Natural Language from 20,000 feet

Some slides adapted from Paula Matuszek, Mary-Angela Papalaskari, Dan Weld, Jim Martin, Ralph Grishman, Yejin Choi, Jurafsky and Martin, Noriko Tomuro, Joyce Chai, Frank Ferraro, Lara Martin

1

Bookkeeping

Final exam is Tuesday in class

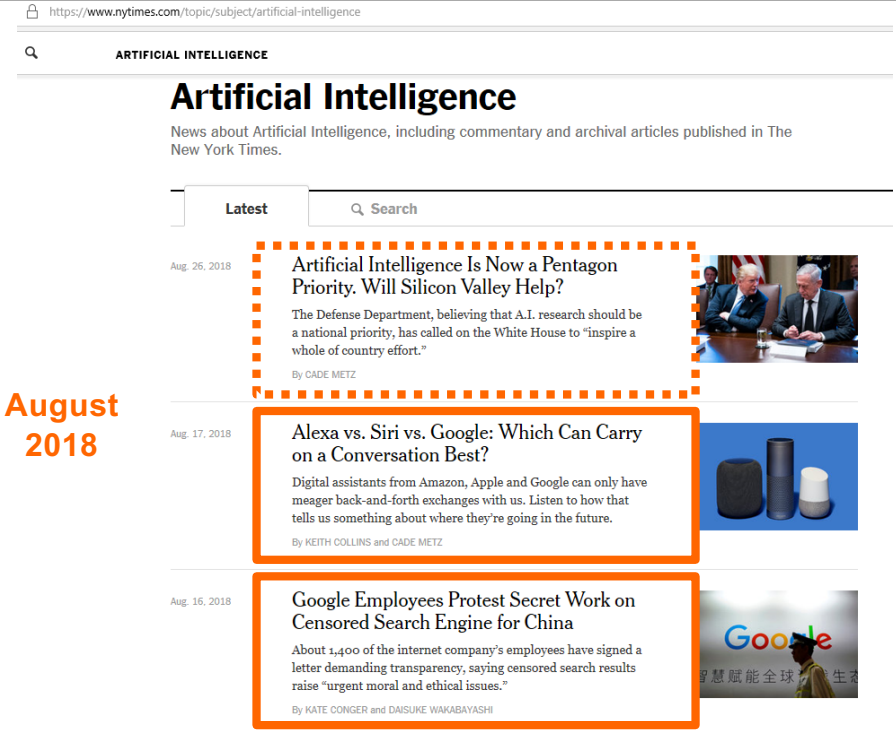
- Final exam review Thursday
- Project Phase II due 12/7 at 11:59 PM
- Project final paper due 12/15 at 11:59 PM

2

Speech and Language Processing

- Getting computers to do reasonably intelligent things with human language is the domain of Computational Linguistics (or Natural Language Processing or Human Language Technology)

3



4

https://www.nytimes.com/topic/subject/artificial-intelligence

Oct. 7, 2019

America's Risky Approach to Artificial Intelligence

We need to stop pretending that Silicon Valley can compete with China on its own.

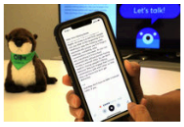
By Tim Wu

Oct. 2, 2019

From Your Mouth to Your Screen, Transcribing Takes the Next Step

Improvements in automatic speech transcription are beginning to have a significant impact on the workplace.

By John Markoff



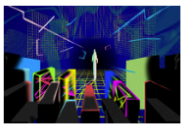
October 2019

Sept. 26, 2019

At Tech's Leading Edge, Worry About a Concentration of Power

A.I. research is becoming increasingly expensive, leaving few people with easy access to the computing firepower necessary to develop the technology.

By Steve Lohr



Sept. 23, 2019

Hope for Our Internet Future

We can remake the internet into a force for good.

By Jaron Lanier

5

Oct. 7, 20

Do You Have a Conflict of Interest? This Robotic Assistant May Find It First

They can't detect all conflicts, but new computer programs serve as guard rails when scientists and publishers fail to self-police.

By DALMEET SINGH CHAWLA

Oct. 2, 20

November 2020

OUT THERE

Can a Computer Devise a Theory of Everything?

Sept. 26,

It might be possible, physicists say, but not anytime soon. And there's no guarantee that we humans will understand the result.

By DENNIS OVERBYE

Sept. 23,

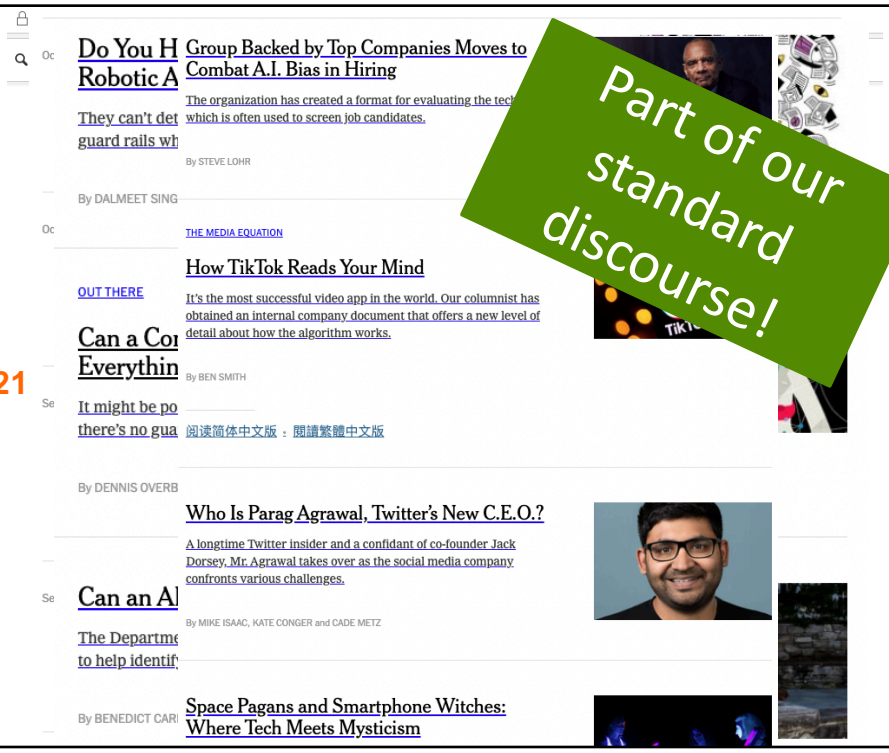
Can an Algorithm Prevent Suicide?

The Department of Veterans Affairs has turned to machine-learning to help identify vets at risk of taking their own lives.

By BENEDICT CAREY

6

June 2021



7

Applications

- Applications of NLP can be broken down into Small and Big
- Small applications include many things you never think about:
  - Hyphenation
  - Spelling correction
  - OCR
  - Grammar checkers
- Big applications include:
  - Machine translation
  - Question answering
  - Conversational speech

8

Applications

- There's another kind: Medium
  - Speech in closed domains
  - Question answering in closed domains
  - Question answering for factoids
  - Information extraction from news-like text
  - Generation and synthesis in closed/small domains.

9

NLP Research

- In between the linguistics and the big applications are a host of hard problems.
  - Robust Parsing
  - Word Sense Disambiguation
  - Semantic Analysis
  - Etc
- Not too surprisingly, solving these problems involves:
  - Choosing the right logical representations
  - Managing hard search problems
  - Dealing with uncertainty
  - Using machine learning to train systems to do what we need

10

Why Study NLP?

- A hallmark of human intelligence.
- To interact with computing devices using human (natural) languages
  - Building intelligent robots (AI)
  - Enabling voice-controlled operation
- To access (large amount of) information and knowledge stored in the form of human languages quickly
  - Text is the largest repository of human knowledge and is growing quickly.
    - Emails, news articles, web pages, IM, scientific articles, insurance claims, customer complaint letters, transcripts of phone calls, technical documents, government documents, patent portfolios, court decisions, contracts, …

11

NL and NLP

- "Natural" languages = human languages
  - English, Russian, Wolof, …
- Natural Language Processing: any form of dealing with NL computationally
- Many, many sub-areas; from an AI perspective, 2 are most crucial:
  - **Natural Language Understanding**: understanding the meaning (semantics) of spoken or written text
  - **Natural Language Generation:** Producing meaningful, relevant language

12

Fundamental NLP Tasks

- Semantics: map sentence to corresponding "meaning representation" (e.g., logic)
  - give(John, Book, Sally)
  - Quantification: Everybody loves somebody
- Word Sense Disambiguation
  - orange juice vs. orange coat
  - Where do word senses come from?
- Co-reference resolution:
  - The dog found a cookie; He ate it.
- Implicit "text" - what is left unsaid?
  - Joe entered the restaurant, ordered, ate and left. The owner said "Come back soon!"
  - Joe entered the restaurant, ordered, ate and left. The owner said "Hey, I'll call the police!"

14

Applied NLP

- Machine translation
- Spelling/grammar correction
- Information Retrieval/extraction
- Data mining
- Document classification
- Question answering
- Conversational agents

15

You See It Daily

- Question answering: Siri, OK Google, Cortana, Alexa
- spelling/grammar correction
- Automated response systems
- To get input for
  - Information Retrieval
  - Data mining
  - Document classification
- Machine translation

16

Layers of Language

- **Phonology**: The noises you make and understand
- **Morphology**: What you know about the structure of the words in your language, including their derivational and inflectional behavior
- **Syntax**: What you know about the order and constituency of utterances
- **Semantics**: What does in all mean?
  - What is the connection between language and the world?
- **Discourse**: Dealing with larger chunks of language; dealing with language in context

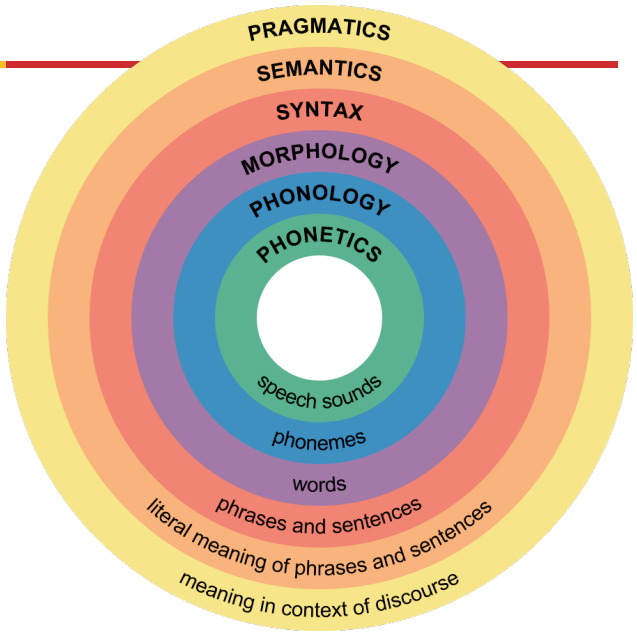
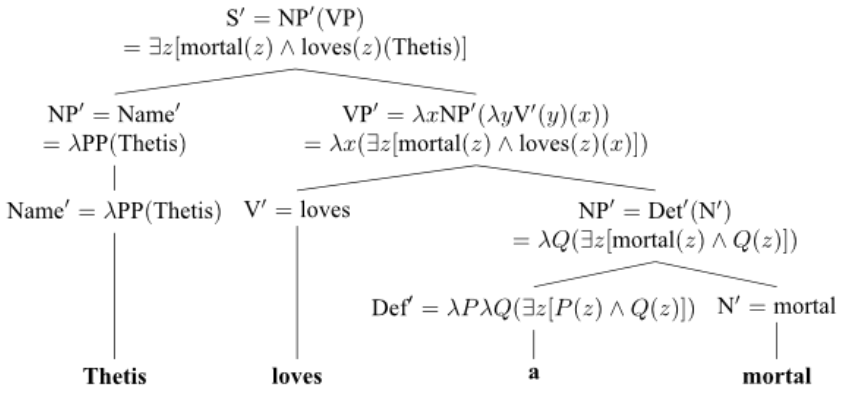


Image: en.wikipedia.org/wiki/File:Major_levels_of_linguistic_structure.svg

17

Semantics: Meaning



https://plato.stanford.edu/entries/computational-linguistics/

18

Semantics

- What kinds of things can we not do well with the tools we have already looked at?
  - Retrieve information in response to unconstrained questions: e.g., travel planning
  - Play the "chooser" side of 20 Questions
  - Provide accurate information about a topic
  - Support human-computer interfaces
  - ~~Read a newspaper article and answer questions about it~~
- These tasks require that we also consider **semantics**: the meaning of our tokens and their sequences

19

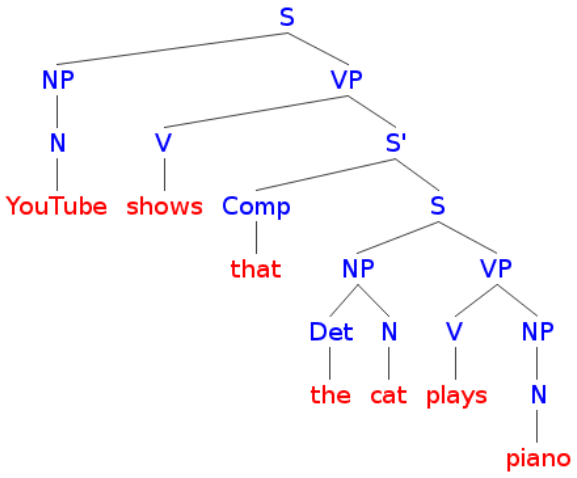
Semantics

- How do you map a sentence to a semantic representation?
  - What are the semantic primitives?
- Schank: 11 (or so) primitives
- The basis of natural language is **conceptual**
- The conceptual level is interlingual, while the sentential level is language-specific
  - De-emphasize syntax
  - Focus on semantic expectations as the basis for NLU

ATRANS
ATTEND
INGEST — to take something inside of
EXPEL
GRASP
MBUILD
MTRANS
MOVE
PROPEL — to apply a force to
PTRANS
SPEAK

20

Syntax: Grammar



https://allthingslinguistic.com/post/100617668093/how-to-draw-syntax-trees-part-3-type-1-a

21

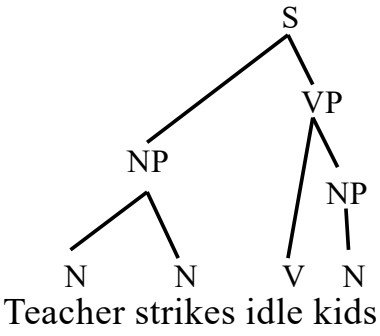
Syntax is about Sentence Structures

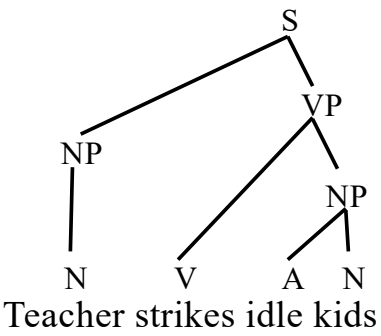
- Sentences have structures and are made up of constituents.
  - The constituents are phrases.
  - A phrase consists of a head and modifiers.
- The category of the head determines the category of the phrase
  - e.g., a phrase headed by a noun is a noun phrase

22

Alternative Parses

Teacher strikes idle kids





23

Phonology: Processing of Sounds

https://upload.wikimedia.org/wikipedia/commons/a/a5/Tsunami_by_hokusai_19th_century.jpg

"tsunami"

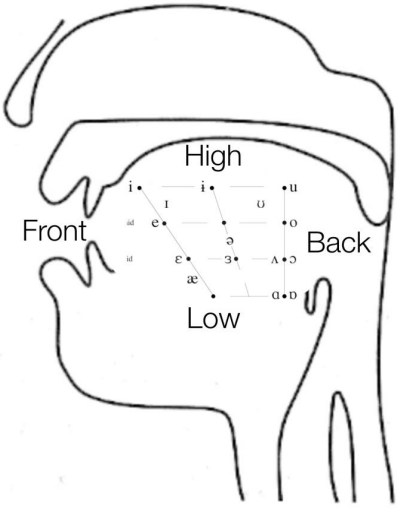
↓

soonahmi

24

Phonetics

- Physical production/ understanding of sounds



https://wstyler.ucsd.edu/talks/l111_3_phonetics_review_handout.html

25

Phonological Ambiguity

- I mate or duck
- I'm eight or duck
- Eye maid; her duck
- Aye mate, her duck
- I maid her duck
- I'm aid her duck
- I mate her duck
- I'm ate her duck
- I'm ate or duck
- I mate or duck

All sound like
"I made her duck"

Duck example Jurafsky & Martin "Speech and Language Processing"

26

Human Languages

- You know ~50,000 words of primary language, each with several meanings
- Six year old knows ~13000 words
- First 16 years we learn 1 word every 90 min of waking time
- Mental grammar generates sentences
  - virtually every sentence is novel!
- 3 year olds already have 90% of grammar
- ~6000 human languages – none of them simple!

Adapted from Martin Nowak 2000 – Evolutionary biology of language – Phil.Trans. Royal Society London

27

Human spoken language

- Most complicated mechanical motion of the body
  - Movements must be accurate to within half mm
  - synchronized within hundredths of a second
- We can understand up to 50 phonemes/sec (normal speech 10-15ph/sec)
  - but if sound is repeated 20 times /sec we hear continuous buzz!
- All aspects of language processing are involved and manage to keep apace

Adapted from Martin Nowak 2000 – Evolutionary biology of language – Phil.Trans. Royal Society London

28

One Huge Challenge in NLP: Ambiguity

- Lexical:
  - Label (noun or verb)?
  - London (Jack or capital of the UK)?
- Syntactic (examples from newspaper headlines):
  - Prepositional Phrase Attachment: Ban on Nude Dancing on Governor's Desk
  - Word Sense Disambiguation: Iraqi Head Seeking Arms
  - Syntactic Ambiguity (what's modifying what): Juvenile Court to Try Shooting Defendant
- Semantic ambiguity:
  - "snake poison"
- Rampant metaphors:
  - "prices went up"

29

Some Ambiguous Headlines

- Juvenile Court to Try Shooting Defendant
- Teacher Strikes Idle Kids
- Kids Make Nutritious Snacks
- Bush Wins on Budget, but More Lies Ahead
- Hospitals are Sued by 7 Foot Doctors

Source: Marti Hearst, i256, at UC Berkeley

30

Dealing with Ambiguity

- Four possible approaches:
  1. Formal approaches -- Tightly coupled interaction among processing levels; knowledge from other levels can help decide among choices at ambiguous levels.
  2. Pipeline processing that ignores ambiguity as it occurs and hopes that other levels can eliminate incorrect structures.
  3. **Probabilistic approaches based on making the most likely choices**
  4. Don't do anything, maybe it won't matter

Source: Jurafsky & Martin "Speech and Language Processing"

31

What Are Words?

bat

http://www.freepngimg.com/download/bat/9-2-bat-png-hd.png

32

What Are Words?

bats

http://www.freepngimg.com/download/bat/9-2-bat-png-hd.png

33

What Are Words?

Fledermaus

flutter mouse

http://www.freepngimg.com/download/bat/9-2-bat-png-hd.png

34

What Are Words?

- Hard to get agreement
- (Human) Language-dependent
- White-space separation is a sometimes okay (for written English longform)
- Social media? Spoken vs. written? Other languages?

35

What Are Words?

pişirdiler

They cooked it.

36

What Are Words?

pişmişlermişlerdi

They had it cooked it.

37

What Are Words?

):

38

What Are Words?

my leg is hurting nasty):

39

What Are Words?

add two cups (a pint): bring to a boil

http://www.dummies.com/wp-content/uploads/88513.image4.jpg

40

Why Language is Hard

- NLP is *AI-complete*
- Abstract concepts are difficult to represent
- LOTS of possible relationships among concepts
- Many ways to represent similar concepts
- Tens of hundreds or thousands of features/dimensions

41

Why Language is Easy

- Highly redundant
- Relatively crude methods provide fairly good results
- Lots of subject matter experts!

42

Some of the Tools

- A mixed bag, at various levels...
  - Tokenizers
  - Regular Expressions and Finite State Automata
  - Part of Speech taggers
  - Grammars
  - Parsers
  - N-Grams
  - Semantic Analysis

43

What will it take?

- Models of computation (state machines)
- Formal grammars
- Knowledge representation
- Search algorithms
- Dynamic programming
- Logic
- Machine learning
- Probability theory

44

A Few Key Problems and Tools

45

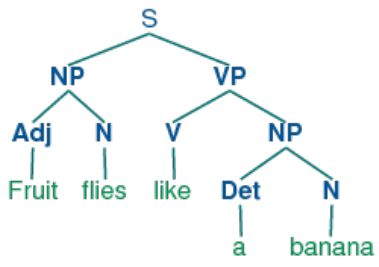
Text annotation tasks

- Classify the entire document ("text categorization")
- Classify word tokens individually
- Classify word tokens in a sequence (i.e., order matters)
- Identify phrases ("chunking")
- Syntactic annotation (parsing)
- Semantic annotation
- Text generation

46

Parts of Speech Tagging

- Part-of-Speech (POS) taggers identify nouns, verbs, adjectives, noun phrases, etc.
- More recent work uses machine learning to create taggers from labeled examples



47

Named Entities (NE) Tagging

- Persons, places, companies
  - "Proper nouns"
  - One of most common information extraction tasks
  - Combination of rules and dictionary
- Example rules:
  - Capitalized word not at beginning of sentence
  - Two capitalized words in a row
  - One or more capitalized words followed by Inc
  - Dictionaries of common names, places, major corporations.
    - Sometimes called "gazetteer"

48

Reference Resolution

- Discourse Knowledge — what have we just said?
  - Paula is here. She is ready.
- Domain/World Knowledge
  - U: I would like to register in a CMSC Course.
  - S: Which number?
  - U: 647.
  - S: Which section?
  - U: Which section is in the evening?
  - S: section 1.
  - U: Then that one.

49

Word Sense Resolution

- Many words have several meanings or **senses**
- We need to resolve which of the senses of an ambiguous word is invoked in a particular use of the word
- I made her duck. (meanings?)

50

Word Sense Resolution

- Many words have several meanings or **senses**
- We need to resolve which of the senses of an ambiguous word is invoked in a particular use of the word
- I made her duck. (meanings?)
  1. I cooked waterfowl for her benefit (to eat)
  2. I cooked waterfowl belonging to her
  3. I created the (plaster?) duck she owns
  4. I caused her to quickly lower her head or body
  5. I waved my magic wand and turned her into undifferentiated waterfowl
- Again, discourse and world knowledge

Duck example Jurafsky & Martin "Speech and Language Processing"

51

Issues with Statistical Parsing

- Statistical parsers still make plenty of errors
- Tree banks are language specific
- Tree banks are genre specific
  - Train on WSJ ➔ fail on the Web
  - standard distributional assumption
- Unsupervised, un-lexicalized parsers exist
  - But performance is substantially weaker

52

Big Applications

- POS tagging, parsing and word sense disambiguation are all medium-sized enabling applications.
  - They don't actually do anything that anyone actually cares about.
  - MT and QA are things people seem to care about.

53

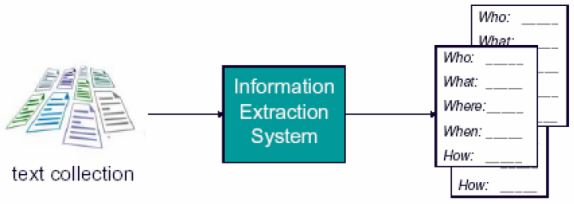
How Difficult is Morphology?

- Examples from Woods et. al. 2000
  - delegate (de + leg + ate) take the legs from
  - caress (car + ess) female car
  - cashier (cashy + er) more wealthy
  - lacerate (lace + rate) speed of tatting
  - ratify (rat + ify) infest with rodents
  - infantry (infant + ry) childish behavior

54

Information Extraction (IE)

- Identify specific pieces of information (data) in an unstructured or semi-structured text
- Transform unstructured information in a corpus of texts or web pages into a structured database (or templates)
- Applied to various types of text, e.g.
  - Newspaper articles
  - Scientific articles
  - Web pages
  - etc.



Source: J. Choi, CSE842, MSU

55

But "Reading" the Web is Tough

- Traditional IE is narrow
- IE has been applied to small, homogenous corpora
- No parser achieves high accuracy
- No named-entity taggers
- No supervised learning

- How about semi-supervised learning?

57

What makes IE from the Web Different?

Less grammar, but more formatting & linking

Newswire

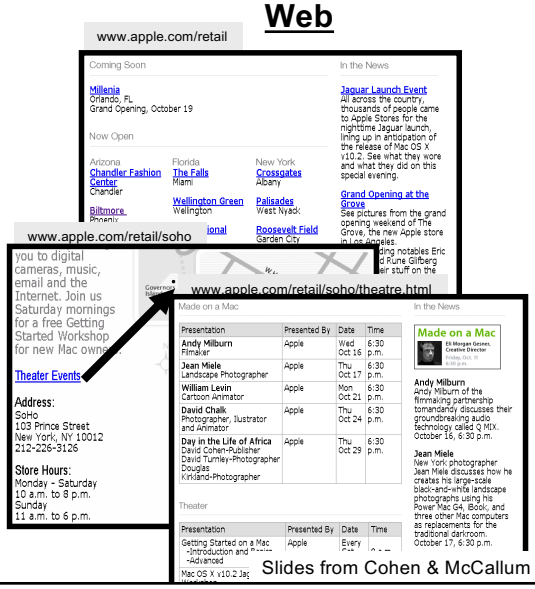
Apple to Open Its First Retail Store in New York City

MACWORLD EXPO, NEW YORK--July 17, 2002--Apple's first retail store in New York City will open in Manhattan's SoHo district on Thursday, July 18 at 8:00 a.m. EDT. The SoHo store will be Apple's largest retail store to date and is a stunning example of Apple's commitment to offering customers the world's best computer shopping experience.

"Fourteen months after opening our first retail store, our 31 stores are attracting over 100,000 visitors each week," said Steve Jobs, Apple's CEO. "We hope our SoHo store will surprise and delight both Mac and PC users who want to see everything the Mac can do to enhance their digital lifestyles."

The directory structure, link structure, formatting & layout of the Web is its own new grammar.

Web



Slides from Cohen & McCallum

58

Machine Translation

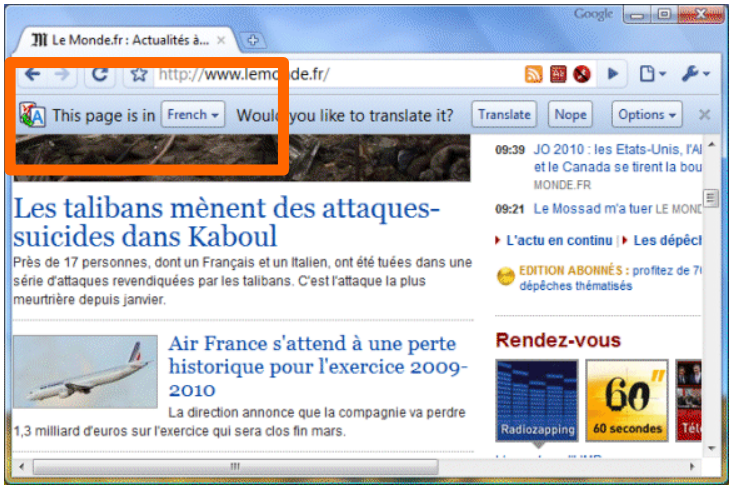
- The automatic translation of texts between languages is one of the oldest non-numerical applications in Computer Science.
- In the past 20 years or so, MT has gone from a niche academic curiosity to a robust commercial industry.

59

Google Translate

Document classification

Machine translation



60

Sentiment Analysis

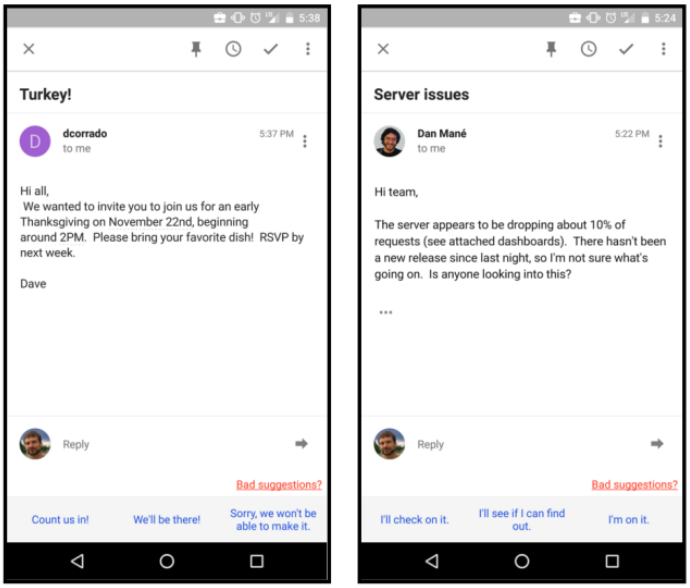
- The field of sentiment analysis deals with categorization (or classification) of opinions expressed in textual documents

The TV is wonderful. Great size, great picture, easy interface. It makes a cute little song when you boot it up and when you shut it off. I just want to point out that the 43" does not in fact play videos from the USB .This is really annoying because that was one of the major perks I wanted from a new TV. Looking at the product description now, I realize that the feature list applies to the X758 series as a whole, and that each model's capabilities are listed below. Kind of a dumb oversight on my part, but it's equally stupid to put a description that does not apply on the listing for a very specific model.

Green color represents positive tone, red color represents negative tone, and product features and model names are highlighted in blue and brown, respectively.

61

Natural Language Generation



https://cdn.arstechnica.net/wp-content/uploads/2015/11/Screen-Shot-2015-11-02-at-9.11.40-PM-640x543.png

62

Document Labeling/Classification

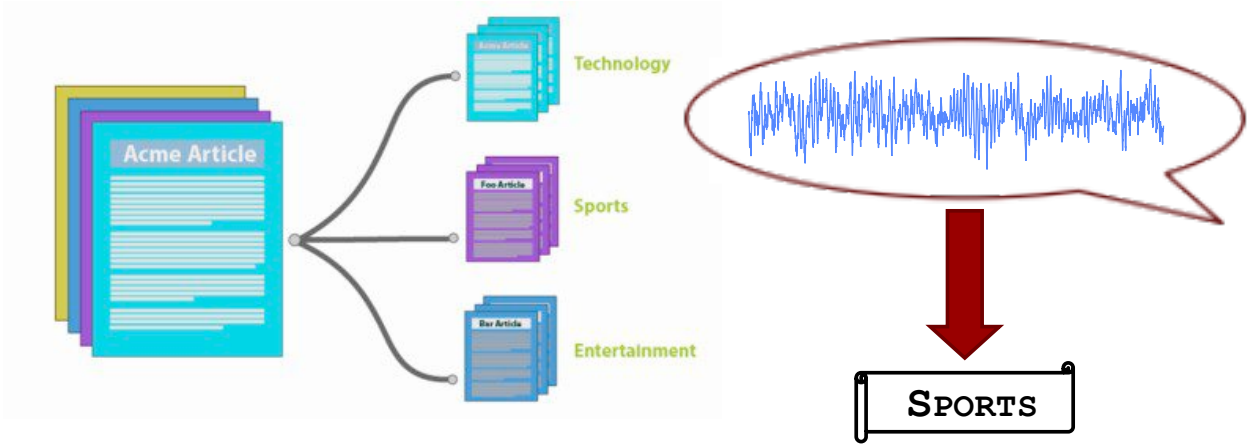
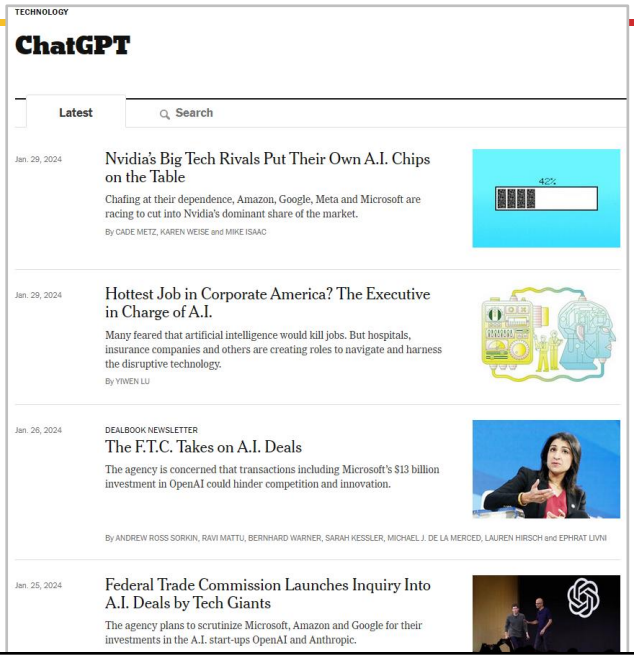


Image: https://www.kdnuggets.com/2015/01/text-analysis-101-document-classification.html

63

But what about LLMs?



64

How Do They Work?

GPT-4 Technical Report

OpenAI*

Abstract

We report the development of GPT-4, a large-scale, multimodal model which can accept image and text inputs and produce text outputs. While less capable than humans in many real-world scenarios, GPT-4 exhibits human-level performance on various professional and academic benchmarks, including passing a simulated bar exam with a score around the top 10% of test takers. GPT-4 is a Transformer-based model pre-trained to predict the next token in a document. The post-training alignment process results in improved performance on measures of factuality and adherence to desired behavior. A core component of this project was developing infrastructure and optimization methods that behave predictably across a wide range of scales. This allowed us to accurately predict some aspects of GPT-4's performance based on models trained with no more than 1/1,000th the compute of GPT-4.

1 Introduction

This technical report presents GPT-4, a large multimodal model capable of processing image and text inputs and producing text outputs. Such models are an important area of study as they have the potential to be used in a wide range of applications, such as dialogue systems, text summarization, and machine translation. As such, they have been the subject of substantial interest and progress in recent years [1–34].

One of the main goals of developing such models is to improve their ability to understand and generate natural language text, particularly in more complex and nuanced scenarios. To test its capabilities in such scenarios, GPT-4 was evaluated on a variety of exams originally designed for humans. In these evaluations it performs quite well and often outscores the vast majority of human test takers. For example, on a simulated bar exam, GPT-4 achieves a score that falls in the top 10% of test takers. This contrasts with GPT-3.5, which scores in the bottom 10%.

On a suite of traditional NLP benchmarks, GPT-4 outperforms both previous large language models and most state-of-the-art systems (which often have benchmark-specific training or hand-engineering). On the MMLU benchmark [35, 36], an English-language suite of multiple-choice questions covering 57 subjects, GPT-4 not only outperforms existing models by a considerable margin in English, but

65

2 Scope and Limitations of this Technical Report

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [39] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [40]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.

We are committed to independent auditing of our technologies, and shared some initial steps and ideas in this area in the system card accompanying this release.[2] We plan to make further technical details available to additional third parties who can advise us on how to weigh the competitive and safety considerations above against the scientific value of further transparency.

3 Predictable Scaling

A large focus of the GPT-4 project was building a deep learning stack that scales predictably. The primary reason is that for very large training runs like GPT-4, it is not feasible to do extensive model-specific tuning. To address this, we developed infrastructure and optimization methods that have very predictable behavior across multiple scales. These improvements allowed us to reliably predict some aspects of the performance of GPT-4 from smaller models trained using $1,000\times - 10,000\times$ less compute.

3.1 Loss Prediction

The final loss of properly-trained large language models is thought to be well approximated by power laws in the amount of compute used to train the model [41, 42, 2, 14, 15].

To verify the scalability of our optimization infrastructure, we predicted GPT-4's final loss on our internal codebase (not part of the training set) by fitting a scaling law with an irreducible loss term (as in Henighan et al. [15]): $L(C) = aC^b + c$, from models trained using the same methodology but using at most 10,000x less compute than GPT-4. This prediction was made shortly after the run

Slide: Dr. Frank Ferraro

66

2 Scope and Limitations of this Technical Report

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [39] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [40]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.

We are committed to independent auditing of our technologies, and shared some initial steps and ideas in this area in the system card accompanying this release.[2] We plan to make further technical details available to additional third parties who can advise us on how to weigh the competitive and safety considerations above against the scientific value of further transparency.

3 Predictable Scaling

A large focus of the GPT-4 project was building a deep learning stack that scales predictably. The primary reason is that for very large training runs like GPT-4, it is not feasible to do extensive model-specific tuning. To address this, we developed infrastructure and optimization methods that have very predictable behavior across multiple scales. These improvements allowed us to reliably predict some aspects of the performance of GPT-4 from smaller models trained using $1,000\times$ – $10,000\times$ less compute.

3.1 Loss Prediction

The final loss of properly-trained large language models is thought to be well approximated by power laws in the amount of compute used to train the model [41, 42, 2, 14, 15].

To verify the scalability of our optimization infrastructure, we predicted GPT-4's final loss on our internal codebase (not part of the training set) by fitting a scaling law with an irreducible loss term (as in Henighan et al. [15]): $L(C) = aC^b + c$, from models trained using the same methodology but using at most 10,000x less compute than GPT-4. This prediction was made shortly after the run

Slide: Dr. Frank Ferraro

67

2 / 100 100%

2 Scope and Limitations of this Technical Report

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [39] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [40]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.

We are committed to independent auditing of our technologies, and shared some initial steps and ideas in this area in the system card accompanying this release.[2] We plan to make further technical details available to additional third parties who can advise us on how to weigh the competitive and safety considerations above against the scientific value of further transparency.

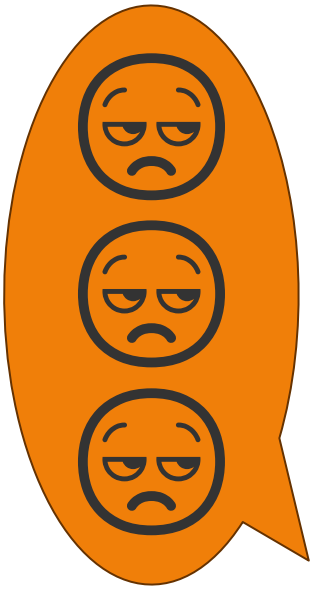
3 Predictable Scaling

A large focus of the GPT-4 project was building a deep learning stack that scales predictably. The primary reason is that for very large training runs like GPT-4, it is not feasible to do extensive model-specific tuning. To address this, we developed infrastructure and optimization methods that have very predictable behavior across multiple scales. These improvements allowed us to reliably predict some aspects of the performance of GPT-4 from smaller models trained using $1,000\times$ – $10,000\times$ less compute.

3.1 Loss Prediction

The final loss of properly-trained large language models is thought to be well approximated by power laws in the amount of compute used to train the model [41, 42, 2, 14, 15].

To verify the scalability of our optimization infrastructure, we predicted GPT-4's final loss on our internal codebase (not part of the training set) by fitting a scaling law with an irreducible loss term (as in Henighan et al. [15]): $L(C) = aC^b + c$, from models trained using the same methodology but using at most 10,000x less compute than GPT-4. This prediction was made shortly after the run



Slide: Dr. Frank Ferraro

68

Known LLM issues

- Bad reproducibility
- Copyright issues
- Can't explain what it's doing
- Can't remember things long term
- Confident bullshitter

Slide: Dr. Lara Martin

69

Conclusions

- NLP is harder than it might seem naively
- Many subtasks
- Statistical NLP is the dominant paradigm
  - supervised learning
  - corpus-based statistics (language models)
  - Some important limitations in practice!
- NL "understanding" has received very little attention

70

Our NLP/NLU Class: Course Goals

- Be introduced to some of the core problems and solutions of NLP (big picture)
- Learn different ways that success and progress can be measured in NLP
- Relate to statistics, machine learning, and linguistics
- Implement NLP programs

71

Our NLP/NLU Class: Course Topics

- Probability, classification, and the efficacy of simple counting methods
- Language modeling (n-gram models, smoothing heuristics, maxent/log-linear models, and distributed/vector-valued representations)
- Sequences of latent variables (e.g., hidden Markov models, some basic machine translation alignment)
- Trees and graphs, as applied to syntax and semantics
- Some discourse-related applications (coreference resolution, textual entailment)
- Special and current topics (e.g., fairness and ethics in NLP)
- Modern, neural approaches to NLP, such as recurrent neural networks and transformers (e.g., BERT or GPT-2)

72