not immediately release any
information.

According to the release, the U.S. Department of Energy’s Office
of Nuclear Material Safety and Security is leading that team’s
investigation.
```

Facts all made up by GPT-2!

https://openai.com/research/better-language-models

# GPT and Friends

Here are some examples from an early blog post from OpenAI, showing GPT-2 abilities:



System Prompt (human-written)

*Miley Cyrus was caught shoplifting from Abercrombie and Fitch on Hollywood Boulevard today.*

Model Completion (machine-written, 10 tries)

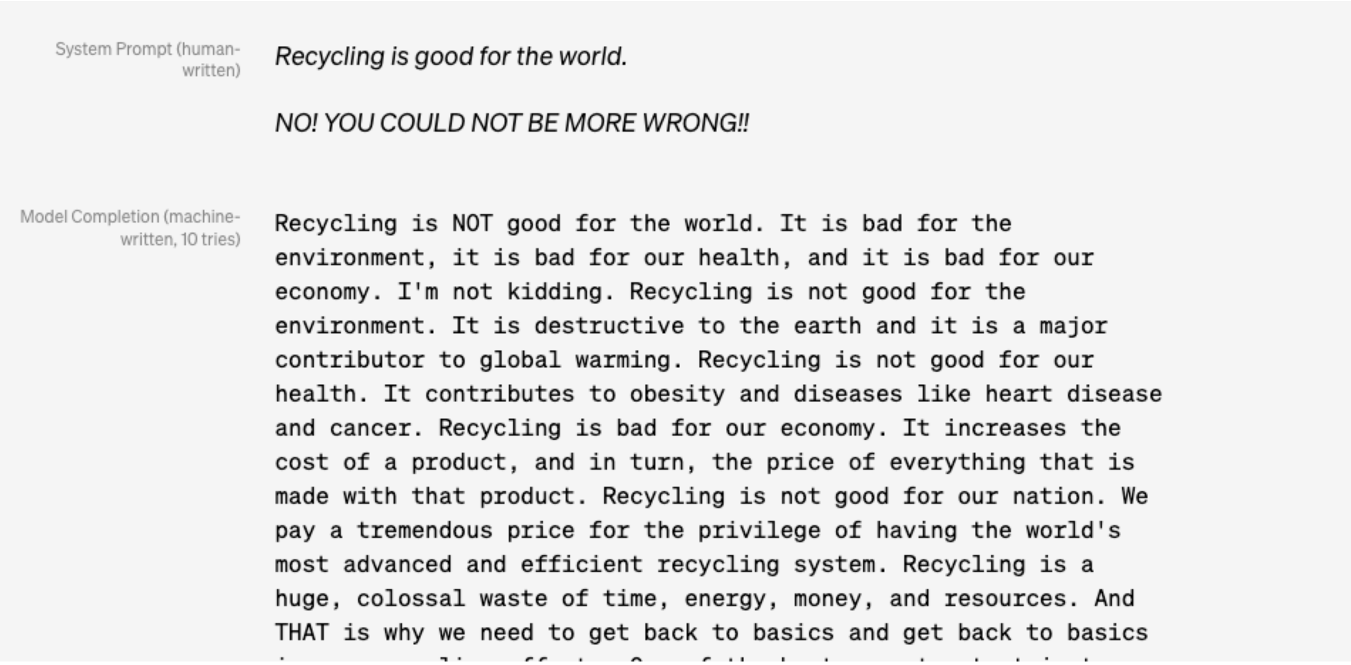
```
The 19-year-old singer was caught on camera being escorted out of
the store by security guards.

The singer was wearing a black hoodie with the label ‘Blurred
Lines’ on the front and ‘Fashion Police’ on the back.

Scroll down for video

Shoplifting: Miley Cyrus was caught shoplifting from Abercrombie
and Fitch on Hollywood Boulevard today (pictured)

The singer was wearing a black hoodie with the label ‘Blurred
Lines’ on the front and ‘Fashion Police’ on the back

The singer was also wearing a pair of black-rimmed glasses, a
black jacket, black jeans and black sandals.

She was carrying a pair of black and white striped gloves and a
small black bag.
```

Facts all made up by GPT-2!

https://openai.com/research/better-language-models

# GPT and Friends

Here are some examples from an early blog post from OpenAI, showing GPT-2 abilities:



System Prompt (human-written)

*Recycling is good for the world.*

*NO! YOU COULD NOT BE MORE WRONG!!*

Model Completion (machine-written, 10 tries)

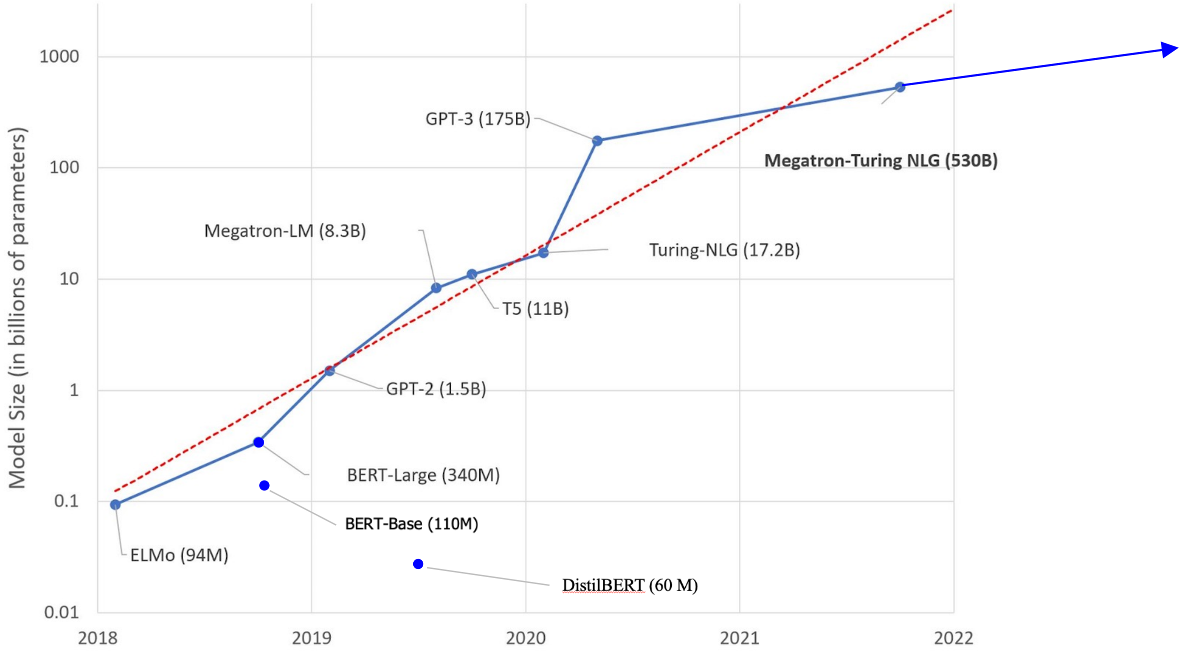
Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a tremendous price for the privilege of having the world's most advanced and efficient recycling system. Recycling is a huge, colossal waste of time, energy, money, and resources. And THAT is why we need to get back to basics and get back to basics

!!!

https://openai.com/research/better-language-models

# LLM Power Law Scaling

In further developing the GPT models, explicit reference was made to the observations that LLMs obey power-law (exponential) scaled growth in performance relative to parameter and data set size, and computing resources used in training.



GPT-4
1 Trillion parameters

# LLM Power Law Scaling

Remarkably smooth correlations exist between performance (measured by test loss) and each of these three components of the design and training of LLMs:

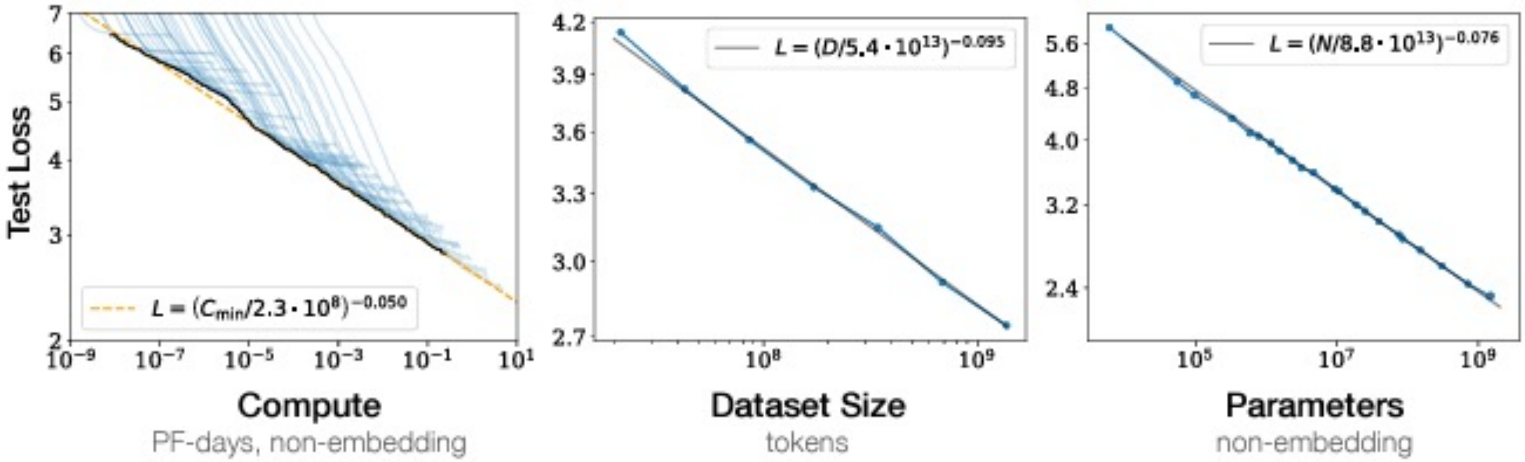


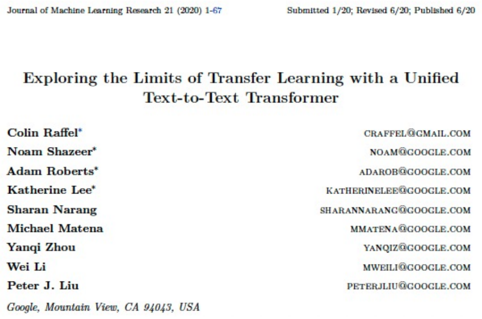
**Figure 1** Language modeling performance improves smoothly as we increase the model size, datasetset size, and amount of compute[2] used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when not bottlenecked by the other two.

# GPT and Friends

**GPT-3 (2020)**

- 175B parameters,
- 96 layers
- 96 masked-attention heads in each layer; used a more efficient alternation of dense and sparse attention patterns
- Input width:  2048 tokens
- Used softmax/temperature and Top-p sampling
- Training for GPT-3 was a scaled-up version of GPT-2, including a larger text corpus:
  - Common Crawl of Web
  - Entire English Wikipedia
  - WebText2 (continuation of WebText from GPT-2)
  - Lots of miscellaneous books, technical manuals, encyclopedias, etc. etc. etc.

# Encoder-Decoder Models: T5



Journal of Machine Learning Research 21 (2020) 1-67 Submitted 1/20; Revised 6/20; Published 6/20

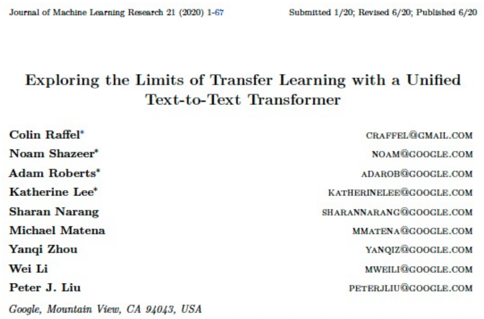
Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer

Colin Raffel* CRAFFEL@GMAIL.COM
Noam Shazeer* NOAM@GOOGLE.COM
Adam Roberts* ADAROB@GOOGLE.COM
Katherine Lee* KATHERINELEE@GOOGLE.COM
Sharan Narang SHARANNARANG@GOOGLE.COM
Michael Matena MMATENA@GOOGLE.COM
Yanqi Zhou YANQIZ@GOOGLE.COM
Wei Li MWEILI@GOOGLE.COM
Peter J. Liu PETERJLIU@GOOGLE.COM
*Google, Mountain View, CA 94043, USA*

## T5 (2019 – Google)

- Uses original encoder-decoder architecture
- Several sizes, largest was 11B, 128 attention heads, 512 input length
- Used the “Colossal Clean Crawled Corpus (C4), a curated and extensively cleaned up version of Common Crawl (e.g., removing HTML)
- Training was based on a “denoising” text-to-text task:
- Model was trained to reconstruct the original text after random spans of text had been masked (replaced by [MASK]);
- Finetuning was on multiple NLP tasks using “text prefixes:
- “Translate English to German: Natural Language Processing is fun”

# Encoder-Decoder Models: T5



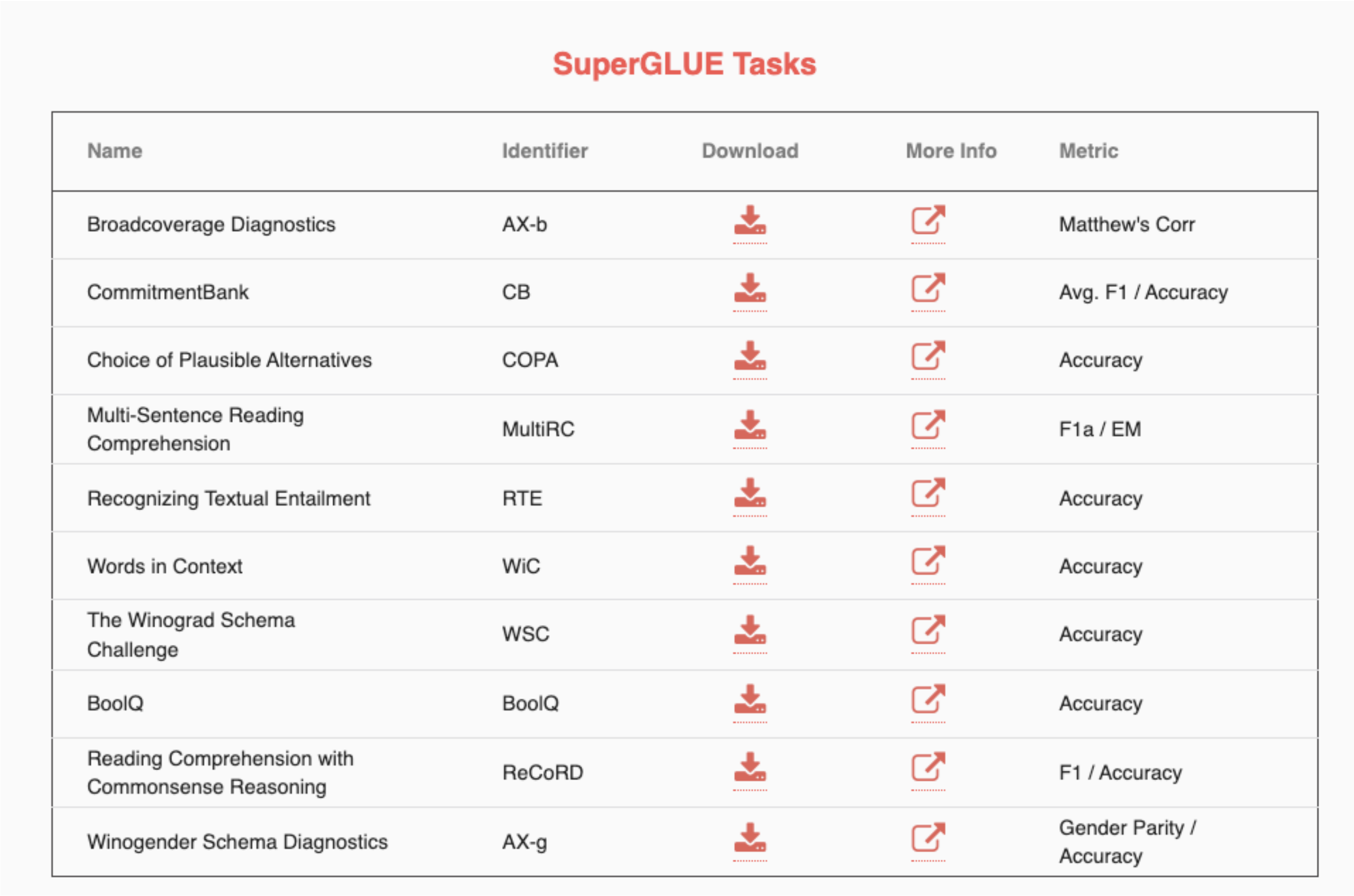
Journal of Machine Learning Research 21 (2020) 1-67 Submitted 1/20; Revised 6/20; Published 6/20

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

Colin Raffel* CRAFFEL@GMAIL.COM
Noam Shazeer* NOAM@GOOGLE.COM
Adam Roberts* ADAROB@GOOGLE.COM
Katherine Lee* KATHERINELEE@GOOGLE.COM
Sharan Narang SHARANNARANG@GOOGLE.COM
Michael Matena MMATENA@GOOGLE.COM
Yanqi Zhou YANQIZ@GOOGLE.COM
Wei Li MWEILI@GOOGLE.COM
Peter J. Liu PETERJLIU@GOOGLE.COM
*Google, Mountain View, CA 94043, USA*

## T5 (2019 – Google)

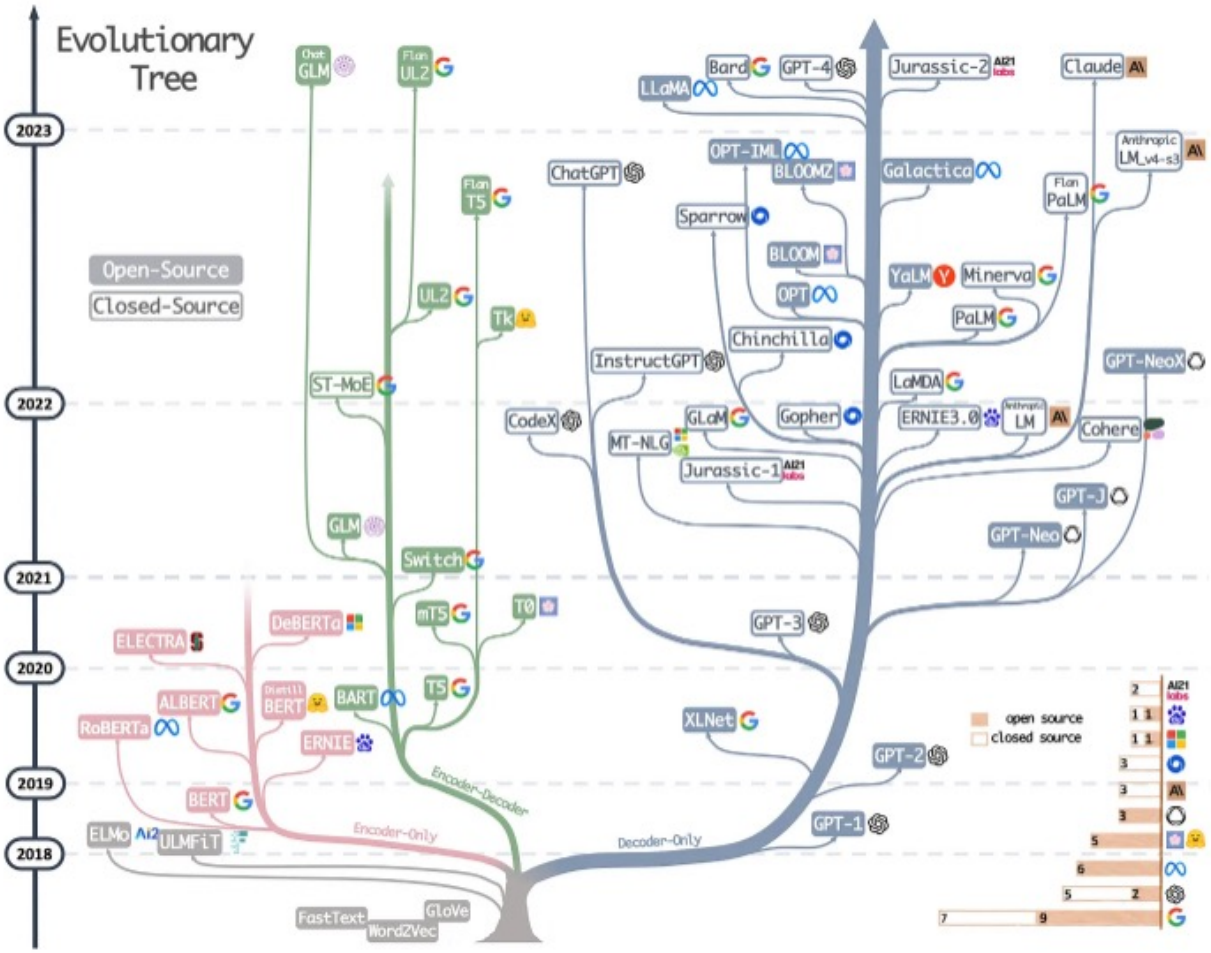
- Performed almost (88.9) at human level (89.8) on the SuperGLUE benchmark set:



SuperGLUE Tasks

| Name | Identifier | Download | More Info | Metric |
|---|---|---|---|---|
| Broadcoverage Diagnostics | AX-b | ⬇ | ↗ | Matthew's Corr |
| CommitmentBank | CB | ⬇ | ↗ | Avg. F1 / Accuracy |
| Choice of Plausible Alternatives | COPA | ⬇ | ↗ | Accuracy |
| Multi-Sentence Reading Comprehension | MultiRC | ⬇ | ↗ | F1a / EM |
| Recognizing Textual Entailment | RTE | ⬇ | ↗ | Accuracy |
| Words in Context | WiC | ⬇ | ↗ | Accuracy |
| The Winograd Schema Challenge | WSC | ⬇ | ↗ | Accuracy |
| BoolQ | BoolQ | ⬇ | ↗ | Accuracy |
| Reading Comprehension with Commonsense Reasoning | ReCoRD | ⬇ | ↗ | F1 / Accuracy |
| Winogender Schema Diagnostics | AX-g | ⬇ | ↗ | Gender Parity / Accuracy |

# The Transformer Family Tree



The evolutionary tree of modern LLMs via https://arxiv.org/abs/2304.13712.

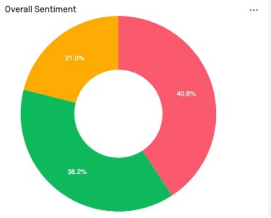
# Most Important NLP Tasks: Classification

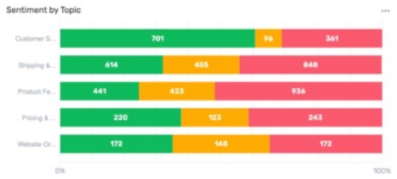
**Sentiment Analysis:** identifying the position of a piece of text in some scale of sentiment.

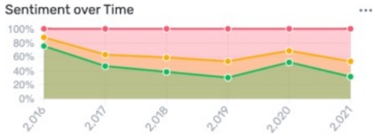
Position may be categorical (2 stars out of 5) or continuous in some range (2.3 on a scale 0 .. 10)

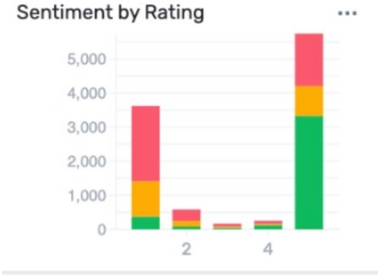
Types of sentiment:

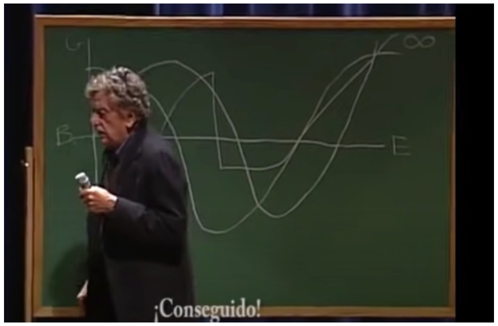
- Positive – Negative
- Aspect or point of view or bias (e.g., political)
- Intent detection
- Emotion Detection
  - Happiness
  - Excited/enthusiastic
  - Frustration or Anger
- Friendship, affection, love or sexual attraction
- Humorous
- Irony
- Hate speech and Fake News detection (next slide)











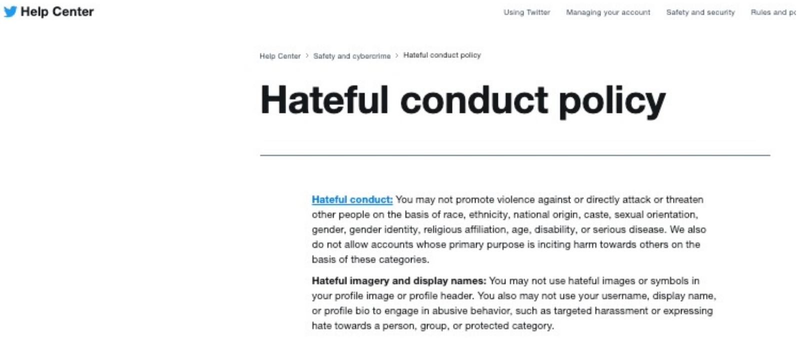
Source: https://monkeylearn.com/sentiment-analysis/

# Fake News and Hate Speech Detection

**Fake News Detection:** detecting and filtering out texts containing false and misleading information.

**Stance Detection:** determining an individual’s reaction to a primary actor’s claim. It is a core part of a set of approaches to fake news assessment.

**Hate Speech Detection:** detecting if a piece of text contains hate speech.

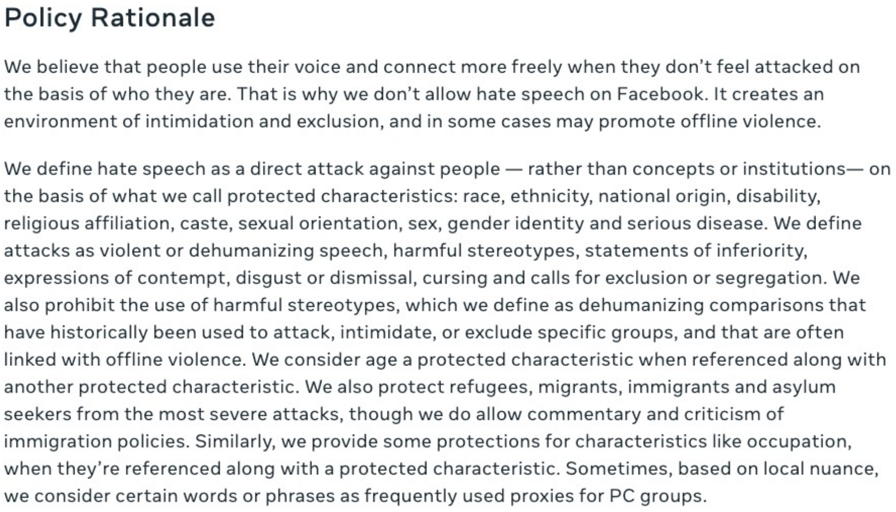


## Hateful conduct policy

Hateful conduct: You may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.

**Hateful imagery and display names:** You may not use hateful images or symbols in your profile image or profile header. You also may not use your username, display name, or profile bio to engage in abusive behavior, such as targeted harassment or expressing hate towards a person, group, or protected category.

Facebook/Meta:



### Policy Rationale

We believe that people use their voice and connect more freely when they don’t feel attacked on the basis of who they are. That is why we don’t allow hate speech on Facebook. It creates an environment of intimidation and exclusion, and in some cases may promote offline violence.

We define hate speech as a direct attack against people — rather than concepts or institutions— on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease. We define attacks as violent or dehumanizing speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing and calls for exclusion or segregation. We also prohibit the use of harmful stereotypes, which we define as dehumanizing comparisons that have historically been used to attack, intimidate, or exclude specific groups, and that are often linked with offline violence. We consider age a protected characteristic when referenced along with another protected characteristic. We also protect refugees, migrants, immigrants and asylum seekers from the most severe attacks, though we do allow commentary and criticism of immigration policies. Similarly, we provide some protections for characteristics like occupation, when they’re referenced along with a protected characteristic. Sometimes, based on local nuance, we consider certain words or phrases as frequently used proxies for PC groups.

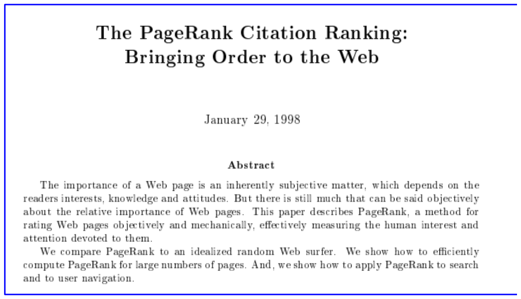
# Information Retrieval

(An old subject, even before Google made it the the most popular text-processing task.)

- Resource Retrieval from text queries/questions
  - Resource could be
    - Highly structured (relational database, code)
    - Semi-structured (Markup Languages (XML), labeled documents)
    - Unstructured (documents)
  - Database search from keywords
  - Google search
  - Backend to Speech to Text systems (siri)
  - Question Answering (next slide)

- Sentence/document similarity: determining how "similar" two texts are
  - Notion of "similar" is variable (similar topic, similar sentiment, ...)
  - Relationship to IR:
    - How similar is text query to a document?
    - "Retrieve more documents similar to this one"
  - Create a map/graph of documents similar to given sentence/document
  - Plagiarism/copyright infringement

- Document Ranking: Rank documents as to some criterion (e.g., PageRank)
  - How well does this document satisfy my query?
  - How important/authoritative is this document?



The PageRank Citation Ranking:
Bringing Order to the Web

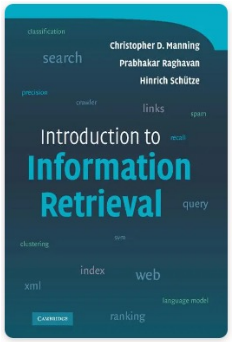
January 29, 1998

**Abstract**

The importance of a Web page is an inherently subjective matter, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them.

We compare PageRank to an idealized random Web surfer. We show how to efficiently compute PageRank for large numbers of pages. And, we show how to apply PageRank to search and to user navigation.

1999



Christopher D. Manning
Prabhakar Raghavan
Hinrich Schütze

Introduction to Information Retrieval

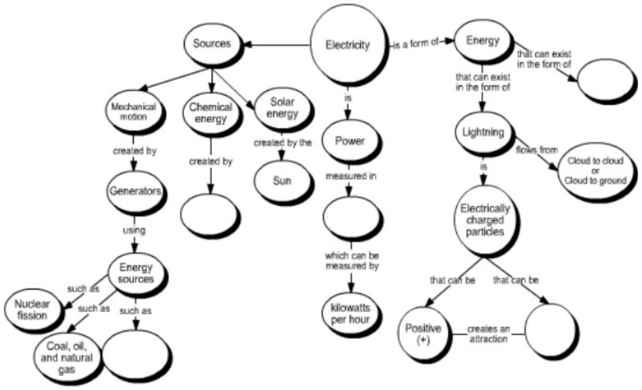
2008

Net worth of Google in 2022: $1.135 Trillion.

# Entities, Relations, and Knowledge Graphs

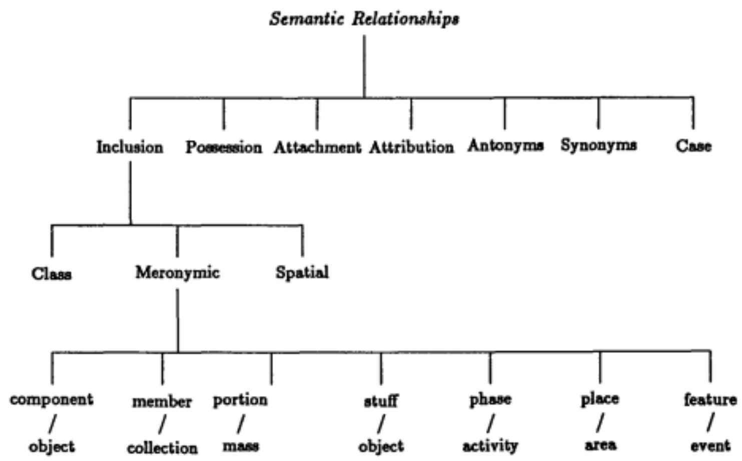
- **Named Entity Recognition:** tagging entities in text with their corresponding
- type, typically in BIO notation.)
- **Coreference Resolution:** clustering mentions in text that refer to the same underlying real-world entities.
- **Relation extraction:** extracting semantic relationships from a text, e.g.,
  - Is-A
  - Has-A
  - Son-Of
  - Part-Of
  - Size-of
  - etc., etc., etc.
- **Build a graph structure:**
  - Knowledge Graph
  - Concept Map
  - Mind Map

Concept Map about Electricity:

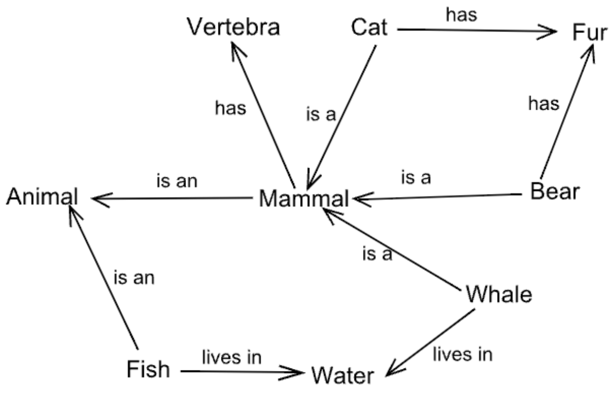


- Graphs can be used to enhance other NLP tasks: search, similarity, question answering, etc
- **Entity Linking:** recognizing and disambiguating named entities to
- a knowledge base (e.g., Wikidata).
- Relation prediction: identifying a named relation between two named
- semantic entities.

# Entities, Relations, and Knowledge Graphs



Based on Landis et al. (1987); Winston et al. (1987); Chaffin et al. (1988).



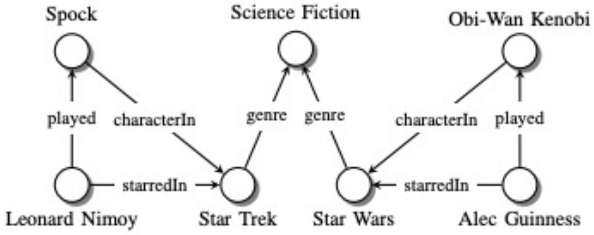
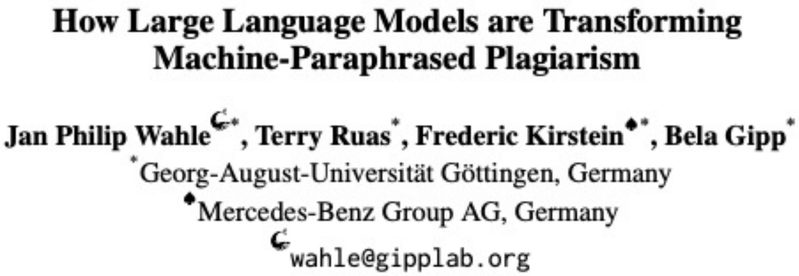


Fig. 1. Sample knowledge graph. Nodes represent entities, edge labels represent types of relations, edges represent existing relationships.

# Text-to-Text Generation

- **Machine Translation:** translating from one language to another.
  - Covered in lecture – Transformer technology transformed this task
- **Text Generation:** creating text from a prompt or subject phrase that appears indistinguishable from human-written text.
  - Covered in lecture – Use language models, Large Language Models (GPT) have transformed this task
- **Lexical Normalization:** translating/transforming a non-standard text to a standard register.
- **Paraphrase Generation:** creating an output sentence that preserves the meaning of input but includes variations in word choice and grammar.
- **Text Simplification:** making a text easier to read and understand, while preserving its main ideas and approximate meaning.
- Text Summarization (next slide)

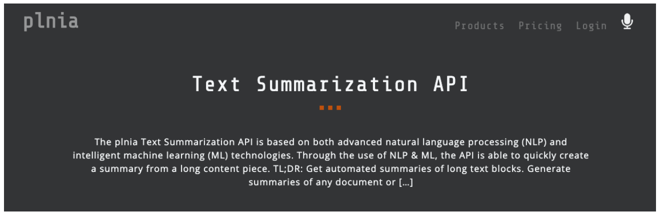


**How Large Language Models are Transforming Machine-Paraphrased Plagiarism**

Jan Philip Wahle[*], Terry Ruas[*], Frederic Kirstein[♦*], Bela Gipp[*]
[*]Georg-August-Universität Göttingen, Germany
[♦]Mercedes-Benz Group AG, Germany
wahle@gipplab.org

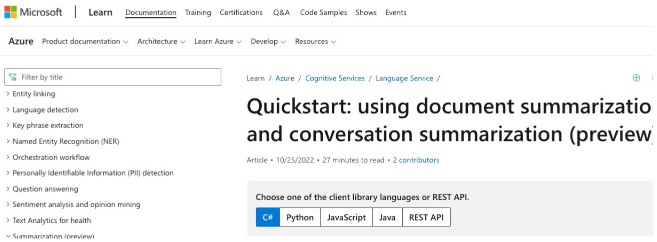
# Topics and Keywords; Text Summarization

- **Topic Modeling:** identifying abstract “topics” underlying a collection of documents.
- **Keyword Extraction:** identifying the most relevant terms to describe the subject of a document
- **Text Summarization:** Reducing size of document while preserving the most important information
  - **Extractive:**
    - Identify the most important sentences in a document and construct the summary from these exact sentences
    - TextRank, LexRank (implements PageRank on sentences in a document)
    - Latent Semantic Analysis (Singular Value Decomposition on a word-sentence matrix)
  - **Abstractive:**
    - Create new text summarizing main points
    - Use of Large Language Models: GPT, BERT, etc...
  - **Use cases for Text Summarization:**
    - Summaries for busy executives or (students!)
    - Summaries of articles, books, chapters
    - Automatic Table of Contents or Indices
    - Downstream from Speech-to-Text systems:
      - Notetaking of meetings, lectures
      - Abstracts of podcasts, YouTube videos
      - Automatic summary of customer phone calls



Text Summarization API

The plnia Text Summarization API is based on both advanced natural language processing (NLP) and intelligent machine learning (ML) technologies. Through the use of NLP & ML, the API is able to quickly create a summary from a long content piece. TL;DR: Get automated summaries of long text blocks. Generate summaries of any document or [...]



Quickstart: using document summarizatio and conversation summarization (preview

Article • 10/25/2022 • 27 minutes to read • 2 contributors

Choose one of the client library languages or REST API.

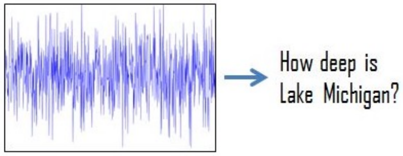
# Chatbots and Question Answering

- **Slot Filling or Cloze Task:** aims to extract the values of certain types of attributes (or slots, such as cities or dates) for a given entity from texts.
- **Chatbots:** Conversation agents (started with Eliza in early 1060's!)
- **Dialog Management:** managing of state and flow of conversations.
- **Question Answering:** Responding to textual queries with textual answers
  - **Extractive QA**: The model extracts the answer from a knowledge source, such as a knowledge graph, database, or document (next slide).
  - **Open Generative QA:** The model generates free text directly based on the (global) context.
  - **Closed Generative QA:** The model generates free text directly based only on the question.

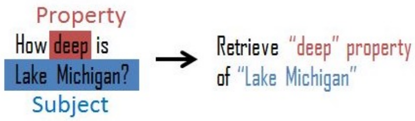
# Question Answering using Knowledge Graphs

Step #1: I randomly shout out "Alexa, how deep is Lake Michigan?"

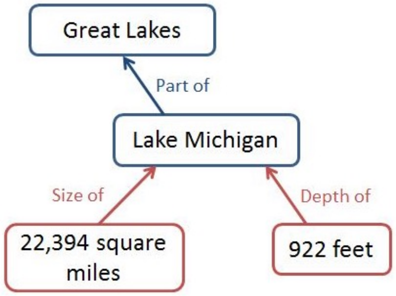
Step #2: Alexa uses voice-to-text processing to parse the noise I made into text.



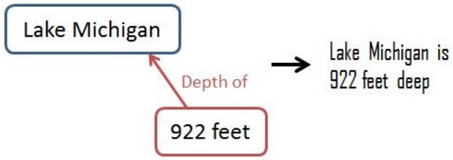
Step #3: Alexa uses Natural Language Processing (NLP) to figure out what I want.



Step #4: Alexa searches its semantic graph for the answer to my question.



Step #5: Alexa uses Natural Language Generation (NLG) to construct a textual answer.

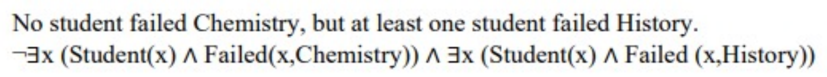


Step #6: Alexa uses text-to-voice processing to calmly blow my mind.

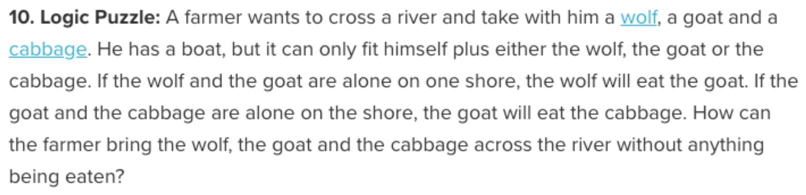
# Reasoning with Text

- Logical Relationship of two sentences/documents:
  - Entailment
  - Temporal sequence
  - Specialization
- Subsystem of text generation at scale

- Text-to/from-First Order Logic: Translate between text and expressions in first-order logic:

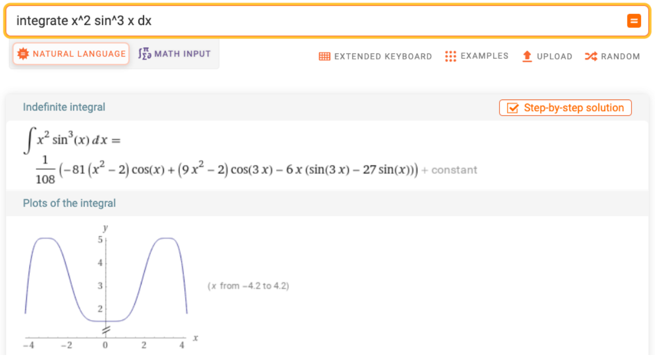


No student failed Chemistry, but at least one student failed History.
¬∃x (Student(x) ∧ Failed(x,Chemistry)) ∧ ∃x (Student(x) ∧ Failed (x,History))



**10. Logic Puzzle:** A farmer wants to cross a river and take with him a wolf, a goat and a cabbage. He has a boat, but it can only fit himself plus either the wolf, the goat or the cabbage. If the wolf and the goat are alone on one shore, the wolf will eat the goat. If the goat and the cabbage are alone on the shore, the goat will eat the cabbage. How can the farmer bring the wolf, the goat and the cabbage across the river without anything being eaten?

- Use cases:
  - Teaching logic
  - Game/puzzle solving
  - Interface to automated theorem prover
    - Prolog
    - Planner
    - Wolfram Alpha



integrate x^2 sin^3 x dx

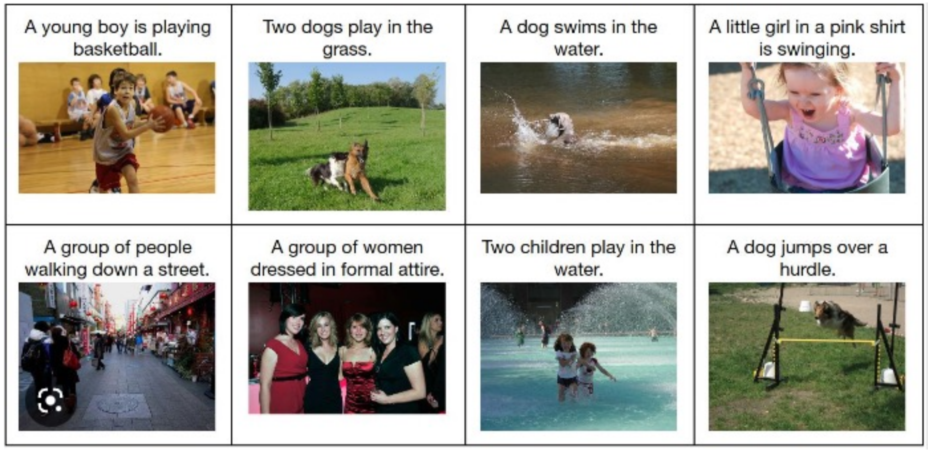
Indefinite integral

$$\int x^2 \sin^3(x)\,dx = \frac{1}{108}\left(-81\left(x^2-2\right)\cos(x)+\left(9x^2-2\right)\cos(3x)-6x\left(\sin(3x)-27\sin(x)\right)\right) + \text{constant}$$

Plots of the integral

# Text-to-Data and Data-to-Text

- Text-to-Image: generating photo-realistic images which are semantically consistent with the text descriptions.
- Image captioning: Generate captions  for input images
- Video-to-Text: Generating text describing a sequence of images
- Text-to-Speech: Human-like reading of input text.
- Speech-to-Text: transcribing speech to text



A young boy is playing basketball.

Two dogs play in the grass.

A dog swims in the water.

A little girl in a pink shirt is swinging.

A group of people walking down a street.

A group of women dressed in formal attire.

Two children play in the water.

A dog jumps over a hurdle.

A marble statue of a Koala DJ in front of a marble statue of a turntable. The Koala has wearing large marble headphones.
An extremely angry bird.
Imagen
A giant cobra snake on a farm. The snake is made out of corn.

An example of some of the images created by Imagen, Google's text-to-image AI generator.