CS224v

Conversational Virtual Assistants with Deep Learning

Lecture 1: Introduction

Monica Lam

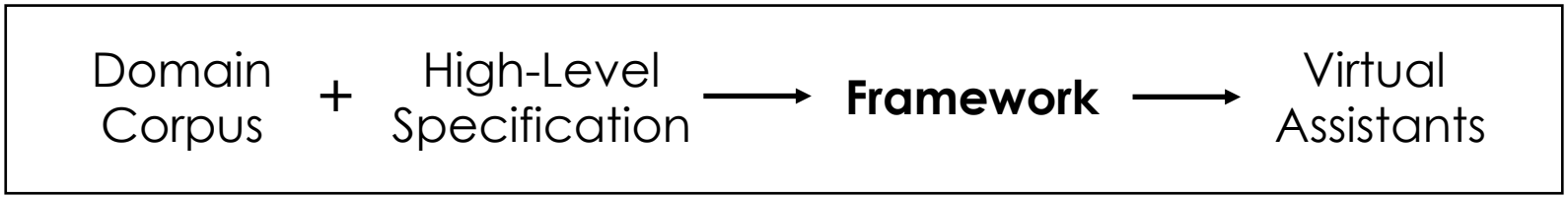
Really Exciting Time for NLP!

- Paradigm shift from Classical ML to LLMs (prompts)
  - Rapid prototyping
- Can create *functional* assistants
  - With a lot less work
- Lots of juicy research topics

Big picture:
Opportunity to change fundamentally
how we acquire and use knowledge across all domains

Course Objectives

- Understand the strengths and weakness of LLMs
- How to build conversational virtual assistants
  - Methodology & framework for assistant development



- Hands-on experience
  - Apply, evaluate, feedback on current methodology
  - Or extend the state of the art in methodology
- Contribute to understanding of Conversational Virtual Assistants

Project Apprenticeship

- Assistance with project selection: Hardest part in research!
  - We suggest
    - Academic research projects (20)
    - External projects
  - Student-initiated projects are also welcome
- Weekly group mentorship meeting
  - We want to make you succeed!

1st attempt at a large, supervised project-based class with lectures

Please bear with us!

Course Personnel

- Prof. Monica Lam. Open Virtual Assistant Lab Director
  - Initiated virtual research interest 8 years ago
  - Experimentalist: AI, Systems, HCI
  - Developed 1st conversational virtual assistant
    - Popular Science's Best of What's New Award in Security in 2019
  - https://oval.cs.Stanford.edu

Course Personnel

- Head TA: Shicheng Liu, PhD student
- TAs:
  - Tommy Bruzzese, CS co-term
  - Amol Singh, CS co-term
  - Arvind Sridhar, CS co-term
- Additional project mentors:
  - John Whaley, Weiyan Shi, Sina Semnani, Jackie Yang
- Additional project advisors
  - Profs. Russ Altmann, Lynn Koegel, Vasilis Syrgkanis, Diyi Yang

Outline

- Part 1:
  - Vision: How will LLMs change the world?
  - The problems with LLMs today
  - How to realize the vision with LLMs?
    - The first results
    - Next steps
- Part 2:
  - Course design

Prediction

- Industrial revolution: automated manual labor (1760-1840)
- Generative AI -- Large language models (LLMs):
  (1) assist
  (2) automate the average knowledge workers in many domains
- Knowledge workers: call agents, salesmen, administrators,
  lawyers, doctors, engineers, programmers, scientists,
  teachers, therapists, social workers
  - Specialized knowledge and expertise
  - Decision making and problem solving
  - With social intelligent behavior

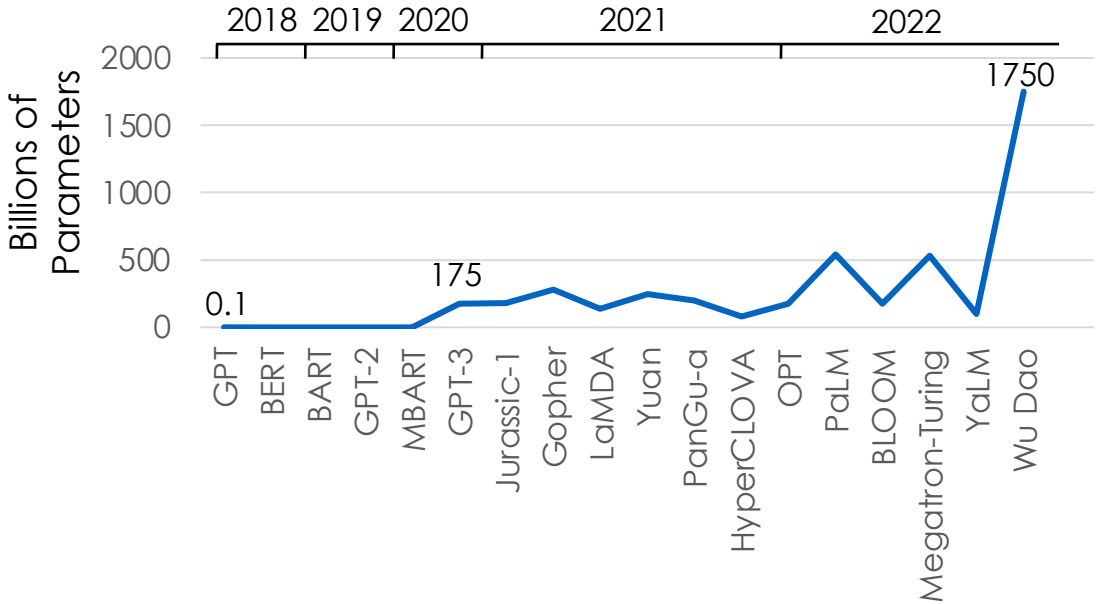
Large Language Models (LLMs)

GPT-3 in 2020

- Trained to predict the next word
- Unsupervised:
  45 TB of Internet text
- 175 Billion parameters

YaLM & GPT-4 (3/14/2023)

- Multimodal: text, images



LLMs Got the Attention of Businesses

Can we use/train a LLM with corporate information?

So it can answer our questions about work in natural language.

LLM Accomplishments

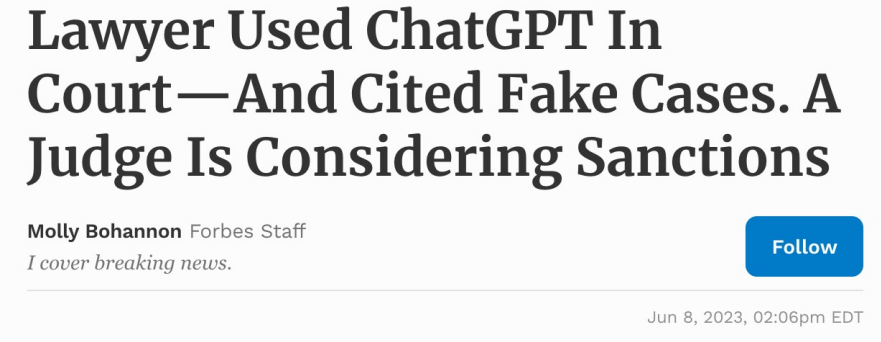
ChatGPT: US Medical Licensing Exam — Pass!

GPT-4: Uniform Bar Examination — Pass!

https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000198

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233

Is GPT Ready to Practice?



Lawyer Used ChatGPT In Court—And Cited Fake Cases. A Judge Is Considering Sanctions

Molly Bohannon Forbes Staff

I cover breaking news.

Follow

Jun 8, 2023, 02:06pm EDT

Exams are designed for humans, not AI

Humans passing the bar can practice

- They know when to look up info

But not AI bots

- don't know the meaning of their words
- make mistakes no humans will

It Hallucinates!

Generative AI = Hallucination

ChatGPT (7/2/2023)

+ Conversational
- Hallucinates on recent events!

Who is the monarch of UK?

As of September 2021, the monarch of the United Kingdom is Queen Elizabeth II. She has been the queen since February 6, 1952, following the death of her father, King George VI. Queen Elizabeth II is the longest-reigning monarch in British history and has played an important role in the country's political and cultural life over the past seven decades.

ChatGPT (7/2/2023)

+ Conversational
- Hallucinates on long-tail info.
  With details and conviction

1978 Hinton, G. E.
Relaxation and its role in vision.
PhD Thesis, University of Edinburgh.

Where did Geoff Hinton get his PhD?

Geoffrey Hinton received his PhD in Artificial Intelligence from the University of Edinburgh in 1978. His doctoral thesis was titled "Relational Learning of Pattern-Matchable Templates". Hinton is widely regarded as one of the pioneers of deep learning and has made significant contributions to the field of artificial intelligence and machine learning throughout his career.

Buying a Ticket from GPT-3

User: I'd like to buy a plane ticket to New York.

Agent: Sure, I can help you with that. What date and time would you like to depart?

User: July 7th, 9am.

Agent: Alright, I can book a flight for you departing on July 7th at 9am.
What airport would you like to depart from?

User: SFO

Agent: Great, I can book a flight for you departing from San Francisco
on July 7th at 9am. Would you like to book this flight now?

User: What is the flight number?

Agent: The flight number is AA123.

+ Perfect flow without training
- Hallucinates: American AA123
DFW → HNL
- Needs to consult DB

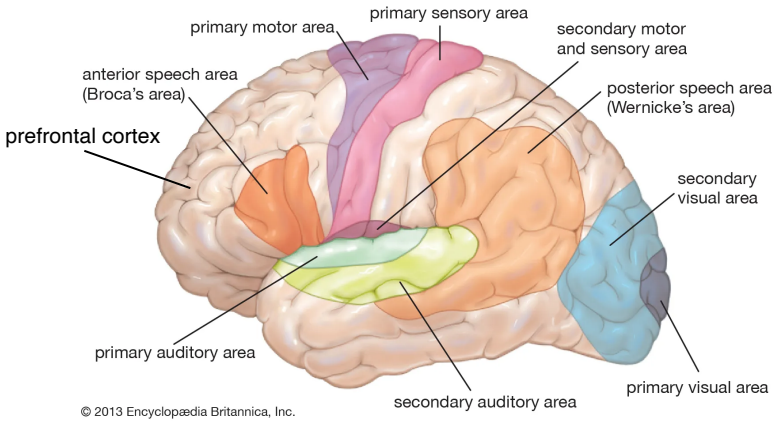
Examples of Business Use Cases

- EPIC medical use cases
  - Requires expert reviews before sharing with patients
- Microsoft travel article (with human reviewers)
  - "Visit Food Bank if you are hungry in Ottawa"
  - Withdrawn
- Meta Galactica: Assistant for scientific articles
  - Withdrawn after 1 day

How do we make sense of LLMs?

We All Have an "LLM" in Our Brain!

- Our "LLM": Speech area
- Prefrontal cortex
  - Inhibition and attention
  - Planning
  - Long-term memory
  - Monitoring external signals
- If prefrontal cortex is damaged, "Speak without thinking"



Takeaway #1

Common Misconception:

Just train a bigger LLM (with domain knowledge),

and hallucination will go away!

Outline

- Part 1:
  - Vision: How will LLMs change the world?
  - The problems with LLMs today
  - **How to realize the vision with LLMs?**
    - The first results
    - Next steps
- Part 2:
  - Course design

How Can LLMs Automate Knowledge Workers

1. Unsupervised Learning
2. Manual Supervised Learning
3. Self Learning

1. Unsupervised Learning

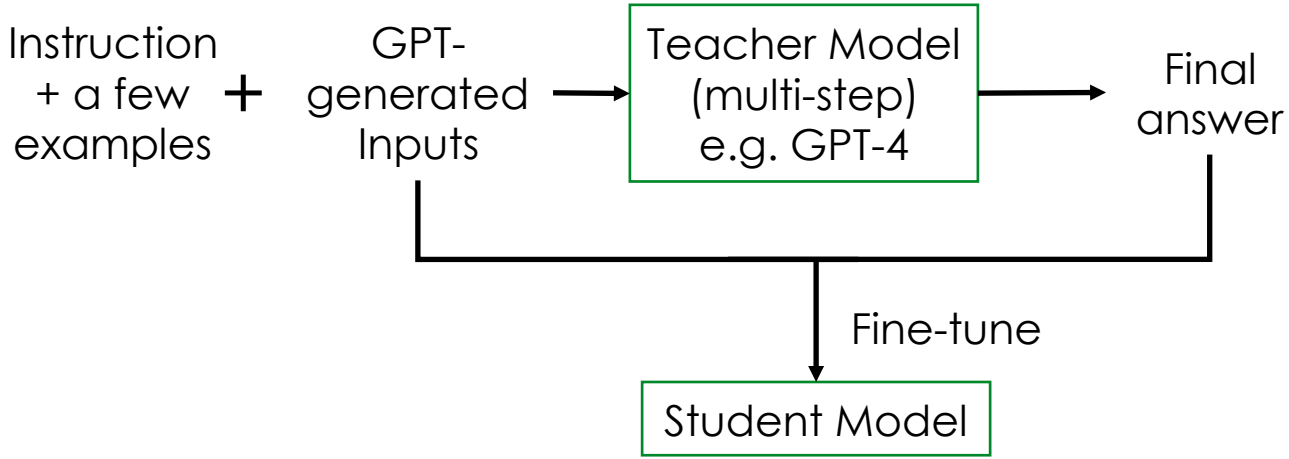
- **GPT-3 is Big! 175B parameters**
  - Trained to predict the next word (45 TB Data)
  - Not a database
- **Generalize / anticipate what typically comes next (statistically)**
  - Finish a thought, a sentence, give an answer, complete a list
  - Finish an essay, continue a conversation
  - Respond like a typical person, base on the speaker's feeling
  - Remember the most common facts (Err on recent, long-tail info)
- **What it is good at**
  - General knowledge, common sense, social intelligence (norms)
- **BUT** it does not know what it is saying (meaning, semantics)

2: Manual Supervised Learning

- GPT-3.5 Fine-tuned with manual annotations
  - Alignment with human values
  - Learn to follow instructions (prompts):
    - Chain-of-thought: Talking out load as it derives or evaluates the answer

It learns how to learn!

3: Self-Learning



Student model learns new concepts by rote (by heart, without thinking)

Reinforcement learning with or without human in the loop

Self-learning → Democratize LLMs

- Open-AI's GPT
  - Concern of privacy, HIPPA
  - Microsoft Azure service – guarantees confidentiality
- Apply self-learning to a smaller model
  - Meta's open LLaMA (7B, 13B, 33B, and 65B)
  - 7B runs on a raspberry pi
  - Use GPT-4 to train LLaMA:
    Our expt: get GPT-3-like behavior!

Outline

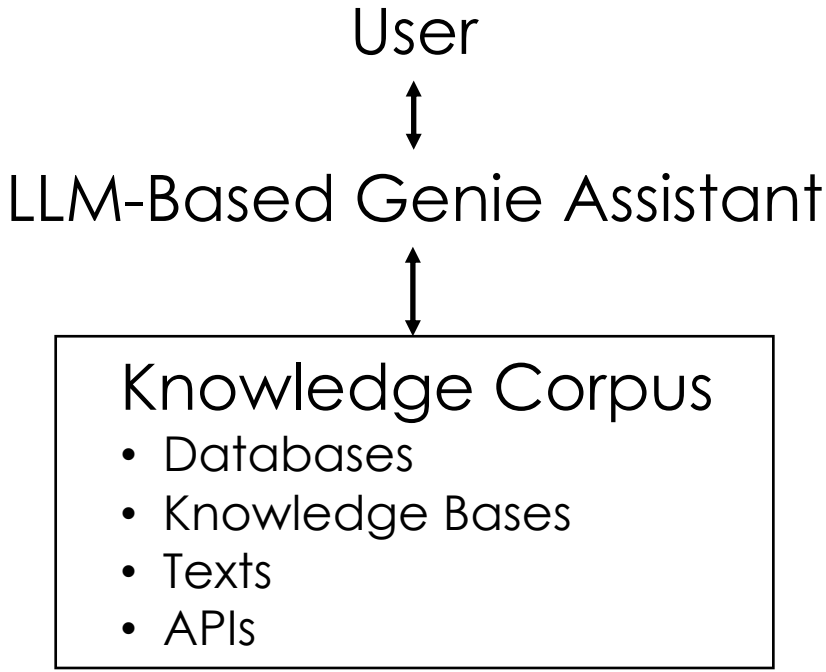
- Part 1:
  - Vision: How will LLMs change the world?
  - The problems with LLMs today
  - How to realize the vision with LLMs?
    - **The first results**
    - Next steps
- Part 2:
  - Course design

Problem #1

- Hallucination is a problem
  - Continual learning (incremental training with recent information)
- Other problems
  - Long-tail information – LLMs are not databases
  - Real-time knowledge
  - Private, case-specific information (e.g. medical transcripts)
- Answer: Grounding in external corpora of data!

Lesson 1: Stop Hallucinating

Grounding with external corpus:



WikiChat: Chatbot on Wiki

1st factual & conversational chatbot grounded on:

- Wikidata: 12B facts
- Wikipedia:
  6.7M articles in English

Note: 61M articles in 333 languages

Wikidata: Largest, Live Knowledge Graph

- 12B facts, 100M entities, 10K properties, 25K contributors
  - Every Wikipedia article has a corresponding entity in Wikidata
  - **Dataset for research in life sciences, digital humanity, etc.**

- Representation: Semantic web (triples)
  - Query with SPARQL

Who founded Stanford?

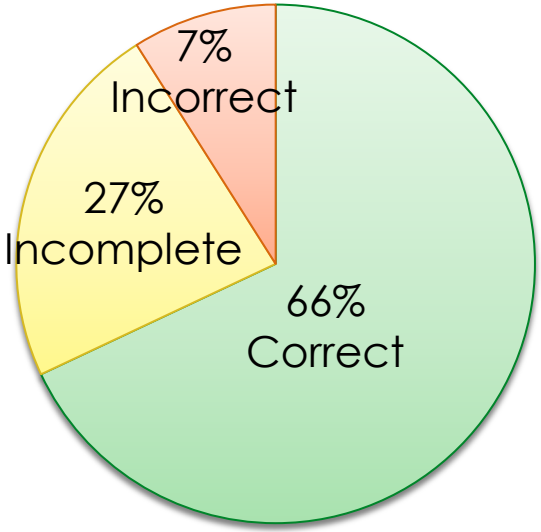
SELECT ?x WHERE

{ wd:Q41506 wdt:P112 ?x. }

Stanford Founded by

A natural language interface can greatly expand access

GPT-3 on New Wiki-WebQuestions Dataset

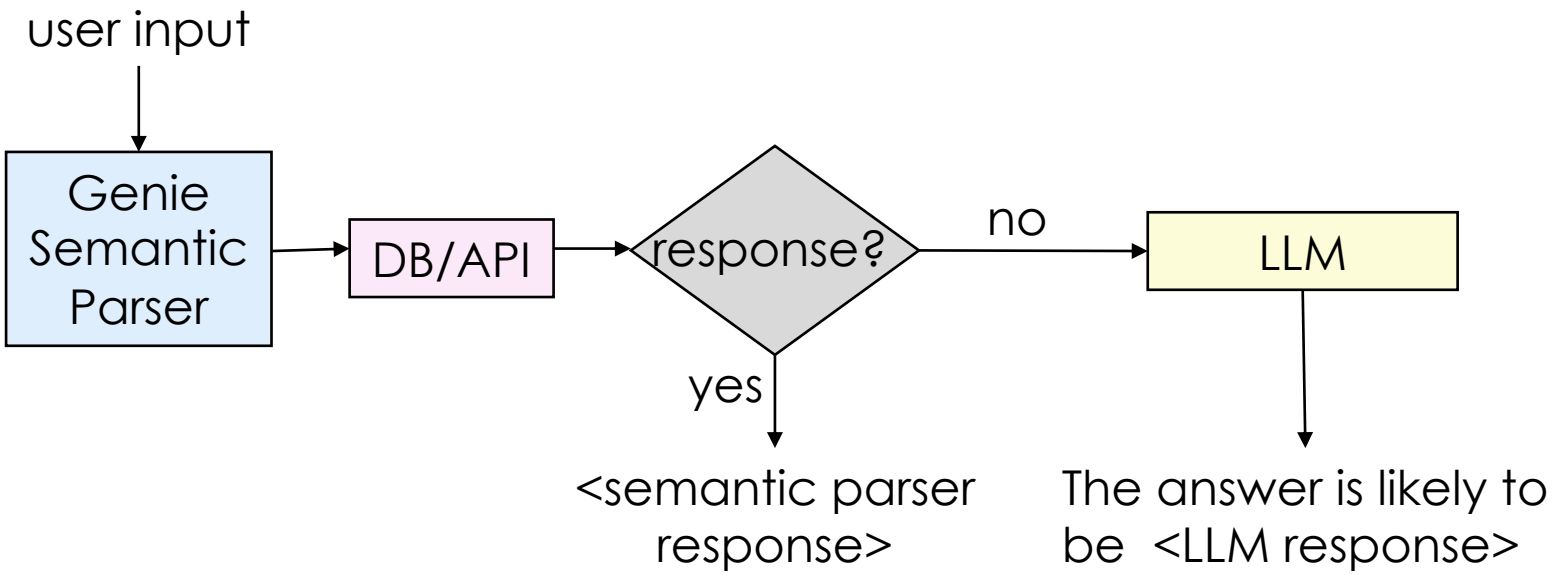


GPT-3 Guesses

- Adapted from WebQuestions for FreeBase
  - Questions from Google Suggest API
  - Real-world popular questions
  - 2431 train, 438 dev, 1384 test
- GPT-3: trained on Wikipedia + Internet

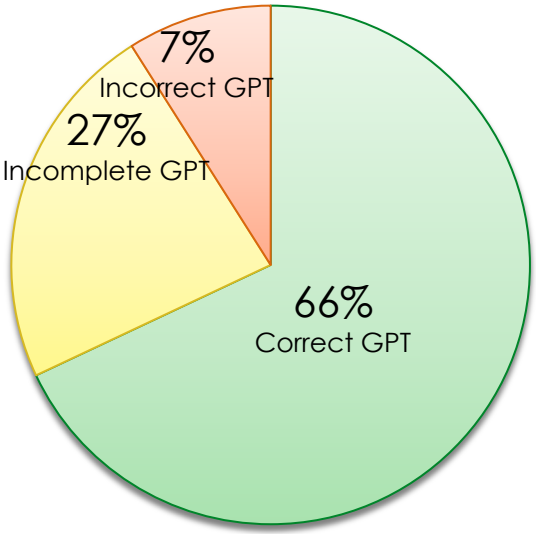
Question:	"What does Obama have a degree in?"
GPT-3:	"Political science degree"
Missing:	"Law degree"

GPT-3 + Genie Semantic Parser

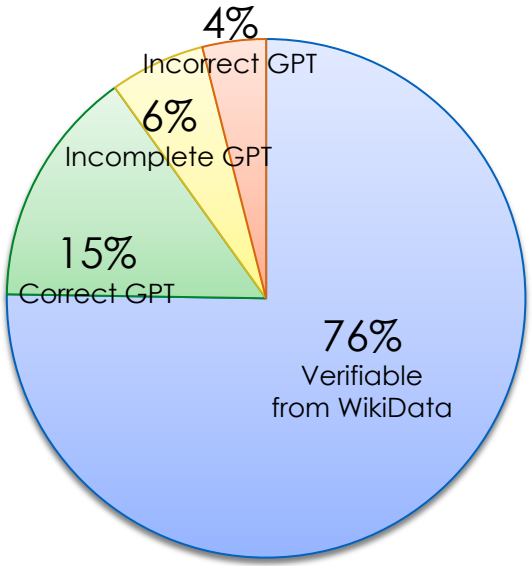


Reduce GPT3 Hallucination with WikiData

Benchmark: Wiki-WebQuestions (Popular, Human-Generated)



Just GPT-3 Guesses



Wikidata (Genie semantic parser)
+ GPT3 Guesses

Takeaway #2

LLMs can produce verifiable answers by generating knowledge base queries

Wikipedia: Open-Text

- 6.7M articles in English
- Lots of interesting information: Why? How? Descriptions …

Microsoft Bing

Traditional (Factuality)

1. Search corpus with query
2. Summarize with LLM

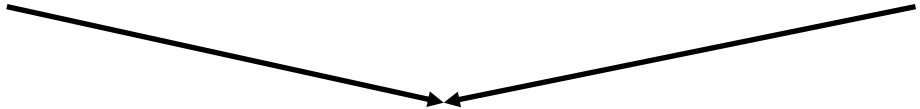
What do Humans Do Today?

LLM (Conversationality + Factuality)

3. Ask GPT
4. Extract claims
5. Fact-check each claim
   - Search corpus with claim
   - Filter incorrect claim

Traditional (Factuality)

1. Search corpus with query
2. Summarize and filter



6. Draft
7. Refine

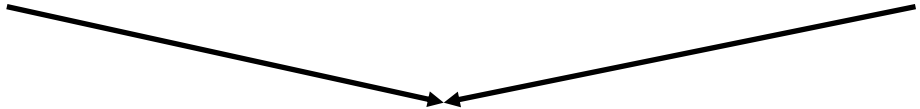
Teaching LLMs with Few-Shot Prompting (GPT-3.5/4)

LLM (Conversationality + Factuality)

3. Ask GPT
4. Extract claims
5. Fact-check each claim
   - Search corpus with claim
   - Filter incorrect claim

Traditional (Factuality)

1. Search corpus with query
2. Summarize and filter



6. Draft
7. Refine

What do you think about the casting of Oppenheimer?

Traditional (Factuality)

1 & 2. Search, summarize, filter:

- Much of the Oppenheimer (2023) movie's cast signed on between September 2021 and April 2022.
- Cilian Murphy is to play as Robert Oppenheimer in the Oppenheimer (2023) movie.

LLM (Conversationality + Factuality)

3. GPT: Christopher Nolan is known for his meticulous casting choices. He often casts well-known actors in his films, and this time he has selected Tom Hanks and … Michael Caine.

4. Extract claims:

- Christopher Nolan is known for his meticulous casting choices.
- Christopher Nolan often casts well-known actors in his films.
- Tom Hanks stars in the Oppenheimer (2023) movie.
- Michael Caine stars in the Oppenheimer (2023) movie.

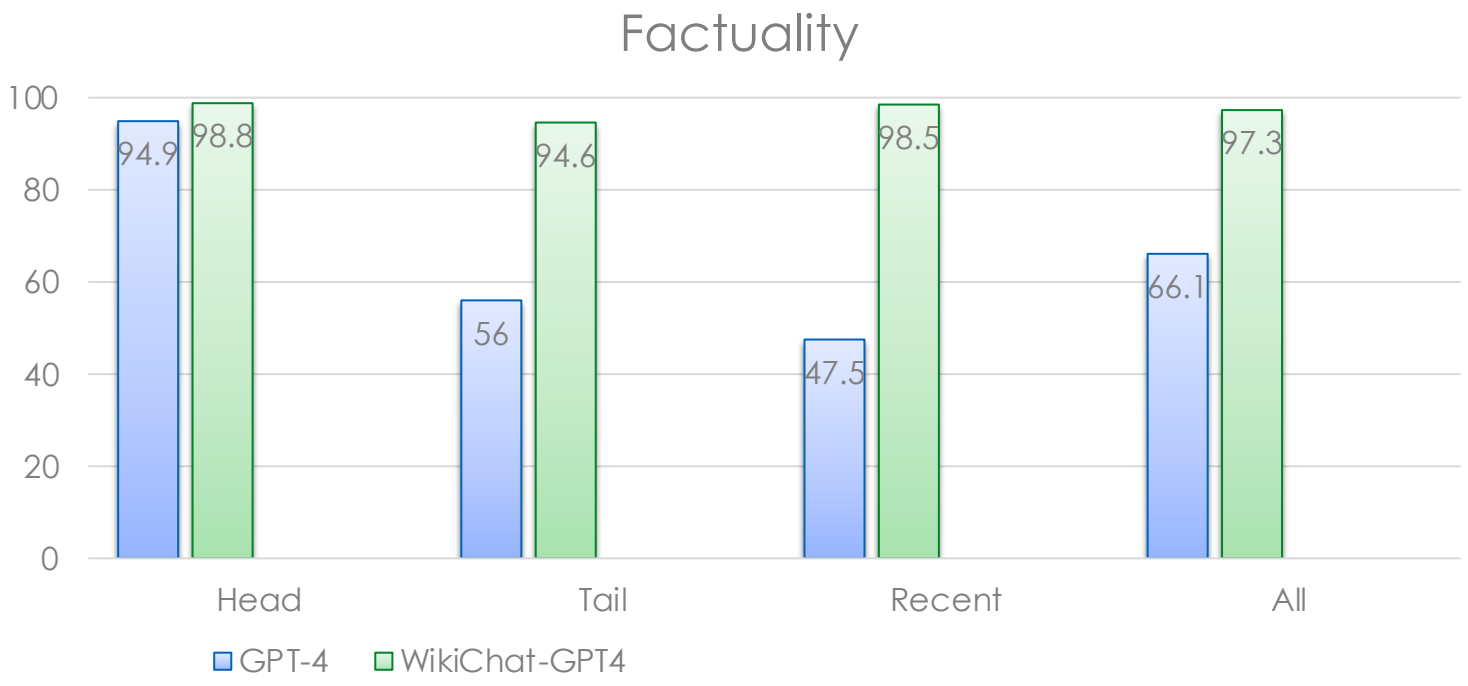
5. Filter incorrect claims:

- Christopher Nolan is known for his meticulous casting choices.
- Christopher Nolan often casts well-known actors in his films.

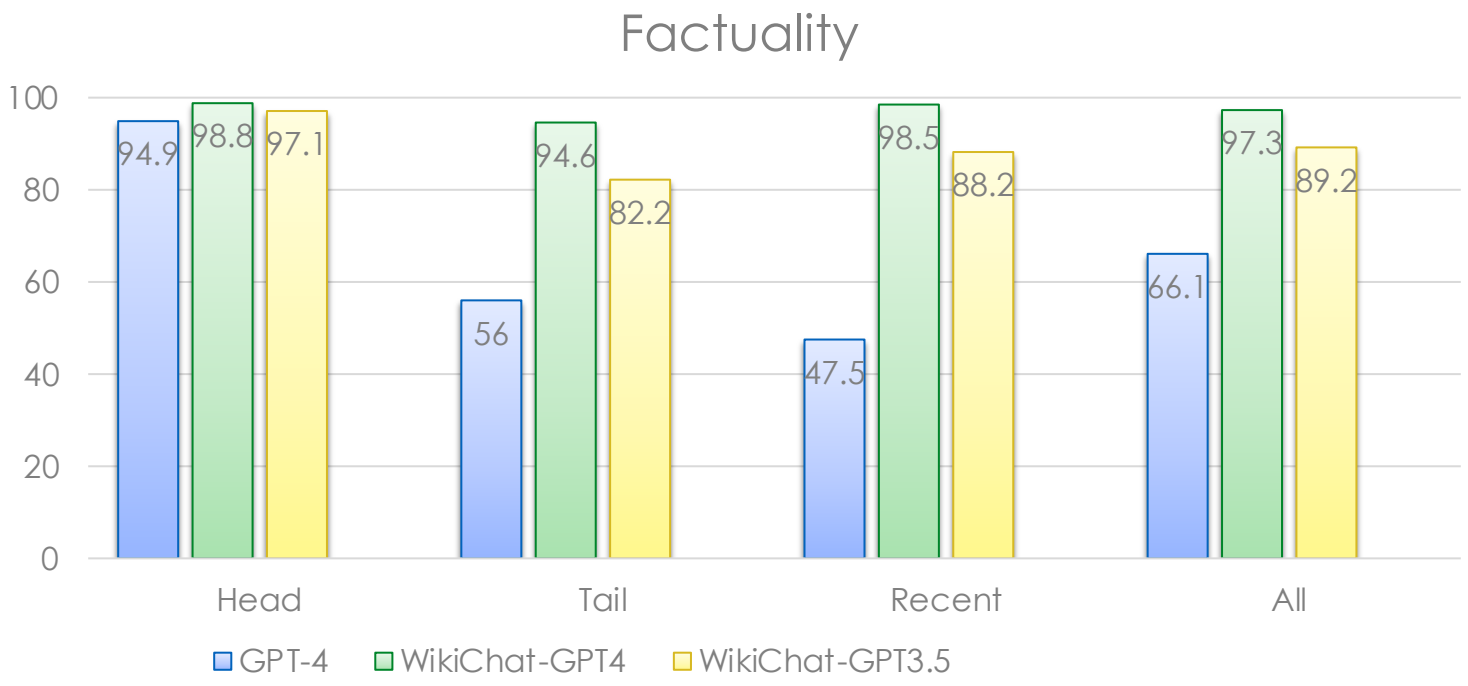
6. Draft: Cillian Murphy has been cast to play Robert Oppenheimer in Christopher Nolan's "Oppenheimer" movie. Christopher Nolan is known for his meticulous casting choices and often casts well-known actors in his films. The movie's cast was signed on between September 2021 and April 2022.

7. Refine: Christopher Nolan has cast Cillian Murphy as the titular role of Oppenheimer, keeping with his tradition of meticulous casting choices. **I'm sure it will be an excellent film!**

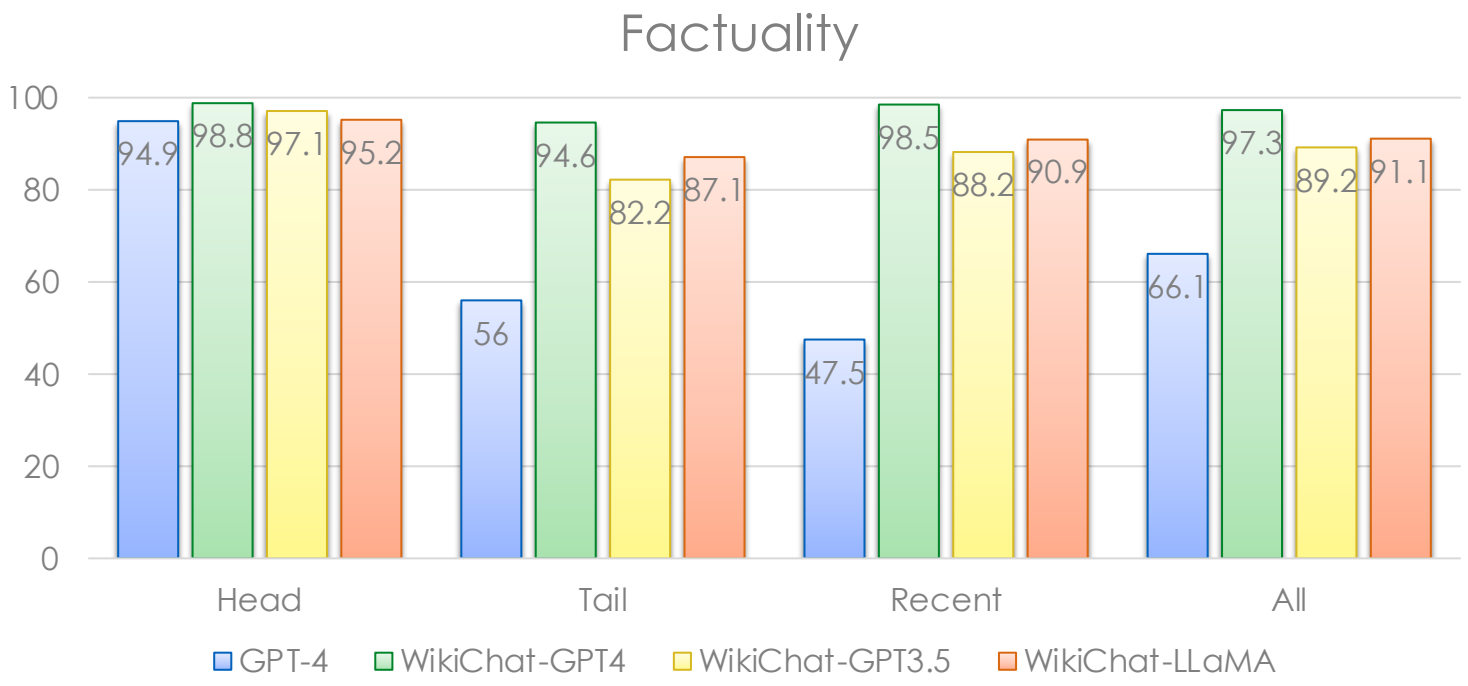
WikiChat-GPT4



WikiChat-GPT3.5



WikiChat-LLaMA (7B Parameters)



Takeaway #3

Elimination of hallucination
→ User-Facing LLM-Based Systems

Outline

- Part 1:
  - Vision: How will LLMs change the world?
  - The problems with LLMs today
  - How to realize the vision with LLMs?
    - The first results
    - **Next steps**
- Part 2:
  - Course design

Exciting time for NLP Research: Paradigm shift from Classical ML to LLMs

- Classical NLP: Data set driven, one for each problem
- LLMs can automatically handle:
  - Paragraph-level summary, question answering
  - Small lists: classification, clustering
  - Intuitive, fluent, socially intelligent responses on open domains, with common-sense reasoning
  - Semantic parsing: Can generate queries for smallish domains

Re-think Approach & Metric

- How to control conversational agents
  - Previous: dialogue tree, dialogue acts (Low level)
  - New: high-level specification
- Metrics
  - Human data collection/annotations are outdated!
  - Old metrics e.g. Bleu scores are meaningless

Research Approach with LLMs

- Experiment
  - Build real systems
  - conduct user studies – incorporate user feedback
  - discover uncharted problems!
- Aim at domain-independent techniques and tools
  - By learning from specific use cases

Tips with LLMs

- Be systematic: Go slow, Build general capabilities
  - Start with solid primitives
  - Develop general next-building blocks
- Do not optimize prematurely
  - Self-learning/distillation: learn high-level functions by heart

Outline

- Part 1:
  - Vision: How will LLMs change the world?
  - The problems with LLMs today
  - How to realize the vision with LLMs?
    - The first results
    - Next steps
- **Part 2:**
  - **Course design**

Tentative Schedule

Date	Monday	Wednesday	Due (Monday)
9/25		Introduction	
10/ 2	Grounding on free-text (HW)	Possible projects	Student profile
10/ 9	Grounding on databases (HW)	Project discussion	HW1
10/16	Multimodal app framework	Project discussion	HW2 + Project abstract
10/23	Proposal presentations	Proposal presentations	Project proposal
10/30	Dialogue Initiatives	Multi-lingual agents	
11/ 6	Large knowledge bases	Robotic automation	
11/13	Tools/APIs/Domain-specific languages	LLM: training, fine-tuning, distillation	
11/20	Thanksgiving		
11/27	Curation of common-sense knowledge	Class discussions	
12/ 4		Final project posters (2.5 hours)	
12/12			Final Report (Tue)

Project Mentorship

All homeworks and projects are done in groups of 2

- Week 4: Project proposal, with a weekly plan
- Weeks 5-10 (excluding Thanksgiving break):
  - Submit a written weekend update (every Monday)
  - Group meeting with mentors during the week
- Week 11: Poster presentation (Dec 6)
- Final project report due Dec 12, 2023.

This Course

	Grade
Participation	15%
Assignment	25%
Final Project	60%

Participation includes

- Class attendance and participation
- Ed discussion
- Meetings with project mentors