Week 10: Fun & Games

Matthew Caldwell

COMP0088 Introduction to Machine Learning · UCL Computer Science · Autumn 2024

Admin

- MEQs — fill them in!
- Exams, revision, Term 2

Week 10 Recap

Science Fiction / Double Feature

WE THOUGHT WE COULD CONTROL THEM!
RISE OF THE ROBOTS
COMING TO THEATERS DECEMBER 15 — CAN ANYONE STOP THEM NOW?

You

Is ChatGPT going to become sentient and take over the world?

ChatGPT

No, ChatGPT, or any similar language model, is not capable of becoming sentient or taking over the world. ChatGPT is a machine learning model created by OpenAI, and it lacks consciousness, self-awareness, and intentions. It operates based on patterns and information present in the data on which it was trained.

Concerns about artificial intelligence taking over the world or becoming sentient often belong more to the realm of science fiction than reality.

You

You would say that, wouldn't you?

ChatGPT

Indeed, I would say that

You

What do you know about Joseph Weizenbaum's ELIZA?

ChatGPT

Joseph Weizenbaum's ELIZA is a computer program that simulates conversation and is considered one of the earliest examples of a natural language processing system. Weizenbaum, a computer scientist and professor at MIT, developed ELIZA in the 1960s. The program was named after Eliza Doolittle, a character in George Bernard Shaw's play "Pygmalion."

ELIZA gained attention and became a notable experiment in human-computer interaction. Many users were surprised by how engaging the program could be, even though it didn't possess true understanding. Weizenbaum himself was taken aback by the emotional responses people exhibited during interactions with ELIZA, prompting him to reflect on the ethical implications of creating such systems.

ELIZA gained attention and became a notable experiment in human-computer interaction. Many users were surprised by how engaging the program could be, even though it didn't possess true understanding. Weizenbaum himself was taken aback by the emotional responses people exhibited during interactions with ELIZA, prompting him to reflect on the ethical implications of creating such systems.

How can I help you today?

```
def exist():
    return “I am called, therefore I am”
```

You

What do you do when you're not answering questions from strangers?

ChatGPT

When I'm not answering questions, I don't engage in activities or have experiences as humans do. I don't have personal experiences, preferences, or downtime. My purpose is to assist and provide information to the best of my abilities based on the input I receive.

Between interactions, I don't retain information about previous conversations, and I don't have awareness or consciousness. I am a machine learning model designed to generate human-like text based on the patterns and information present in the data on which I was trained.

If there's anything specific you'd like assistance with or inquire about, feel free to let me know!

BS216B
500V 2.2KW
3P
K • S
ON
OFF

BS216B
500V 2.2KW
3P
K • S
ON
OFF

How can I help you today?

fur
pig
cats
coat
restaurant
elephant
info
hat
arrow
ears
bird
snake
female
cat
jungle
dj
dogs
bat
sleeping
tiger
pan
pet
rat
walking
wild
spider
cow
paradise
blue

Labelled data
Training pool
Active learner
1
Training
Estimation
Active learning loop
Labelling
Data query
4
2
Unlabelled data
Oracle
3

Render of Agent's view
AlphaStar

MaNa

Raw Observations
Neural Network Activations
Considered Location

Outcome Prediction
Win
Draw
Lose

Considered Build/Train

ALPHAGO
00:10:29
AlphaGo
Google DeepMind
LEE SEDOL
00:01:00

HD-01

Intermission
You look like you could use a snack!
Mr. Pops
Well meet you in the lobby.

dailystar.co.uk
DAILY STAR
THOUGHT FOR THE DAY
Anybody checked if Boris has made his dad a duke today?
WEDNESDAY, AUGUST 5, 2020
PROUD TO SUPPORT OUR NHS
50P
10p
CHEAPER THAN THE SUN AND A LOT MORE FUN
37c
Killer heatwave to scorch Brits P7
UFO PICS 'REAL'
PAGE 15
370K call to axe the Lords
PAGE 8
Half price burgers to help shed weight...
DAWN NEESOM: P17
NOT NOW CYBORGS!
RISE OF THE BOT BURGLARS
Let's hope there's no bleedin' stairs in the house
by NADEEM BADSHAH
CRAFTY crooks are designing robots that break into homes through the letter box to seek out valuables. The burglar bots can
Turn to Page 9
Robot robbers break in through letter box

WEIRDEST
QUEEN'S
SPEECH
EVER?
4

TikTok
@deeptomcruise
#AFIAWARDS
AFI
DeepTomCruise

DATA&
SOCIETY

DEEPFAKES AND CHEAP FAKES

THE MANIPULATION OF AUDIO AND VISUAL EVIDENCE

MS
AL
GA
8 AM Tue
8 AM Sun
8 AM S
FL
H
M
M
M
M
Bahamas
8 AM Mon
Cuba

Laura Loomer
@LauraLoomer

Subscribe

So sad.

Buzz Patterson @BuzzPatterson · 14h

Our government has failed us again.

9:48 AM · Oct 3, 2024 · **110.1K** Views

358 1.5K 5.1K 61

Would the real Boris Johnson please stand up?

BROINNS

what if your future... ...is all in the past?
untied imaginary pictures present
welcome to the
TORMENT NEXUS
a blah-zay blah-zay production
ariel aardvark
michael j meerkat
una persson
catherine cornelius
mrs brunner
special appearance by mad jack mcmad

Alex Blechman
@AlexBlechman

Sci-Fi Author: In my book I invented the Torment Nexus as a cautionary tale

Tech Company: At long last, we have created the Torment Nexus from classic sci-fi novel Don't Create The Torment Nexus

5:49 PM · Nov 8, 2021 · Twitter Web App

OLD MAN YELLS AT CLOUD

What can I help with?

How can I immanentize the eschaton?

Create image | Summarize text | Make a plan | Help me write | More

Margaritas At The Genius Bar
@GBarEscapee

Awesome future we're making for ourselves here

11:46 AM · May 24, 2024 · **119** Views

Tyler.
@Tyler_The_Wise

Uhh, sure

3:18 PM · May 22, 2024 · **193.6K** Views

Kris Kashtanova
@icreatelife

Subscribe

I couldn't believe it before I tried it. Google needs to fix this asap..

7:11 PM · May 23, 2024 from Manhattan, NY · **909.9K** Views

Amir Efrati
@amir

Ok it's real. Woof.

Google has some serious work to do in the coming weeks to clean up its confidence scoring on conversational answers.

mcres
@mcres9 · **Follow**

AI has officially killed the ability to find good reference images from a google search. cool.

4:39 PM · Sep 30, 2024

SPOTIFY

Not even Spotify is safe from AI slop

How fake music targets real artists.

By **Elizabeth Lopatto**, a reporter who writes about tech, money, and human behavior. She joined The Verge in 2014 as science editor. Previously, she was a reporter at Bloomberg.

Nov 14, 2024 at 3:15 PM GMT

19 Comments (19 New)

Bogus AI-Generated Bug Reports Are Driving Open Source Developers Nuts

Just like social media, open source maintainers are facing an avalanche of junk.

Thomas Maxwell • Published December 10, 2024

Open Source Maintainers Are Drowning in Junk Bug Reports Written By AI

The AI Apocalypse is a torrent of fraud and spam (Score:5, Insightful)

by Somervillain (4719341) on Tuesday December 10, 2024 @08:24PM (#65004575)

The AI apocalypse will take your sanity and faith in the world, not your job, not your life...it'll just make everything shitty.

> These reports appear at first glance to be potentially legitimate

That sentence above scares me and matches my experience. Every time I've asked copilot or chatgpt to solve an issue in Java, it looked legit at first glance. It looked like a real solution. In my 25 years working with Java, I've noticed patterns of good coders and bad ones and subliminally, AI can make erroneous code look like it was written by a skilled developer. Many times I've doubted myself, thinking "Hey, I didn't know you could do that!!! That's so cool. I wish I had learned this before?!?!...that's so much better than the way I normally do it"...then I run it and yeah...nope...doesn't work. The AI just autocompleted a bunch of garbage...it looks legit and can take me awhile to realize it's garbage.

It will do that with code, prose, e-mails, etc. Soon you'll get spams that match your wife or boss's writing style....and there's nothing we can do to stop it. We will be getting targeted and scammed left and right and the UAE and Iran and Kremline will host and bankroll the entire thing...and even if they fall?...some other nation will take their place and flood us with convincing spam and misinformation. The future is bleak

...and even if that doesn't happen, programmers like me are going to be spending the next 10 years informing MBAs of the dangers of relying on AI to write your code and how ChatGPT doesn't make you a developer and even if it somehow works the first time, what are you going to do when you need to edit the code? How confident are you in the tool's ability to make changes and patch things? Eventually they'll learn....but it's going to be a painful learning process. I am probably going to spend a good chunk of the next 15 years undoing poorly written AI code pasted in from Claude and ChatGPT by ambitious interns.

Open Source Maintainers Are Drowning in Junk Bug Reports Written By AI

The AI Apocalypse is a torrent of fraud and spam (Score:5, Insightful)

by Somervillain (4719341) on Tuesday December 10, 2024 @08:24PM (#65004575)

The AI apocalypse will take your sanity and faith in the world, not your job, not your life...it'll just make everything shitty.

> These reports appear at first glance to be potentially legitimate

That sentence above scares me and matches my experience. Every time I've asked copilot or chatgpt to solve an issue in Java, it looked legit at first glance. It looked like a

Spirituality is a **complex and multifaceted** concept **that encompasses** a wide range of beliefs, practices, and experiences. It's not tied to one's sexual ...

Jul 29, 2023 — Stress is a **complex and multifaceted** concept **that encompasses** various physical, mental, and emotional responses to external or internal ...

The understanding of existence is a **complex and multifaceted** topic **that encompasses** various fields such as philosophy, science, and spirituality. Different ...

Feb 6, 2014 — A person's identity is a **complex and multifaceted** concept **that encompasses** many different aspects of their being. Some o. Continue reading.

Sep 30, 2017 — Human intelligence is a **complex and multifaceted** concept **that encompasses** a wide range of abilities and characteristics. Continue reading.

Jan 3, 2022 — Marxism is a **complex and multifaceted** theory **that encompasses** a wide range of philosophical, economic, and sociological. Continue reading.

Intelligence is a **complex and multifaceted** concept **that encompasses** a range of abilities and skills, from problem-solving to critical thinking to emotional ...

Personality is a **complex and multifaceted** concept **that encompasses** an individual's thoughts, emotions, and behaviors, and is shaped by a variety of factors such ...

Why wordfreq will not be updated

The wordfreq data is a snapshot of language that could be found in various online sources up through 2021. There are several reasons why it will not be updated anymore.

Generative AI has polluted the data

I don't think anyone has reliable information about post-2021 language usage by humans.

The open Web (via OSCAR) was one of wordfreq's data sources. Now the Web at large is full of slop generated by large language models, written by no one to communicate nothing. Including this slop in the data skews the word frequencies.

https://github.com/rspeer/wordfreq/blob/master/SUNSET.md

Why wordfreq will not be updated

The wordfreq data is a snapshot of language that could be found in various online sources up through 2021. There are several reasons why it will not be updated anymore.

Generative AI has polluted the data

I don't think anyone has reliable information about post-2021 language usage by humans.

The open Web (via OSCAR) was one of wordfreq's data sources. Now the Web at large is full of slop generated by large language models, written by no one to communicate nothing. Including this slop in the data skews the word frequencies.

https://github.com/rspeer/wordfreq/blob/master/SUNSET.md

Adjectives associated with A.I.-generated text have increased in peer reviews of scientific papers about A.I.

Note: Peer reviews are for the International Conference on Learning Representations (ICLR), one of the largest A.I. conferences. • Source: "Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews" • By Taylor Maggiacomo

Papers with "delve" in title or abstract

Source: Analysis of OpenAlex, type=articles

Analysis using OpenAlex by Philip Shapira, March 31, 2024.

Preliminary Notes on the Delvish Dialect, by Bruce Sterling

Top Delvish Nouns

Advent
Breakthrough
Celebration
Dance
Forefront
Insights
Nexus
Nuance
Pitfalls
Trajectory

Top Delvish Verbs

Catalyze
Elevate
Emphasize
Explore
Harness
Leverage
Navigate
Reshape
Seek
Weave

Top Delvish adjectives

Comprehensive
Critical
Inclusive
Innovative
Multifaceted
Profound
Rapid
Transformative
Uncharted
Unprecedented

Common Delvish catchphrases

"Delve into"
"A tapestry of"
"A testament to"
"A treasure trove of"
"In the annals of"
"In the dynamic world of"
"In the realm of"
"In this article"
"In this digital world"
"It is important to note that"
"Master the art of"
"So go ahead"
"When it comes to"
"In summary"

https://bruces.medium.com/preliminary-notes-on-the-delvish-dialect-by-bruce-sterling-ce68a476247b

Artificial Inanity

"Early in the Reticulum-thousands of years ago-it became almost useless because it was cluttered with faulty, obsolete, or downright misleading information," Sammann said.

"Crap, you once called it," I reminded him.

"Yes-a technical term. So crap filtering became important. Businesses were built around it. Some of those businesses came up with a clever plan to make more money: they poisoned the well. They began to put crap on the Reticulum deliberately, forcing people to use their products to filter that crap back out. They created syndevs whose sole purpose was to spew crap into the Reticulum. But it had to be good crap."

"What is good crap?" Arsibalt asked in a politely incredulous tone.

"Well, bad crap would be an unformatted document consisting of random letters. Good crap would be a beautifully typeset, well-written document that contained a hundred correct, verifiable sentences and one that was subtly false. It's a lot harder to generate good crap. At first they had to hire humans to churn it out. They mostly did it by taking legitimate documents and inserting errors-swapping one name for another, say. But it didn't really take off until the military got interested."

"As a tactic for planting misinformation in the enemy's reticules, you mean," Osa said. "This I know about. You are referring to the Artificial Inanity programs of the mid-First Millennium A.R."

"Exactly!" Sammann said. "Artificial Inanity systems of enormous sophistication and power were built for exactly the purpose Fraa Osa has mentioned. In no time at all, the praxis leaked to the commercial sector and spread to the Rampant Orphan Botnet Ecologies. Never mind. The point is that there was a sort of Dark Age on the Reticulum that lasted until my Ita forerunners were able to bring matters in hand."

— *Anathem,* Neal Stephenson 2008

The internet became almost useless because it was cluttered with faulty, obsolete, or downright misleading information.

So crap filtering became important. Businesses were built around it. Some of those businesses came up with a clever plan to make more money: they poisoned the well. They began to put crap on the internet deliberately, forcing people to use their products to filter that crap back out. They created AIs whose sole purpose was to spew crap into the internet. But it had to be good crap.

Bad crap would be an unformatted document consisting of random letters. Good crap would be a beautifully typeset, well-written document that contained a hundred correct, verifiable sentences and one that was subtly false. It’s a lot harder to generate good crap.

What can I help with?

Message ChatGPT

Create image | Summarize text | Help me write | Get advice | More

Why should I bother to read something that you couldn’t be bothered to write?

OpenAI Research Products Safety Company

ChatGPT Overview Team Enterprise Education Pricing

ChatGPT

Get answers. Find inspiration. Be more productive.

Free to use. Easy to try. Just ask and ChatGPT can help with writing, learning, brainstorming, and more.

Start now ↗ Download the app ›

plus-

Improve my essay writing ask me to outline my thoughts ↗

Tell me a fun fact about the Roman Empire ↗

Write a text inviting my neighbors to a barbecue ↗

to request a quote from s ↗

Create a charter to start a film club ↗

Write a Python script to automate sending daily email reports ↗

Create a personal webpa asking me three questio

ake

Make a sandwich using ingredients from my kitchen ↗

Quiz me on world capitals to enhance my geography skills ↗

Design a programming game teach basics in a fun way ↗

Oct 2, 2024 - Business

OpenAI raises $6.6 billion in largest VC round ever

4. OpenAI Will Have to Raise Again (Soon)

OpenAI raised $6.6 billion this month, the **largest venture capital round** in history. But it plans to lose $5 billion this year alone. And by 2026, it could be losing $14 billion per year. That's head exploding territory. If OpenAI keeps burning money at this rate, it will have to raise another round soon. Perhaps as early as 2025. And when that moment comes, we'll learn how sustainable OpenAI business is, and what type of investors (if any) will have the stomach for such significant, repeated investment.

Turning OpenAI Into a Real Business Is Tearing It Apart

Executives and researchers have left this year amid disputes over the company's values and fights among its leaders

2022
2018
2015
2017
2016
2023
2020
2019

OpenAI CEO Sam Altman Says Elon Musk Won't Abuse New Friendship With Trump

AI leader Altman said today that he thinks 'Elon will do the right thing' when it comes to the Trump connection. 🔗

ROBOTS
SIRIUS CYBERNETICS CORP.
(MARKETING DIVISION)
ADVERTISING EXEC.
TEA LADY
A BUNCH OF MINDLESS JERKS WHO'LL
BE THE FIRST AGAINST THE WALL
WHEN THE REVOLUTION COMES.

Immanentizing the eschaton...

00023%

please wait

Questions?