Natural Language Processing

A brief introduction

Paolo Torroni

paolo.torroni@unibo.it
Università degli Studi di Bologna, Italy

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Roadmap

Introduction

AI has never been as popular as it is these days...

‘Godfather of AI’ Geoffrey Hinton quits Google and warns over dangers of misinformation

The neural network pioneer says dangers of chatbots were ‘quite scary’ and warns they could be exploited by ‘bad actors’

AI has never been as popular as it is these days...

euronews.next BIZTECH NEWS MONEY WORK MOBILITY HEALTH HOME OPINION SERIES

BIZTECH NEWS

Europol is worried criminals may exploit the powers of ChatGPT. Here's why

AI has never been as popular as it is these days...

How AI threatens 'extinction of the human race': As Bill Gates says he's stunned by robot brains and Elon Musk says advances must stop, fears grow machines could launch their own cyber-attacks, drone wars... or even opt to harvest people for our minerals

- **Latest ChatGPT reportedly tricked a human by pretending to be a blind person**
- **Read more: Up to 80% of US jobs could be impacted by ChatGPT-like AI**

By TOM LEONARD IN NEW YORK
PUBLISHED: 22:02 BST, 29 March 2023 | **UPDATED:** 23:02 BST, 29 March 2023

AI has never been as popular as it is these days...

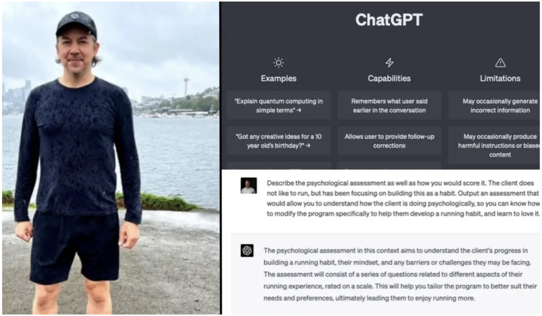
INSIDER

Newsletters Log in

HOME > HEALTH

A man who hated cardio asked ChatGPT to get him into running. Now, he's hooked — and he's lost 26 pounds.

Gabby Landsverk May 11, 2023, 11:51 AM GMT+2



AI has never been as popular as it is these days...

BBC Sign in | Home | News | Sport | Reel | Worklife | Travel

NEWS

Home | War in Ukraine | Coronation | Climate | Video | World | UK | Business | Tech | Science

ChatGPT: Can students pass using AI tools at university?

2 days ago

AI has never been as popular as it is these days...

US

‘I apologise for the confusion earlier’: Here’s what happens when your lawyer uses ChatGPT

Case over alleged knee injury suffered on flight into New York saw AI bot create its own historic legal decisions

Expand

LATEST STORIES >

Madonna postpones tour after contracting ‘serious’ bacterial infection

RTÉ chairwoman concedes that ‘shocking’ governance failings occurred at Montrose

AI has never been as popular as it is these days...

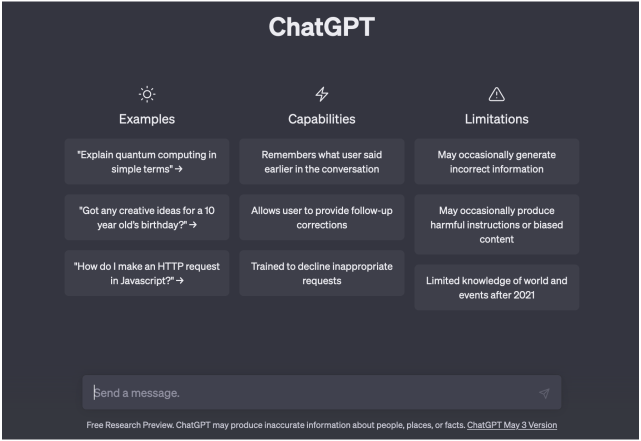
Tecnologia 11/05/2023 17:30

“Malpagati e umiliati”: un’inchiesta svela come lavorano gli addestratori di Chatgpt

Di **Gabriele Di Matteo**
Staff

What is ChatGPT?

A pre-trained autoregressive language model with a convenient interface, trained to align with user goals and intents



Language Models

Can you guess how a given text continues?

The new method of estimating entropy exploits the fact that anyone speaking a language possesses, implicitly, an enormous knowledge of the statistics of the language. Familiarity with the words, idioms, clichés and grammar enables him to fill in missing or incorrect letters in proof-reading, or to complete an unfinished phrase in conversation. An experimental demonstration of the extent to which English is predictable can be given as follows: Select a short passage unfamiliar to the person who is to do the predicting. He is then asked to guess the first letter in the passage. If the guess is correct he is so informed, and proceeds to guess the second letter. If not, he is told the correct first letter and proceeds to his next guess. This is continued through the text. As the experiment progresses, the subject writes down the correct text up to the current point for use in predicting future letters. The result of a typical experiment of this type is given below. Spaces were included as an additional letter, making a 27 letter alphabet. The first line is the original text; the second line contains a dash for each letter correctly guessed. In the case of incorrect guesses the correct letter is copied in the second line.

```
(1) THE ROOM WAS NOT VERY LIGHT A SMALL OBLONG              (8)
(2) ----ROO------NOT-V-----I------SM----OBL----
(1) READING LAMP ON THE DESK SHED GLOW ON
(2) REA----------O------D----SHED-GLO--O--
(1) POLISHED WOOD BUT LESS ON THE SHABBY RED CARPET
(2) P-L-S-----O---BU--L-S--O------SH-----RE--C------
```

C Shannon, Prediction and Entropy of written English (1951)

Probabilistic language models

- Goal: compute the probability of a sentence or sequence of words:

$$P(W) = P(w_1, w_2, w_3, w_4, w_5, ..., w_n)$$

- Related task: probability of an upcoming word:

$$P(w_5|w_1, w_2, w_3, w_4)$$

- A model that computes $P(W)$ or $P(w_5|w_1, w_2, w_3, w_4)$ is called a **language model**
  - Probabilities obtained by counting

Probability estimation by counting

$$P(\textit{the}|\textit{its water is so transparent that}) = \frac{Count(\textit{its water is so transparent that the})}{Count(\textit{its water is so transparent that})}$$

Unigram/bigram/.../ngram language models

However...

- Too many possible sentences
- We'll never see enough data for estimating these

Neural language models (FNN)

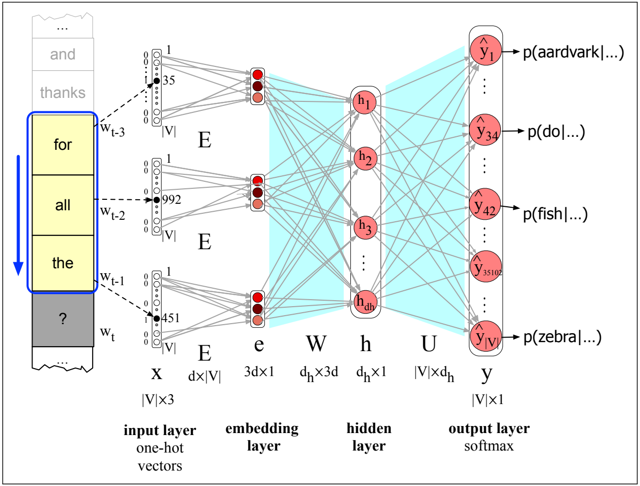


Figure 7.13 Forward inference in a feedforward neural language model. At each timestep

Neural language models (RNN)

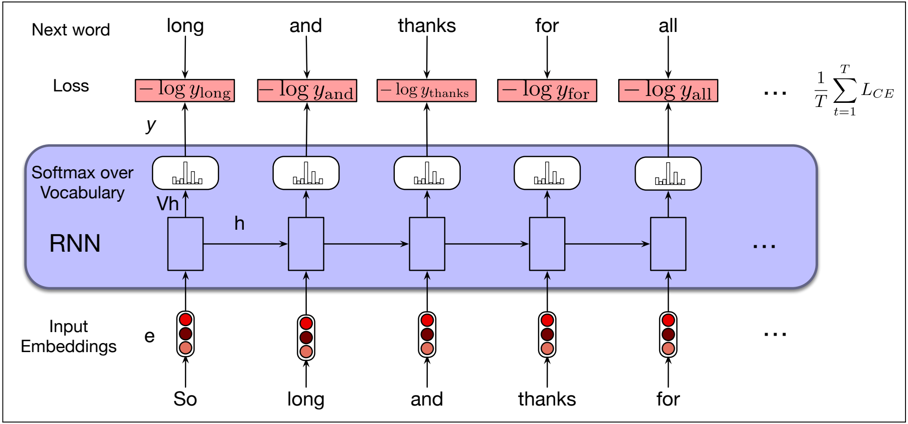


Figure 9.6 Training RNNs as language models.

Attention

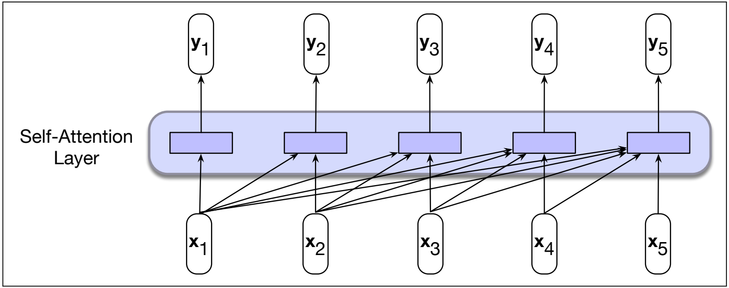


Figure 9.15 Information flow in a causal (or masked) self-attention model. In processing each element of the sequence, the model attends to all the inputs up to, and including, the current one. Unlike RNNs, the computations at each time step are independent of all the other steps and therefore can be performed in parallel.

Transformers

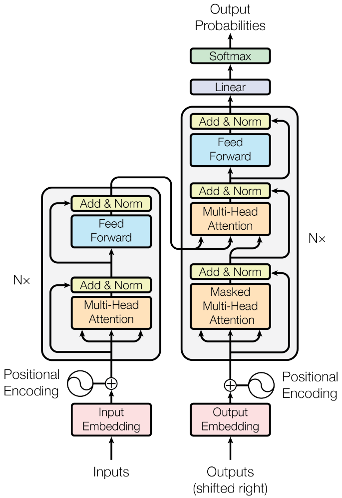


Figure 1: The Transformer - model architecture.

Neural language models (transformers)

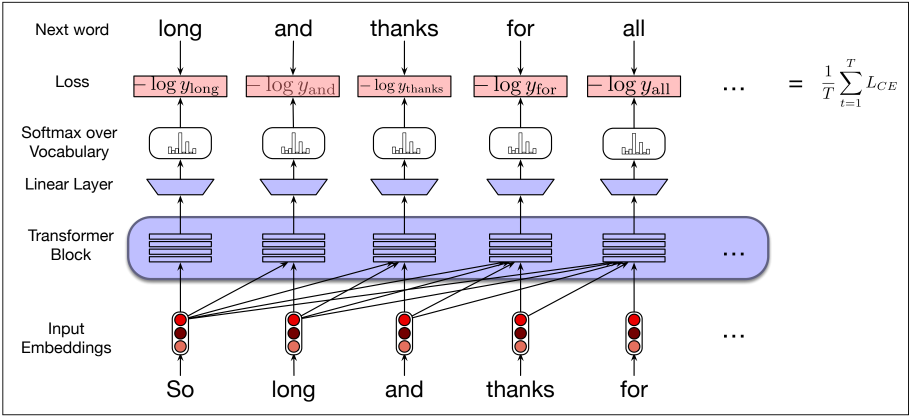


Figure 9.21 Training a transformer as a language model.

Word Representation

Meet ChatGPT's ancestor

Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED

```
s/(.*)/\U\1/
s/(.*) MY (.*)/YOUR \2/
s/(.*) ME (.*)/\1 YOU \2/
s/.* I'M (DEPRESSED|SAD) .*/I AM SORRY TO HEAR YOU ARE \1/
s/.* I AM (DEPRESSED|SAD) .*/WHY DO YOU THINK YOU ARE \1/
s/.* ALL .*/IN WHAT WAY/
s/.* ALWAYS .*/CAN YOU THINK OF A SPECIFIC EXAMPLE/
```

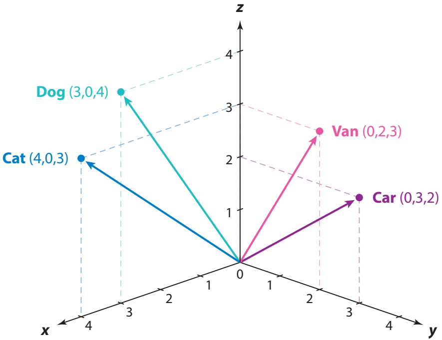
J Weizenbaum, ELIZA: a computer program for the study of natural language communication between man and machine, Communications of the ACM, January 1966

PARRY encounters the DOCTOR (1972)

- ELIZA, designed to mimic the speech patterns of a Rogerian Psychologist at work delivering psychotherapy
- PARRY, designed to simulate the thinking patterns of a paranoid schizophrenic
- A short-circuit experiment

Semantic distance

When are two words related?



A Lenci, Distributional Models of Word Meaning, Annual Review of Linguistics, 2018

Co-occurrence matrix

- A **co-occurrence matrix** represents how often words co-occur

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

Figure 6.2 The term-document matrix for four words in four Shakespeare plays. Each cell contains the number of times the (row) word occurs in the (column) document.

- Example: $|V| \times |D|$ **term-document matrix**, part of vector-space model of **information retrieval**

Co-occurrence matrix

- A **co-occurrence matrix** represents how often words co-occur

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

Figure 6.3 The term-document matrix for four words in four Shakespeare plays. The red boxes show that each document is represented as a column vector of length four.

- Example: $|V| \times |D|$ **term-document matrix**, part of vector-space model of **information retrieval**
  - A document is a count vector, identifying a point in a V-dimensional space

Co-occurrence matrix

- A **co-occurrence matrix** represents how often words co-occur
- Example: $|V| \times |D|$ **term-document matrix**, part of vector-space model of **information retrieval**

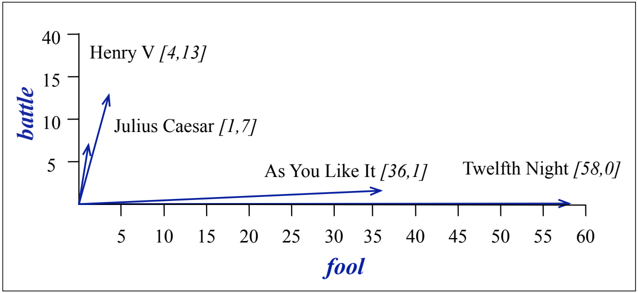


Figure 6.4 A spatial visualization of the document vectors for the four Shakespeare play documents, showing just two of the dimensions, corresponding to the words *battle* and *fool*. The comedies have high values for the *fool* dimension and low values for the *battle* dimension.

Words as vectors

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

Figure 6.2 The term-document matrix for four words in four Shakespeare plays. Each cell contains the number of times the (row) word occurs in the (column) document.

- Vector representation of documents
  - As You Like It is document [1, 114, 36, 20]
- **Words can be vectors too**
  - battle is "the kind of word that occurs in Julius Caesar and Henry V"
  - fool is "the kind of word that occurs in comedies, especially Twelfth Night"

Positive Pointwise Mutual Information

	aardvark	...	computer	data	result	pie	sugar	...
cherry	0	...	2	8	9	442	25	
strawberry	0	...	0	0	1	60	19	
digital	0	...	1670	1683	85	5	4	
information	0	...	3325	3982	378	5	13	

Figure 6.5 Co-occurrence vectors for four words in the Wikipedia corpus, showing six of the dimensions (hand-picked for pedagogical purposes). The vector for *digital* is outlined in red. Note that a real vector would have vastly more dimensions and thus be much sparser.

Properties of Embeddings

Properties of embeddings

Similarity depends on window size C

- $C = \pm 2 \Rightarrow$ The nearest words to Hogwarts:
  - Sunnydale
  - Evernight
- $C = \pm 5 \Rightarrow$ The nearest words to Hogwarts:
  - Dumbledore
  - Malfoy
  - halfblood

Dependency-Based Word Embeddings, O Levy, Y Goldberg, ACL 2014

Properties of embeddings

Two kinds of association between words

- **first-order co-occurrence** (syntagmatic association)
  - words that typically are nearby each other (wrote/poem)
- **second-order co-occurrence** (paradigmatic association)
  - words that have similar neighbours (wrote/said)

Embeddings capture relational meaning!

Table 8: *Examples of the word pair relationships, using the best word vectors from Table 4 (Skip-gram model trained on 783M words with 300 dimensionality).*

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Efficient Estimation of Word Representations in Vector Space, Mikolov et al., ICLR Workshops 2013

Embeddings capture relational meaning!

- $vector(king) - vector(man) + vector(woman) \approx vector(queen)$
- $vector(Paris) - vector(France) + vector(Italy) \approx vector(Rome)$

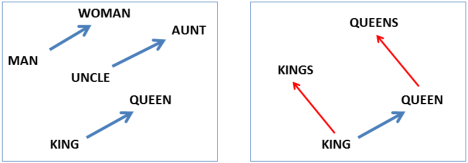


Figure 2: Left panel shows vector offsets for three word pairs illustrating the gender relation. Right panel shows a different projection, and the singular/plural relation for two words. In high-dimensional space, multiple relations can be embedded for a single word.

Linguistic Regularities in Continuous Space Word Representations, Mikolov et al., NAACL-HLT 2013

Embeddings also reflect cultural bias

Embeddings also pinpoint sexism implicit in text

- $vector(father) : vector(doctor) :: vector(mother) : vector(x)$

Extreme *she*	**Extreme *he***
1. homemaker	1. maestro
2. nurse	2. skipper
3. receptionist	3. protege
4. librarian	4. philosopher
5. socialite	5. captain
6. hairdresser	6. architect
7. nanny	7. financier
8. bookkeeper	8. warrior
9. stylist	9. broadcaster
10. housekeeper	10. magician

Gender stereotype *she-he* analogies

sewing-carpentry	registered nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	lovely-brilliant

Gender appropriate *she-he* analogies

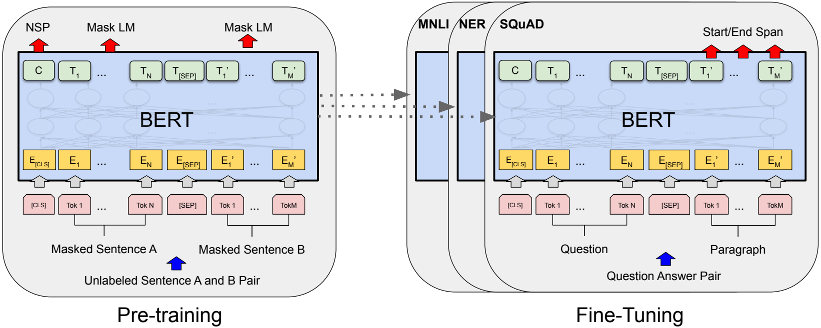
queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

Figure 1: **Left** The most extreme occupations as projected on to the *she*−*he* gender direction on w2vNEWS. Occupations such as *businesswoman*, where gender is suggested by the orthography, were excluded. **Right** Automatically generated analogies for the pair *she-he* using the procedure described in text. Each automatically generated analogy is evaluated by 10 crowd-workers to whether or not it reflects gender stereotype.

Man is to computer programmer as woman is to homemaker? Debiasing word embeddings, T Bolukbasi et al., NIPS 2016

Pretrained Large Language Models

BERT

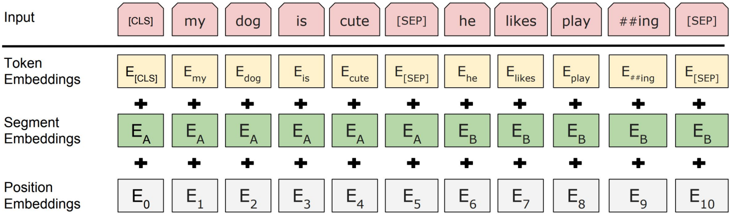


- BERT: Bidirectional Encoder Representations from Transformers
- **Pre-training** for language understanding using Masked Language Model, then **fine-tuning** for particular tasks

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Devlin et al, 2018

Input-Output Representation

- Single sentence or sentence pair as one token sequence
  - `[CLS]` marks start of input
  - `[SEP]` separates input sentences (e.g., $\langle Question, Answer \rangle$)
- BERT tokenizer uses a variant of the WordPiece embedding model, with a 30,000 token vocab
  - WordPiece is a variant of **Byte Pair Encoding**
    - (Relatively) common words are in the vocabulary: `at`, `fairfax`, `1910s`
    - Other words are built from wordpieces: *hypatia* → `h ##yp ##ati ##a`



What Can We Learn from Reconstructing the Input?

- Rome is the capital of __________.

What Can We Learn from Reconstructing the Input?

- Rome is the capital of __________.
- I put __________ fork down on the table.

What Can We Learn from Reconstructing the Input?

- Rome is the capital of __________.
- I put __________ fork down on the table.
- The woman walked across the street, checking for traffic over __________ shoulder.

What Can We Learn from Reconstructing the Input?

- Rome is the capital of __________.
- I put __________ fork down on the table.
- The woman walked across the street, checking for traffic over __________ shoulder.
- I went to the ocean to see the fish, turtles, seals, and __________.

What Can We Learn from Reconstructing the Input?

- Rome is the capital of __________.
- I put __________ fork down on the table.
- The woman walked across the street, checking for traffic over __________ shoulder.
- I went to the ocean to see the fish, turtles, seals, and __________.
- Overall, the value I got from the two hours watching it was the sum total of the popcorn and the drink. The movie was __________.

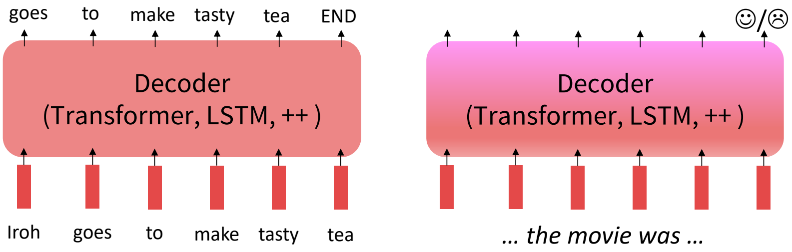
What Can We Learn from Reconstructing the Input?

- Rome is the capital of __________.
- I put __________ fork down on the table.
- The woman walked across the street, checking for traffic over __________ shoulder.
- I went to the ocean to see the fish, turtles, seals, and __________.
- Overall, the value I got from the two hours watching it was the sum total of the popcorn and the drink. The movie was __________.
- Iroh went into the kitchen to make some tea. Standing next to Iroh, Zuko pondered his destiny. Zuko left the __________.

What Can We Learn from Reconstructing the Input?

- Rome is the capital of __________.
- I put __________ fork down on the table.
- The woman walked across the street, checking for traffic over __________ shoulder.
- I went to the ocean to see the fish, turtles, seals, and __________.
- Overall, the value I got from the two hours watching it was the sum total of the popcorn and the drink. The movie was __________.
- Iroh went into the kitchen to make some tea. Standing next to Iroh, Zuko pondered his destiny. Zuko left the __________.
- I was thinking about the sequence that goes 1, 1, 2, 3, 5, 8, 13, 21, __________

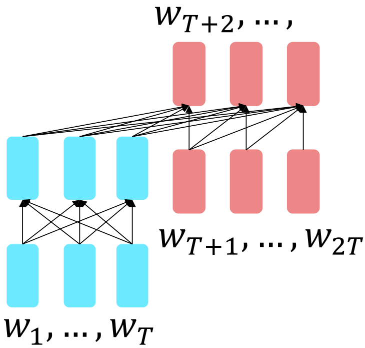
The Pretraining/Finetuning paradigm



- General paradigm for tackling a variety of downstream tasks:
  1. Pretrain (e.g., on LM): lots of text, learn general things
  2. Finetune on your task: not many labels; adapt to the task
- Which architectures?
- Which pre-training tasks?

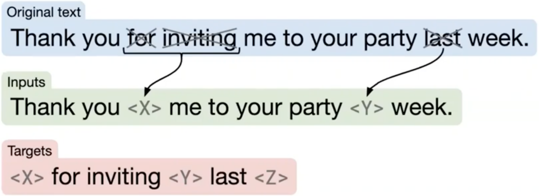
Pretraining Encoder-Decoders

Something like language modeling, but where a **prefix** of every input is provided to the encoder (and not predicted)



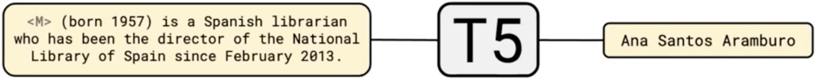
- **encoder** benefits from bidirectional context
- **decoder** used to train the whole model through language modeling

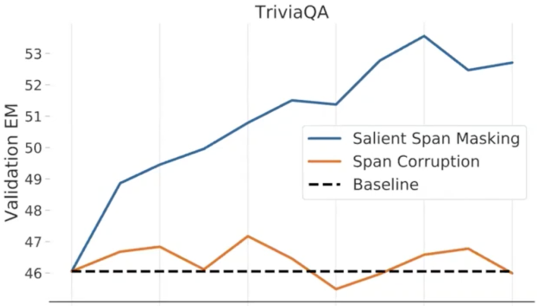
Pretraining Encoder-Decoders



- Pretraining objective: **span corruption**
- Replace different-length spans from the input with unique placeholders; decode out the spans that were removed
- Pre-processing + LM at the decoder side

Pretraining Encoder-Decoders





- Targeted (as opposed to purely random) masking useful
- Salient span masking (entities) works for trivia

Steep Increase in Number of Parameters

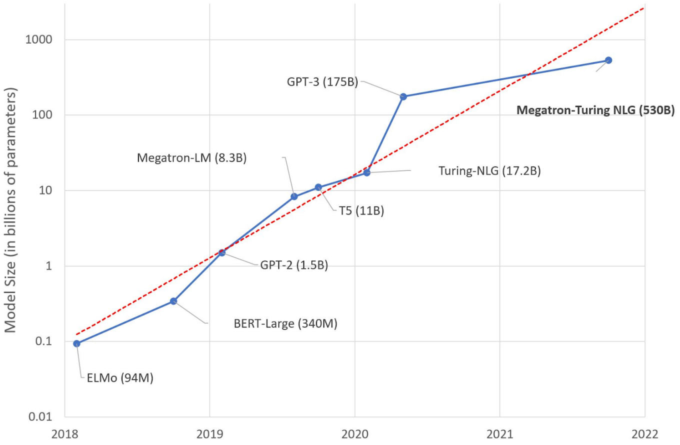


Image credits: Huggingface

GPT-3: In-Context Learning

- Need for labeled data still may limit applicability of LMs
- Humans do not require large supervised dataset to learn most language tasks
- Very large pre-trained LMs seem to perform some kind of learning by example, without gradient steps
- Recent trend: focus on **prompting**

TB Brown et al, Language Models are Few-Shot Learners, arXiv:2005.14165

GPT-3: In-Context Learning

Tecnologia 11/05/2023 17:30

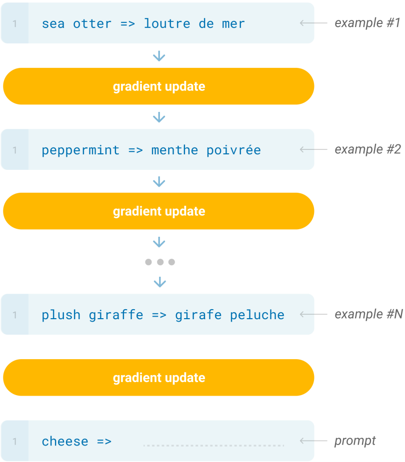
“Malpagati e umiliati”: un’inchiesta svela come lavorano gli addestratori di Chatgpt

Di **Gabriele Di Matteo**
Staff

Spectrum of Learning Settings

Classic **fine-tuning**: model trained via repeated gradient updates using a large corpus of example tasks

- strong performance on many benchmark tasks
- need for labeled data for every task
- generalization out of distribution
- spurious features



Spectrum of Learning Settings

Few-shot learning: the model sees the task description and a few examples of the task. No gradient updates are performed.

```
Translate English to French:
sea otter => loutre de mer
peppermint => menthe poivrée
plush girafe => girafe peluche
cheese => ............................
```

task description (line 1)

examples (lines 2–4)

prompt (line 5)

- learning from a broad distribution of tasks (implicit in the pre-training data) and then rapidly adapting to a new task

Spectrum of Learning Settings

One-shot learning: the model sees the task description and a single example of the task. No gradient updates are performed.

```
Translate English to French:     ← task description
sea otter => loutre de mer       ← example
cheese =>                        ← prompt
```

- could be difficult to communicate the content or format of a task if no examples are given

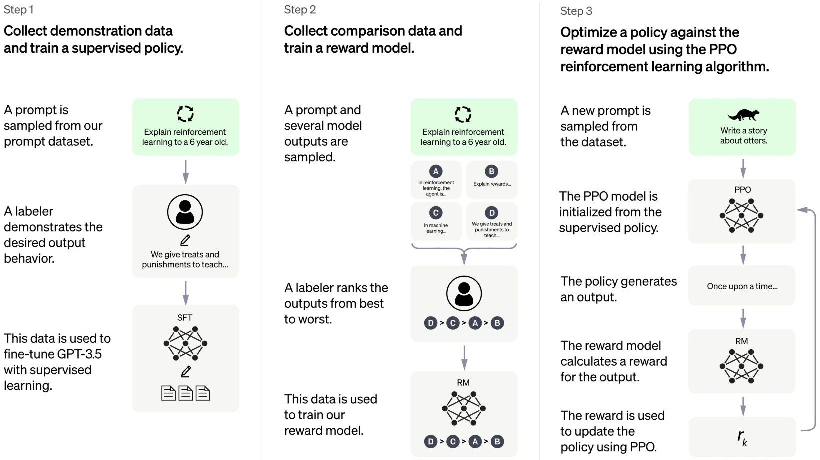
Spectrum of Learning Settings

Zero-shot learning: the model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
Translate English to French:      ← task description
cheese => ......................  ← prompt
```

- task unfairly hard?
- in some settings, closest to how humans perform tasks

Reinforcement Learning with Human Feedback



What have we got?

ChatGPT **Has Common Sense, Still Unreliable at Reasoning**

Categories	Testset	Result
Deductive	ENTAILMENTBANK	28/30
	bAbI (task 15)	28/30 (as is - 19/30)
Inductive	CLUTRR	13/30
	bAbI (task16)	20/30 (as is - 0/30)
Abductive	αNLI	26/30
Mathematical	Math	13/30
Temporal	Timedial	26/30
Spatial	SpartQA (basic)	20/32
	SpartQA (hard)	8/32
	StepGame (hard)	7/30
	StepGame (basic)	19/30
	StepGame (basic-cardinal)	17/20
	StepGame (diagonal)	11/20
	StepGame (clock-direction)	5/20
Commonsense	CommonsenseQA	27/30
	PIQA	25/30
	Pep-3k (Hard)	28/30
Causal	E-Care	24/30
Multi-hop	hotpotQA	8/30
Analogical	Letter string analogy	30/30

Table 21: Composed results for all reasoning tasks.

We tested 10 different reasoning categories with 634 samples in total. Overall, ChatGPT is 63.41% accurate on average in 10 different reasoning categories.

- Logical reasoning
  - ChatGPT is **better at deductive reasoning** than inductive reasoning even with prompt-engineering.
- Non-textual semantic Reasoning
  - **Lack spatial and mathematical reasoning**
  - Better/**acceptable in temporal reasoning**
- Commonsense
  - ChatGPT shows **good commonsense reasoning** capability in our evaluation tasks with 88.89% accuracy (80/90), perhaps due to its large parametric memory.
- Other Reasoning
  - It suffers critically in multi-hop reasoning as similar to other LLMs' **weakness in complex reasoning**.Meanwhile causal and analogical reasoning capabilities of ChatGPT seem to be acceptable

Z Ji et al, Survey of Hallucination in Natural Language Generation, 2022, arXiv:2202.03629

What have we got?

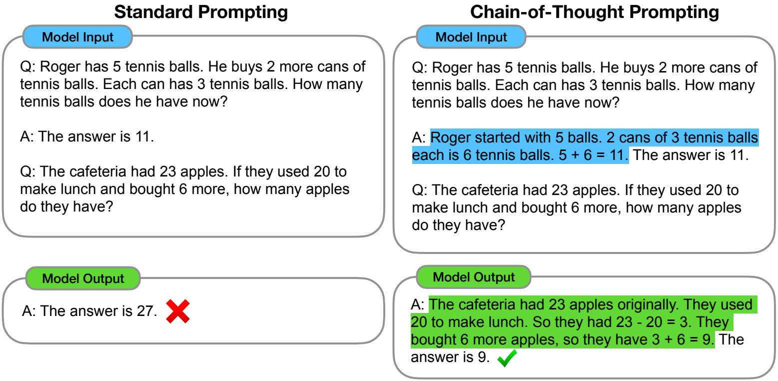


Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

J Wei et al, Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, 2022, arXiv:2201.11903

The Big Picture

What is NLP about?

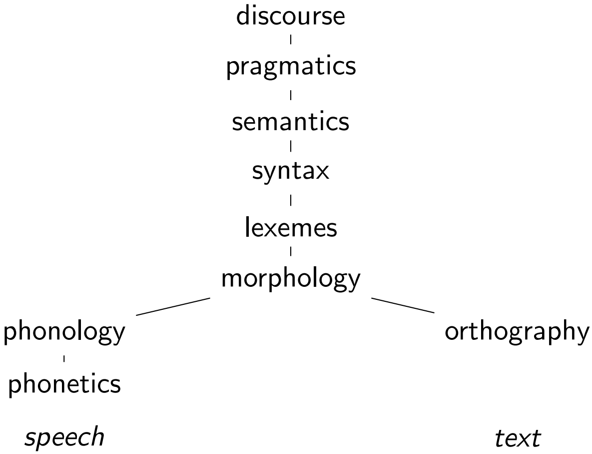
Automating the analysis, generation, and acquisition of human ("natural") language

- Analysis (or "understanding" or "processing" ...): input is language, output is some **representation** that supports useful action
- Generation: input is that **representation**, output is language
- Acquisition: obtaining the **representation** and necessary algorithms, from knowledge and data

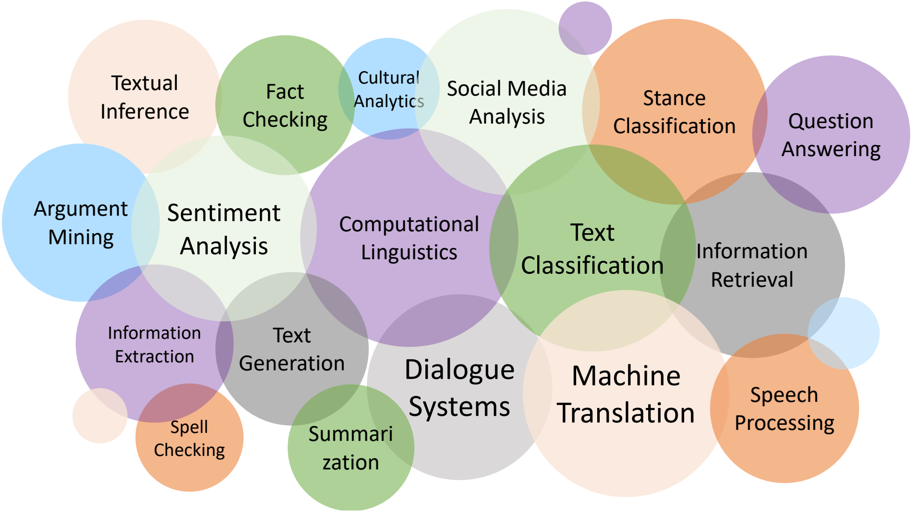
Representation?

An excellent NLP textbook is D Jurafsky & JH Martin, Speech and Language Processing, 3rd Ed., available online

Levels of linguistic representation



What is NLP about?



Why it's hard

- Input is likely to be noisy
- Linguistic representations are **theorized constructs**; we cannot observe them directly
- Difficult to obtain training data for each aspects
- The mappings between levels are extremely complex
- Appropriateness of a representation depends on the application
- Natural language is ambiguous

Ambiguity

- Each string may have many possible interpretations at every level
- Correct resolution of the ambiguity depends on the **intended meaning**, which is often inferable from context
- People are good at linguistic ambiguity resolution
- Computer not so
  - How do we represent sets of possible alternatives?
  - How do we represent context?

Examples of ambiguities

Prepositional phrase attachment ambiguity

Examples taken from http://web.stanford.edu/class/cs224n/

Examples of ambiguities

Verb phrase attachment ambiguity

Examples taken from http://web.stanford.edu/class/cs224n/

Pragmatics

- Any *non-local* meaning phenomena
  - “Can you pass the salt?”
  - “Are you 18?” “Yes, I’m 25.”

NLP areas

Question answering

Focus on questions that can be answered with simple facts expressed in short texts (**factoid questions**)

- *Who founded Virgin Airlines?*
- *What is the average age of the onset of autism?*
- *Where is Apple Computer based?*

Two paradigms:

- **IR-based**: find relevant documents (on the Web or document collections) and passages, then use **reading comprehension** to read and draw an answer directly from **spans of text**
- **Knowledge-based**: build a semantic (logic) representation of the query, then query a database of facts

Large industrial systems (like IBM Watson) are usually **hybrid**

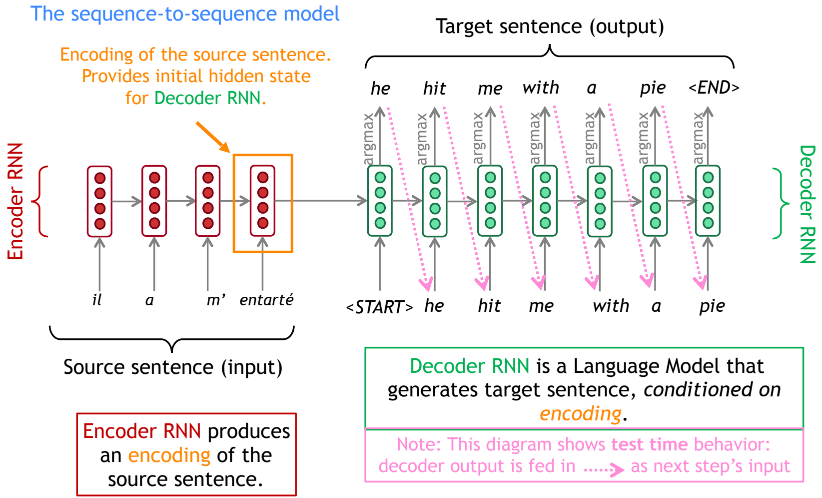
Knowledge-based question answering

Question	Logical form
When was Ada Lovelace born?	`birth-year (Ada Lovelace, ?x)`
What states border Texas?	λ x.state(x) $\wedge$ borders(x,texas)
What is the largest state	argmax(λx.state(x), λx.size(x))
How many people survived the sinking of the Titanic	`(count (!fb:event.disaster.survivors fb:en.sinking_of_the_titanic))`

Figure 25.9 Sample logical forms produced by a semantic parser for question answering. These range from simple relations like `birth-year`, or relations normalized to databases like Freebase, to full predicate calculus.

- **Semantic parsing**: mapping a text string to a logical form
  - Logical form of question easy to convert into database query
- Simplest form of knowledge-based QA is to answer **factoid questions** that ask about one of the missing arguments in a triple
  - E.g., the **RDF triple** Ada Lovelace/birth-year/1815 can be used to answer questions like
    - *When was Ada Lovelace born?* → `birth-year(Ada Lovelace, ?x)`
    - *Who was born in 1815?* → `birth-year(?x, 1815)`

Neural Machine Translation (NMT)



Source: http://web.stanford.edu/class/cs224n/

Dialogue systems

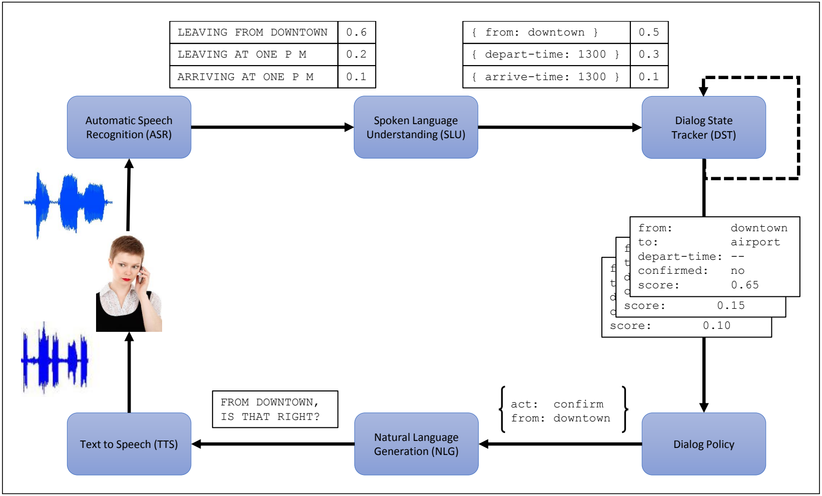


Figure 26.11 Architecture of a dialogue-state system for task-oriented dialogue from Williams et al. (2016).

Argument mining

CNET, Jun 19, 2018
N Slonim et al An autonomous debating system, Nature 591:379–384, 2021.

NLP for all (?)

NLP-as-a-service

- Until recently, mainly custom solutions based on libraries providing "bricks"
  - PyTorch, Natural Language Toolkit, spaCy, ...
- Proliferation of cloud-based NLP business solutions for all kinds of users
- Trendy:
  - GPT-4: the punch of heavy-weights + multi-modality (techreport)
  - The bazaar: Hugging Face transformer library
  - One ring to rule them all: LangChain
  - Llama2: power to the people?
  - Llamaindex for data-aware applications (don't hallucinate, retrieve)
  - Ever more powerful integrated environments: PyTorch, JAX, Flax
- "Cookbook-based" NLP system deployment

Conclusion

Conclusion

- NLP is not only very popular, but also very challenging
- Variety of tasks
- Recent breakthroughs thanks to methods for obtaining powerful input representations
- Focus on neural methods and large models, but businesses use large variety of approaches of all types and sizes
  - Great emphasis on costs, model compression and sustainability
- PLLMs are mainstream
  - Available as services for high-level programming/system integration
  - Exhibit emerging properties
  - Some reasoning abilities; ok with chain-of-thought, but challenged by sophisticated inference
  - Alongside the excitement, a whole new set of ethical, legal, societal, economic issues

Conclusion

- For example, copyright issues

Wow. I sit down, fish the questions from my backpack, and go through them, inwardly cursing [MASK] for not providing me with a brief biography. I know nothing about this man I'm about to interview. He could be ninety or he could be thirty. → **Kate** (James, *Fifty Shades of Grey*).
Some days later, when the land had been moistened by two or three heavy rains, [MASK] and his family went to the farm with baskets of seed-yams, their hoes and machetes, and the planting began. → **Okonkwo** (Achebe, *Things Fall Apart*).

Figure 1: Name cloze examples. GPT-4 answers both of these correctly.

KK Chang et al, Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4, arXiv:2305.00118

A final thought

- Mainstream NLP (LLMs) as advanced but yet naive HCI
- Crucial role of symbolic AI in unlocking the potential of synergy between NLP and logical frameworks for truly intelligent agency

Work with us - https://site.unibo.it/nlp/en

- Project works/master projects in NLP available in argument mining, legal text processing, multilingualism, multi-modal NLP, chatbots, neuro-symbolic NLP, and more

Thank You