

CS 546 – Advanced Topics in NLP

Dilek Hakkani-Tür



UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN



Siebel School of
Computing
and Data Science

Self-Introduction



- Dilek Hakkani-Tür
 - Prof @ UIUC and Amazon Scholar
 - Senior Principal Scientist at Amazon Alexa AI and Visiting Professor at UC Santa Cruz
 - Google Research, Microsoft Research, ICSI @ UC Berkeley and AT&T Research
- Office Hours: Fri, 9-11am
 - Please get an appointment from: <https://calendly.com/dilek-b2w>
 - Will also stay after the class for questions
- Email: dilek@illinois.edu

My Research Focus



- Enabling conversational machines
 - Complete tasks collaboratively by interacting with machines in natural language
 - Safe and accurate interactions
 - Interactions with web and embodied agents
- More information is available on the UIUC ConvAI web pages:

<https://uiuc-conversational-ai-lab.github.io/>

UIUC ConvAI Lab



Teaching Assistants

- Ishika Agarwal
ishikaa2@illinois.edu
Office hours: Friday 10am-12pm near Siebel 1129 (sign up here: <https://docs.google.com/spreadsheets/d/1p2lsCKnjQn6mPRSacTKeKwhlD3IYhFU5fdDtl7Ur9PM/edit?usp=sharing>)
- Janvijay Singh
jvsingh2@illinois.edu
Office Hours: Wednesday 10am-12pm Zoom/Siebel 1129
Book
here: <https://outlook.office.com/book/CS546OfficeHours@uillinoisedu.onmicrosoft.com/>
- Sagnik Mukherjee
sagnikm3@illinois.edu
Office Hours: Friday 1-3PM
Book
here: <https://outlook.office.com/book/SagnikOH@uillinoisedu.onmicrosoft.com/?ismsaljsauthenabled> (edited)

Class Poll



- Please raise hands if:
 - You are a PhD student
 - You are a Masters student
 - You are an Undergraduate student
 - You are enrolled in Siebel School of Computing and Data Science
 - You trained a deep learning model
 - You fine-tuned an LLM
 - You have published any papers
 - You have published papers at NLP conferences (ACL, EMNLP, NAACL, ACL, COLING, LREC,...)
 - This is your first year at UIUC



Technicalities

- Safety first:
 - [a two-minute video available online](#)
 - [Interactive Building Maps](#)
- Students with disabilities

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor as soon as possible and provide the instructor with a Letter of Academic Accommodations from Disability Resources and Educational Services (DRES). To contact DRES, you may visit 1207 S. Oak St., Champaign, call 217-333-1970, e-mail disability@illinois.edu or visit the [DRES website](#). Here is the direct link to apply for services at DRES, <https://www.disability.illinois.edu/applying-services>.

Class Web Page and Resources



- Canvas: <https://canvas.illinois.edu/courses/59757>
- Piazza for offline discussions:
 - Sign up link: <https://piazza.com/illinois/fall2025/cs546>
 - Access code: u66756wqqwm

Goals for this class

- Cover advanced and recent research topics about natural language processing (NLP)
 - Mainly based on **deep learning** and **large language models (LLMs)**
 - Note that this is not an introductory level course!
- Review related founding papers
- The target audience is graduate students who plan to or are already working on these topics.

Expectations



- You have basic knowledge of:
 - Natural language processing
 - Machine learning and deep learning
 - Math, probability, and statistics
 - Your knowledge of Python is already pretty solid.

Class Format



- The classes will mainly be in lecture mode
- There will be
 - 2 homeworks, 2 midterms, and one final project (4-6 people team project)
 - NO final exam!

Important Information



- I expect you to attend class!
- Asking questions is encouraged.
 - Not to each other!
 - Please ask clarification questions immediately and save broader discussion topics, at the end of the class.

Class Policy



- **Academic integrity:** Students should pay particular attention to Article 1, Part 4: Academic Integrity. Read the Code at the following URL: <http://studentcode.illinois.edu/>.
- Late assignments will not be accepted.
- No cell phones and internet surfing are allowed in the lecture room.
- Do not be late to the class.

Grading

- 2 midterms: 30%
- Homeworks (2): 30%
- Final Project: 40%
 - Proposal write-up: 5%
 - Final project write-up: 15%
 - Code & Demo: 10%
 - Presentations: 10% (2 presentations in total: proposal and final)
- **Proportions subject to change: Check canvas for the latest info.**
- **(TBD) Possible to get extra credits from homeworks and the final project.**

Midterms



- Will take place on **September and October 30th, 2025**.
- Will be in the class.
- Will mainly include multiple choice and T/F questions.
- Will be on topics we have covered to date.

Homeworks



- Two coding homeworks with experimental results.
- **TBD** If you get one of the **top N** highest results on experiments, you will get up to **3 grade points bonus**.
- You will submit
 - running code that we can test on our test set and see the performance.
 - a report describing the approach and results on the validation set.

Final Project



- Aim to have publishable work from the class!
- Group projects of **4-6 students**
 - Can also reach out to TAs or PhD students from the UIUC ConvAI lab for:
 - Ideas and mentoring
 - I can also make suggestions once I see the proposals
 - They may reach out to you based on your idea or you can pick their idea as the project topic (more on this in a few slides)

Final Project (cont.)



- **September 23:** Finalize and sign-up project team
 - OK to use piazza to search for project partners.
 - Ask the TAs for help if you have difficulty finding a team
 - Spreadsheet to sign up project teams:
https://docs.google.com/spreadsheets/d/1EJ_5Xby0mRhHFmSRSmxlv6Gws4Qs5T8P5JUiKYKAZcA/edit?usp=sharing
 - PLEASE MAKE SURE TO NOT ERASE OTHER's ENTRIES!

A	B	C	D	E	F	G	H
Team ID	Project Title (can change later)	Brief Description	Team Lead (Name and email)	Team Members (Names and Emails)	TA (or ConvAI PhD student) Mentor	Notes	

Final Project (cont.)

- Project proposal presentations and reports are due on October 7, 9am
- Presentation and report content should include:
 - Problem description and motivation
 - Related work (2-3 papers)
 - Proposed approach(es)
 - Datasets and metrics
- Proposal presentations will be on zoom (traveling to COLM that week)
 - ~2-3 minutes per team, 1 presenter is ok, but all team members should contribute & be present)

Final Project (cont.)



- Final projects presentations are due on December 2. (TBD: Poster versus presentation)
 - Posters will be presented by student teams on December 2, 4, or 9.
- Final Project Reports are due on December 12.
 - Should be in ACL paper submission format (8 pages) & be like a regular paper submission.
 - In addition to the content in the proposal, they should include:
 - Contributions of the work and main findings.
 - Experiments and results, as well as an analysis of results.
 - Limitations of the approach

Final Project (cont.)

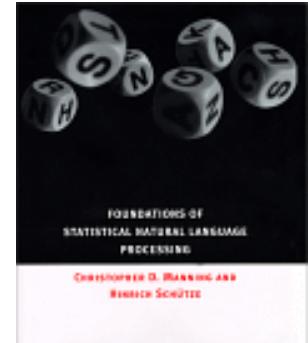
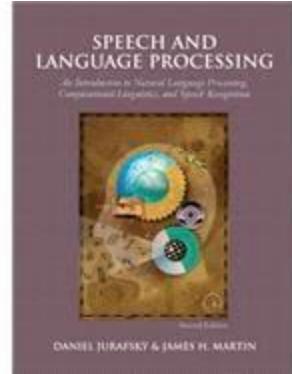


- Start thinking about the proposal soon and reach out to me if you'd like to discuss ideas. Example topics:
 - NL understanding
 - E.g., Converting NL to logical form or converting NL Math Theorems to structured form (i.e., Lean)
 - Domain adaptation
 - E.g., Improving LLM performance in a new domain (without forgetting earlier knowledge)
 - Creative idea generation with LLMs
 - E.g., Math Theorem Proving
- TAs and ConvAI lab PhD students are also sharing a list of high-level ideas that we will share soon.

Books and Reading Material



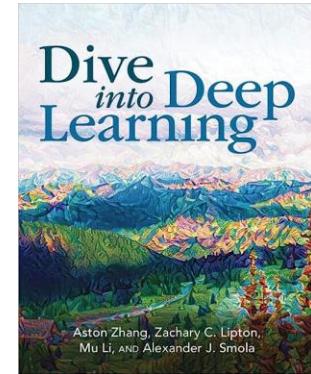
- For general NLP background:
- Daniel Jurafsky and James Martin. [Speech and Language Processing \(3rd draft\)](#)
- Christopher D. Manning and Hinrich Schütze. [Foundations of Statistical Natural Language Processing](#). MIT Press.



Books and Reading Material (cont.)



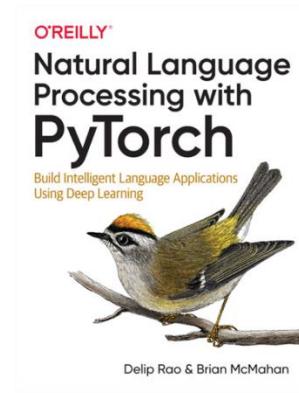
- For deep learning background (and many code examples):
- Aston Zhang, Zack Lipton, Mu Li, and Alex Smola. Dive Into Deep Learning. <https://d2l.ai/>
- Also freely available online.



Books and Reading Material (cont.)



- Delip Rao and Brian McMahan, Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning.
<https://www.oreilly.com/library/view/natural-language-processing/9781491978221/>



Books and Reading Material (cont.)



- I will prepare most of the content based on these books and papers.
- Will share:
 - a list of related papers on my slides (on canvas too)
 - my lecture slides on canvas (right before each class)



Topics for Today

- Introduction to class
 - Class Goals and Logistics
- Introduction to NLP
- Why study NLP?
- Success of LLMs and what is still missing?
- An overview of the topics we will cover

Natural Language Processing (NLP)



- Field at the intersection of many disciplines, such as
 - computer science, linguistics, artificial intelligence, cognitive science, and psychology
- Aims to enable computers to understand, interpret, and generate natural language, leading to improvements in areas, such as
 - summarization, machine translation, question answering, dialogue systems, and so on
- Has applications across several industries, such as
 - virtual personal assistants, customer service, healthcare, data analysis, and so on

Brief History of NLP and Language Modeling



- Claude Shannon, [A Mathematical Theory of Communication](#), The Bell System Technical Journal, 1948.
- Noam Chomsky, [Syntactic Structures](#), 1957.

1950

1960

1970

1980

1990

2000

2010

2020



Brief History of NLP and Language Modeling



- Symbolic/Rule based methods
 - i.e., ELIZA (Weizenbaum, 1966)

```
Welcome to
EEEEE  LL   IIII  ZZZZZZ  AAAAA
EE  LL   II   ZZ  AA  AA
EEEEE  LL   II   ZZZ  AAAAAAA
EE  LL   II   ZZ  AA  AA
EEEEE  LLLLLL  IIII  ZZZZZZ  AA  AA

ELIZA is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true, I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```



Brief History of NLP and Language Modeling



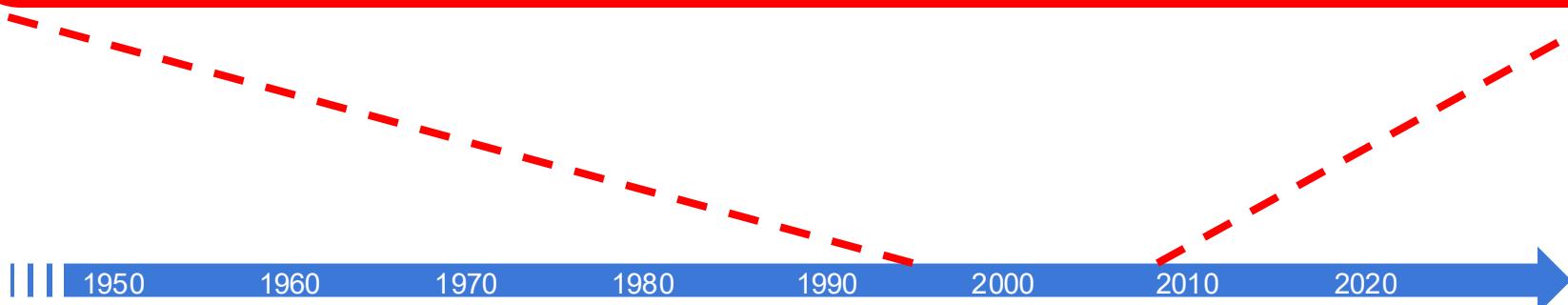
- Hidden Markov Models
 - Baum, L. E.; Petrie, T. (1966). "[Statistical Inference for Probabilistic Functions of Finite State Markov Chains](#)". *The Annals of Mathematical Statistics*. **37** (6): 1554–1563.
 - Jelinek, F.; Bahl, L.; Mercer, R. (1975). "[Design of a linguistic statistical decoder for the recognition of continuous speech](#)". *IEEE Transactions on Information Theory*.
- First Neural Network Architectures
 - Recurrent Neural Networks ([Jordan, 1986](#) and [Elman, 1990](#) Networks)
 - LSTM ([Hochreiter and Schmidhuber, 1995](#))



Brief History of NLP and Language Modeling



- Classification
 - E.g., Support Vector Machines ([Cortes and Vapnik, 1995](#))
- Sequence Tagging
 - E.g., Conditional Random Fields ([Lafferty et al., 2001](#))



Lots of feature engineering!

Brief History of NLP and Language Modeling



- Re-rise of deep neural networks and language models
 - Word embeddings (word2vec, [Mikolov et al., 2013](#))
 - Transformers ([Vaswani et al., 2017](#))
 - BERT ([Devlin et al., 2018](#))



Learning representations!

Brief History of NLP and Language Modeling



- Large Language Models (LLMs)
 - OpenAI's GPT3 ([Brown et al, 2020](#))
 - ChatGPT
 - ...





Topics for Today

- Introduction to class
 - Class Goals and Logistics
- Introduction to NLP
- Why study NLP?
- Success of LLMs and what is still missing?
- An overview of the topics we will cover

Several Real-World Applications

- NLP powers many applications
 - Search engines
 - Virtual personal assistants
 - Chatbots
 - Grammar checkers
 - Writing assistants
 - Question answering
 - Email filtering
 - Summarization
 - Embodied agents

Several Career Opportunities



- Tech companies
- Research labs
- Healthcare (medical transcription, converting interactions to electronic health records, clinical data analysis)
- Law (legal document processing)
- Finance (analyzing reports or news)

Other news (July 1st, 2025):

<https://www.wired.com/story/mark-zuckerberg-meta-offer-top-ai-talent-300-million/>

Screenshots from yesterday:

The image contains three side-by-side screenshots of job search results for 'natural language processing'.

- Google Careers:** Shows 763 jobs matched. Filters include 'natural language processing' in the search bar, 'Locations' dropdown, and 'Skills & qualifications' dropdown.
- Amazon Jobs:** Shows a list of three job postings:
 - Applied Scientist, Sa** at Runway (New York, NY, USA) - Seattle. Job ID: 3065519. Basic qualifications: 1+ years of building, PhD, or Master's deg, Experience in patent, Experience program.
 - Clinical Specialist - P** at Healthcare AI (Santa Clara, CA, USA) - Seattle. Basic qualifications: Relevant medical or
 - Machine Learning Scientist - Natural Language Processing (NLP) - Executive Director** at Machine Learning Center of Excellence (JPMorgan Chase, Seattle, WA) - On-site. Basic qualifications: 15 school alumni work here.
- LinkedIn:** Shows a single job posting for a **Research Scientist, Translate I** at Netflix (United States (Remote)).

Academic Interest: Advancements in AI



- Language is inherently **complex** and **ambiguous**.
- Studying NLP involves solving fascinating problems, such as:
 - How to model and represent meaning?
 - How to deal with ambiguity?
 - How to represent and model context, emotions, tone, or sarcasm?

Academic Interest: Advancements in AI



- NLP is central to modern AI development.
- Innovations in LLMs (like OpenAI's GPTs and Anthropic's Claude models, etc.) are transforming how we think about
 - intelligence, creativity, and automation.

Increasing Academic Interest

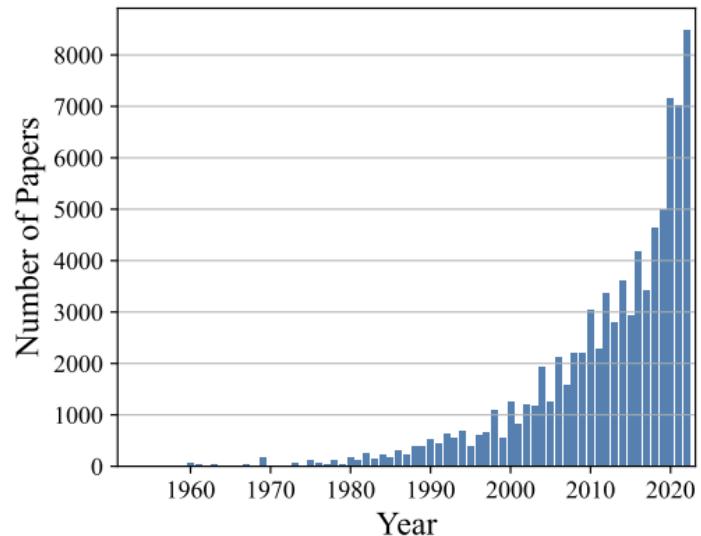


Figure 1: Distribution of number of papers per year in the ACL Anthology from 1952 to 2022.

Figure from [Schopf et al., 2023](#).



Topics for Today

- Introduction to class
 - Class Goals and Logistics
 - Introduction to NLP
 - Why study NLP?
- Success of LLMs and what is still missing?
- An overview of the topics we will cover

Examples of Great Success with LLMs



- **Natural language understanding and generation.** LLMs can:
 - **Answer** questions with context-aware accuracy
 - **Generate** fluent, coherent text (stories, essays, reports)
 - **Summarize** long documents
 - **Translate** languages with near-human quality

Examples of Great Success with LLMs (cont.)



- Code generation and debugging. LLMs can understand and generate **computer code**, and help with:
 - Writing functions
 - Debugging errors
 - Explaining code
 - Completing programs

Conversational Agents and Companions



- LLMs have powered a new generation of **AI chatbots and assistants**, and are able to:
 - Produce very natural sounding and engaging responses
 - Hold multi-turn conversations
 - Remember context (to some degree)
 - Adjust tone and formality

Issues with the current LLMs: Factual Accuracy



- Pretrained language models are known to hallucinate (Ji et al., ACM Comput. Surv., 2022)
- Even for summarization, where the content is given, hallucination rates are around 1-2% (e.g., OpenAI GPT-4o 1.5%)
(<https://huggingface.co/spaces/vectara/Hallucination-evaluation-leaderboard>)
 - And can be higher for other tasks
- Hallucinations are not acceptable for real applications!

Which school did Dilek work at?

Example without retrieval augmentation & tool calling

Dilek Hakkani-Tur has worked at several academic institutions throughout her career. She is currently a professor in the Department of Electrical and Computer Engineering at the University of Arizona, where she leads the Natural Language Processing group.

Prior to her current position, Dr. Hakkani-Tur held several research and academic positions, including working as a research scientist at AT&T Labs, a senior research manager at Microsoft Research, and a professor at the University of Pittsburgh.

Throughout her career, Dr. Hakkani-Tur has made significant contributions to the field of natural language processing, with a particular focus on spoken dialogue systems and multimodal communication. Her work has been recognized with several awards and honors, including the IEEE Speech and Language Processing Technical Committee (SLTC) Award for Distinguished Technical Contributions to the field of Automatic Speech Recognition (ASR) in 2020.

Knowledge Grounded Response Generation

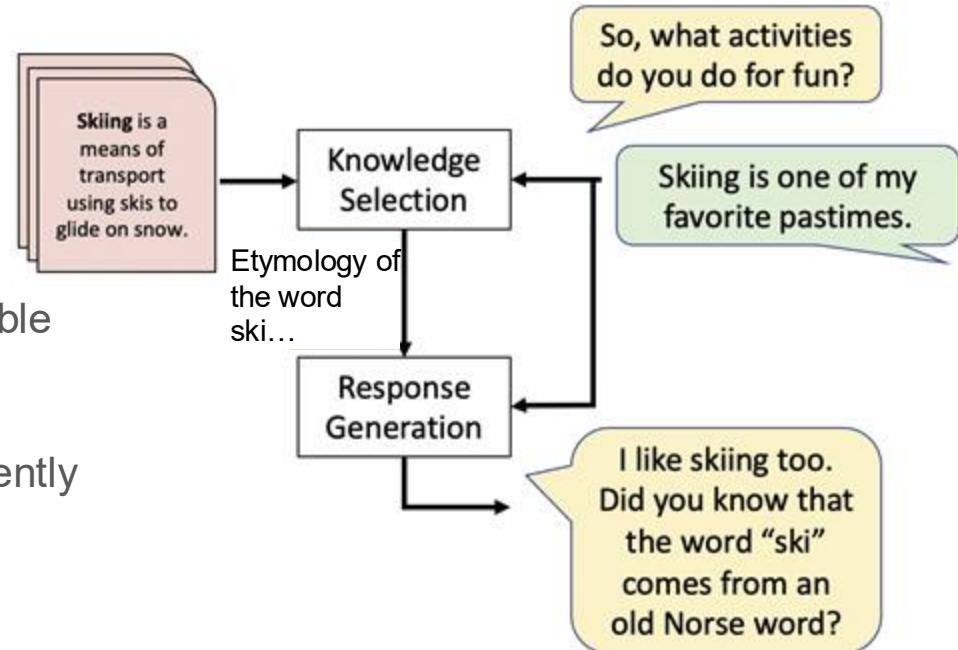


- Retrieving and selecting knowledge to ground on (Eric et al., INLG 2021)
- Generating a response, given the conversation context and selected knowledge

Pros: Text resources may already be available

Cons:

- Unstructured,
- Text segments are modelled independently
- Content is assumed to be reliable



Retrieval Augmentation



- Large language models (like GPTs) are great at generating natural and accurate sounding text, but they don't always *remember specific facts* or stay up to date!
- **Retrieval-Augmented Generation (RAG)** solves this by combining two ideas:
 - **Retrieval:** Searching a knowledge source (such as Wikipedia or company-internal databases) to find relevant information.
 - **Generation:** Using a language model to produce an answer that incorporates that information.
- Instead of making up answers based on model weights, the model looks things up first, then explains them.

Internet-Augmentation



- Earlier datasets use wikipedia (e.g., WoW, TopicalChat), news (e.g., TopicalChat)
- Internet augmented generation
(Komeli et al., ACL 2022)
 - Queries are formulated based on the conversation context
 - Knowledge candidates are retrieved by a search engine

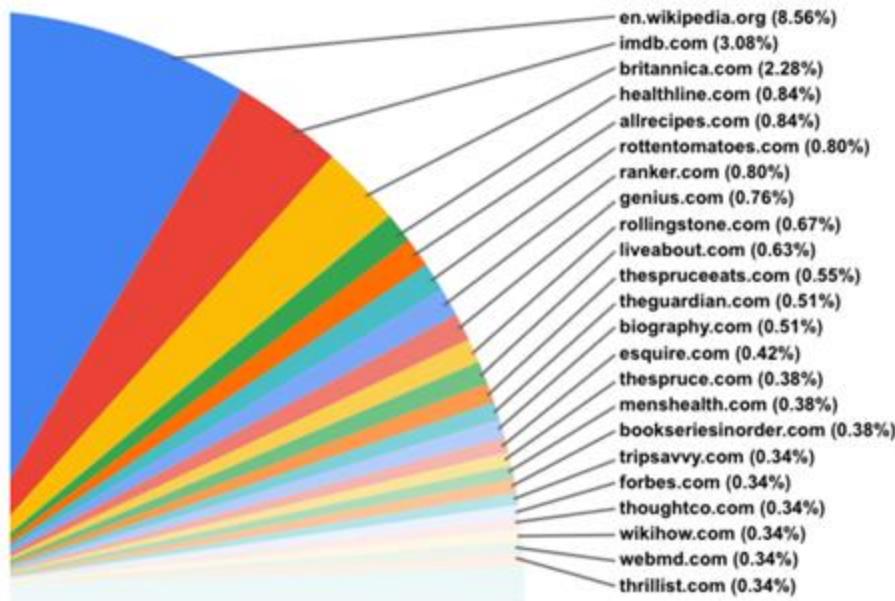


Figure is from (Komeli et al., ACL 2022).

Augmentation with External Tools and APIs



- Large language models can be trained to learn to call tools
- These provide additional knowledge for task completion
 - LAMDA ([Thoppilan et al., arXiv:2201.08239 2022](https://arxiv.org/abs/2201.08239))
 - ToolFormer ([Schick et al., arXiv:2302.04761 2023](https://arxiv.org/abs/2302.04761))
 - LangChain (<https://python.langchain.com/en/latest/>)

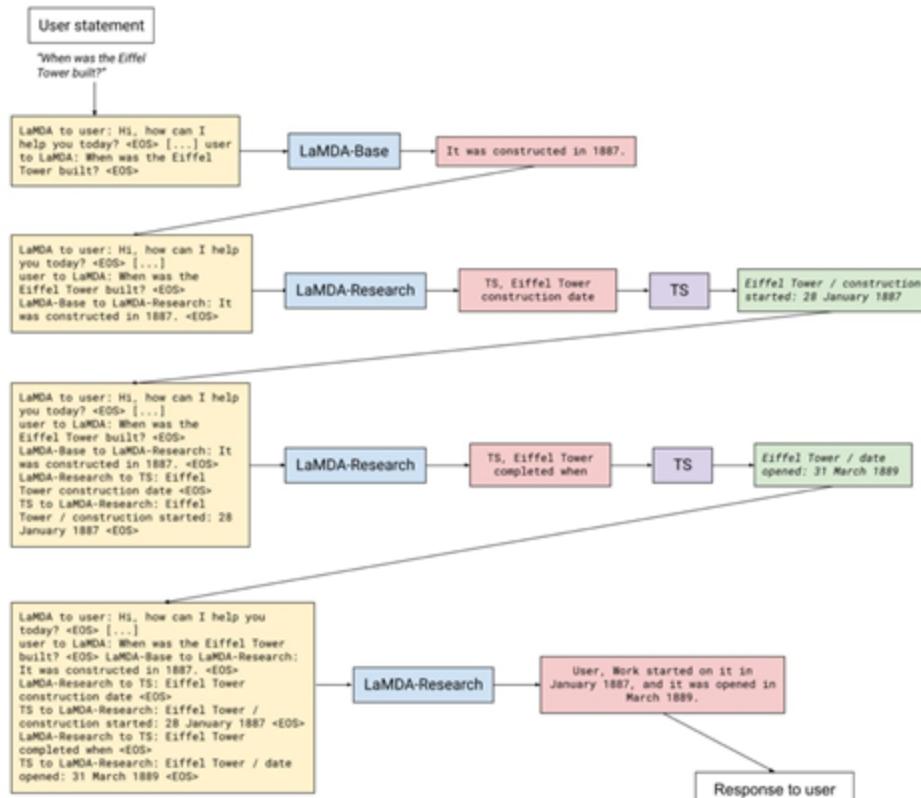


Figure is from ([Thoppilan et al., arXiv:2201.08239 2022](https://arxiv.org/abs/2201.08239))

"Work started on it in January 1887,
and it was opened in March 1889."



Issues with the current LLMs (cont.): Do they really understand?

ChatGPT ▾

how many s's in assistant

The word "assistant" has 2 S's:

- a
- s ← 1
- s ← 2
- i
- s ← (no, this is i)
- t
- a
- n
- t

Total: 2 S's.



Screenshot from 7/28/2025

Tried again last night...
“Thinking longer for a better answer”

Issues with the current LLMs (cont.): Bias, Fairness, and Privacy



- LLMs capture, reflect and amplify:
 - Social biases (gender, race, ethnicity)
 - Cultural or political viewpoints
 - Toxic or offensive content
- They also may unintentionally capture and leak private information.

LLM Safety



- Unsafe responses are not acceptable!
- Detecting and filtering unsafe content from the training data or outputs ([Dinan et al., EMNLP 2019](#))
- Preventing toxic generations during decoding ([Arora et al., ACL 2022](#))
- Reinforcement learning from human feedback ([Bai et al., arXiv:2204.05862, 2022](#))
- Detecting and re-writing unsafe responses ([Bauer et al., EMNLP Findings, 2022](#))
- In-context learning to steer models towards safer outputs ([Meade et al., arXiv:2302.00871, 2023](#))
- Quickly recovering from issues ([Gupta et al., arXiv:2212.10557, 2022](#))

Issues with the current LLMs (cont.): Others



- Lack of full consistency and limited long-term memory:
 - The same instruction can result in **different answers** each time or **contradictions** within a single answer or throughout the same conversation or across multiple sessions.
- AI Overreliance:
 - Users may be misled by plausible sounding mistakes in the outputs or
 - AI sycophancy
- Environmental concerns:
 - Training and inference with LLMs requires massive computational resources and consumes significant energy and raises concerns about carbon footprint.
- Opaque model decisions:
 - Models are often black boxes, and it is hard to explain their decisions.



Topics for Today

- Introduction to class
 - Class Goals and Logistics
- Introduction to NLP
- Why study NLP?
- Success of LLMs and what is still missing?
- An overview of the topics we will cover

Topics we will cover

- First few weeks will mainly include background:
 - an introduction to NLP and deep learning for NLP
 - Neural networks
 - Language Modeling and Word Representations
 - Sequence Modeling
 - Model Architectures and Contextual Embeddings
 - Decoding, Attention
 - Transformers
 - Pre-training and Fine-tuning

Topics we will cover (cont.)

- Following weeks will introduce newer research areas and provide background on them, e.g.,
 - Prompting and Instruction Tuning
 - Reasoning
 - Alignment and Preference Learning
 - Theory of Mind
 - Agentic Systems
 - And so on...

Thanks to several people I borrowed slides from



- Vivian (Yun-Nung) Chen, NTU (especially for the graphics, for several lectures)
- Heng Ji, UIUC
- Marilyn Walker, UCSC
- And others

Topics for Thursday

- Machine Learning (ML) for NLP
- ML Examples: NLP Tasks
- ML Basics
- Calculus Reminders
- Deep Learning
- Linear Regression – Case Study