The Good, the bad, and the ugly of AI and machine learning

HPU Faculty Summit: AI Unplugged

Chong Ho Alex Yu, Ph.D., D. Phil.

Jan. 17, 2025

HAWAIʻI PACIFIC UNIVERSITY

Opt-out option

- In this presentation, I will explore a wide range of complex and sometimes controversial topics that arise in the rapidly evolving field of AI, such as DeepFake and DeepNude.
- These discussions are intended to critically examine ethical issues and societal impacts of AI, but I understand that the nature of the content may be uncomfortable or distressing for some attendees.
- If you feel uncomfortable with these topics, you can opt out of them by logging off.

The good:
Applications in teaching and research

Apply to teaching: Personal tutor

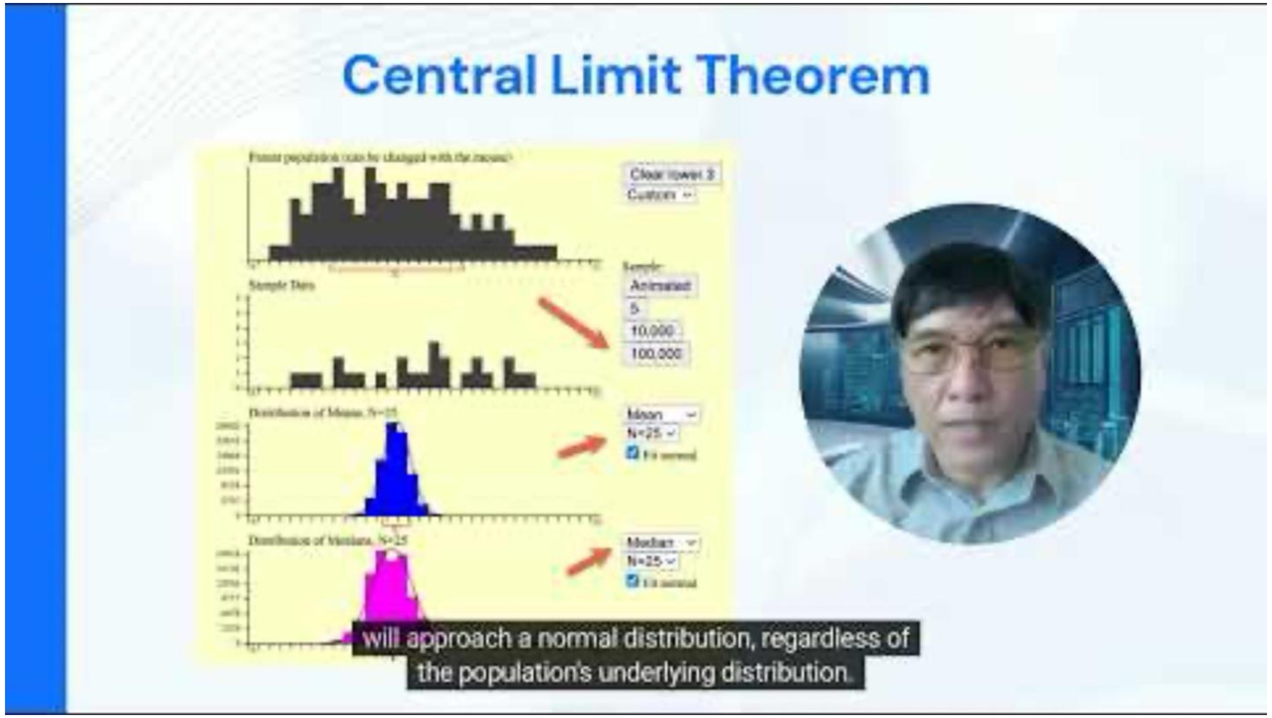
- AI is a **social equalizer** in education and others.
- In the past it was expensive to hire personal tutors or signing up for extra tutorial sessions.
- AI as personal tutor: Research indicates that AI tools can serve as a **nonjudgmental resource** for students, allowing them to ask questions freely without fear of embarrassment or judgment.
- This is especially helpful for students who are hesitant to participate in classroom settings or ask questions to human instructors due to fear of being perceived as incompetent.

Apply to teaching: Media producer

- In the past it was time-consuming and expensive to create instructional videos.
- Today anyone can do it efficiently by recording a short video clip of yourself in the AI video generation website. After you upload a PowerPoint file or script to the system, the AI will use your video clip to generate a lecture where you appear to speak naturally.
- **There is no "evergreen" content**. If you want to update the video using the conventional way, you have to remake the entire video, as inserting new clips into existing footage can compromise the quality and cohesion.
- In the AI system you can simply edit the script and then recreate the video! You can even swap the face and the voice.

Apply to teaching

- Video lectures using myself as the avatar (Example 1):
- https://www.youtube.com/watch?v=OjlNH98dsyA

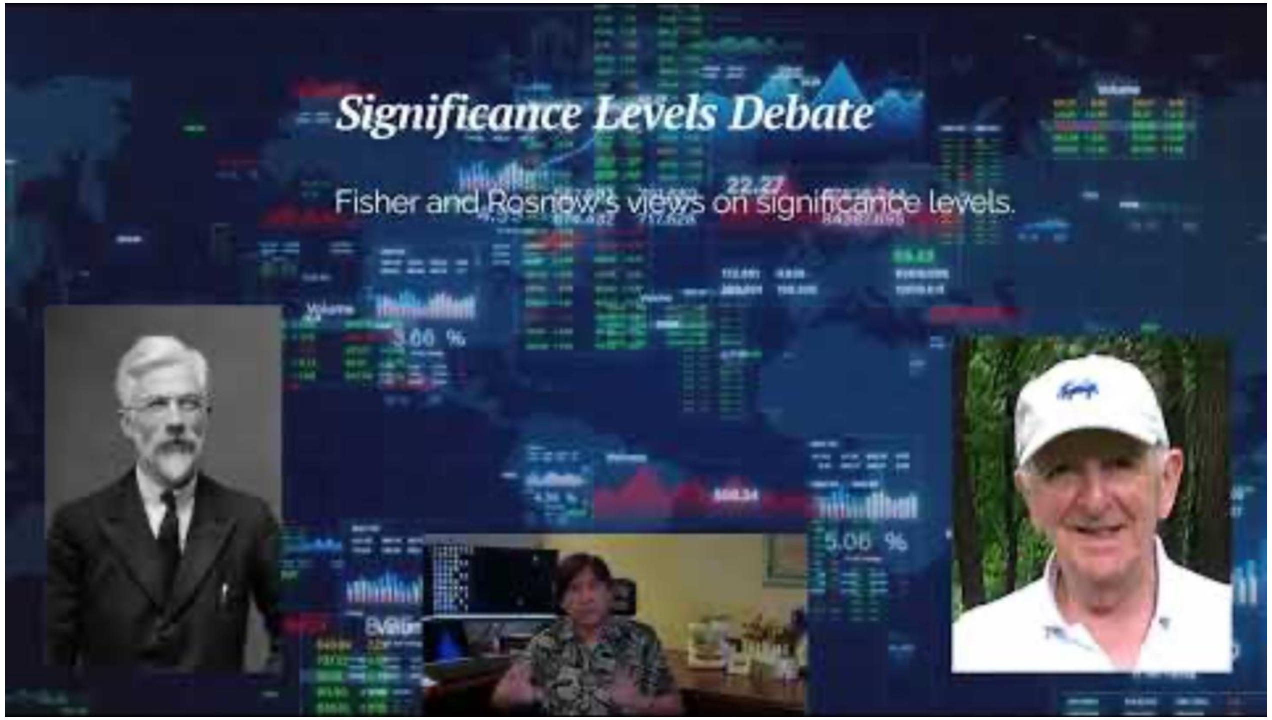


Central Limit Theorem

will approach a normal distribution, regardless of the population's underlying distribution.

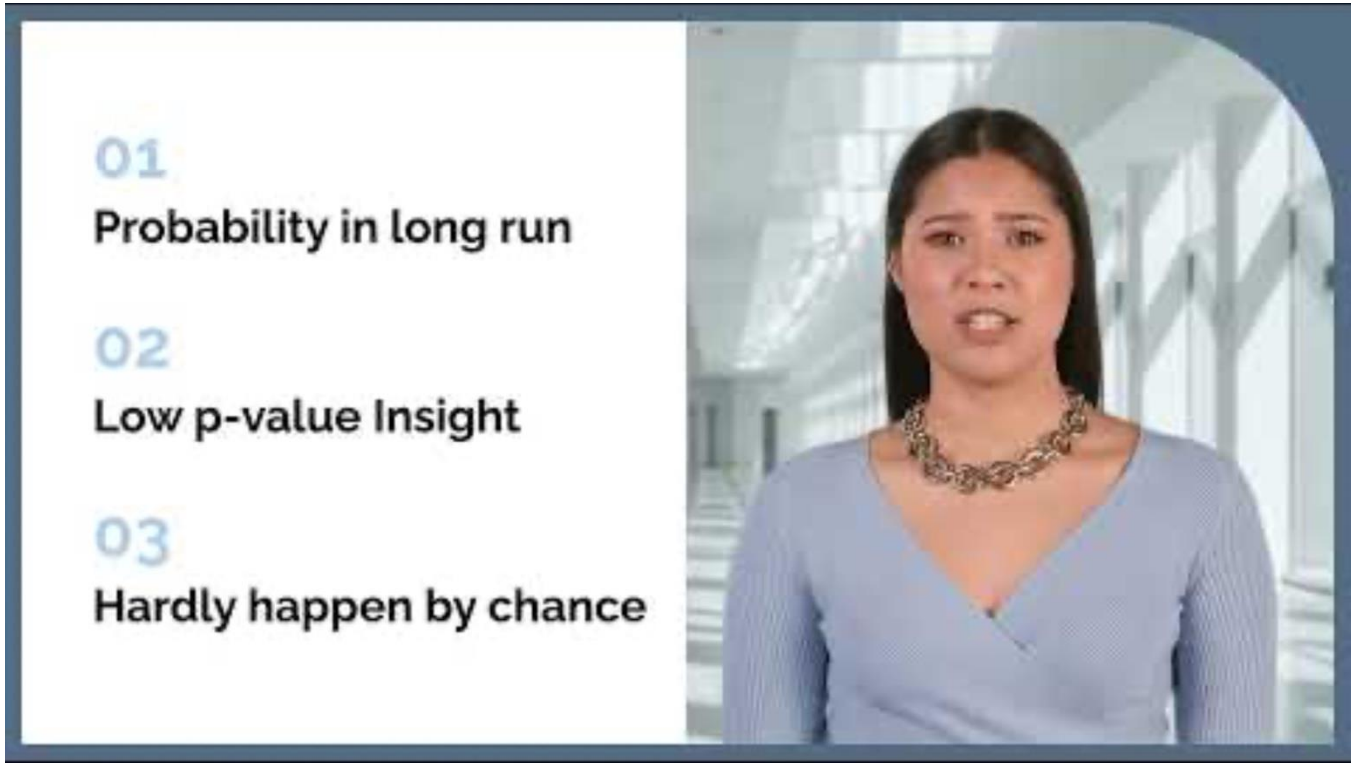
Apply to teaching

- Video lectures using myself as the avatar (Example 2):
- https://www.youtube.com/watch?v=c2QJ9sKb9RA


Significance Levels Debate
Fisher and Rosnow's views on significance levels.

Apply to teaching

- Video lectures using AI avatar:
- https://www.youtube.com/watch?v=iBhzne6xmMU



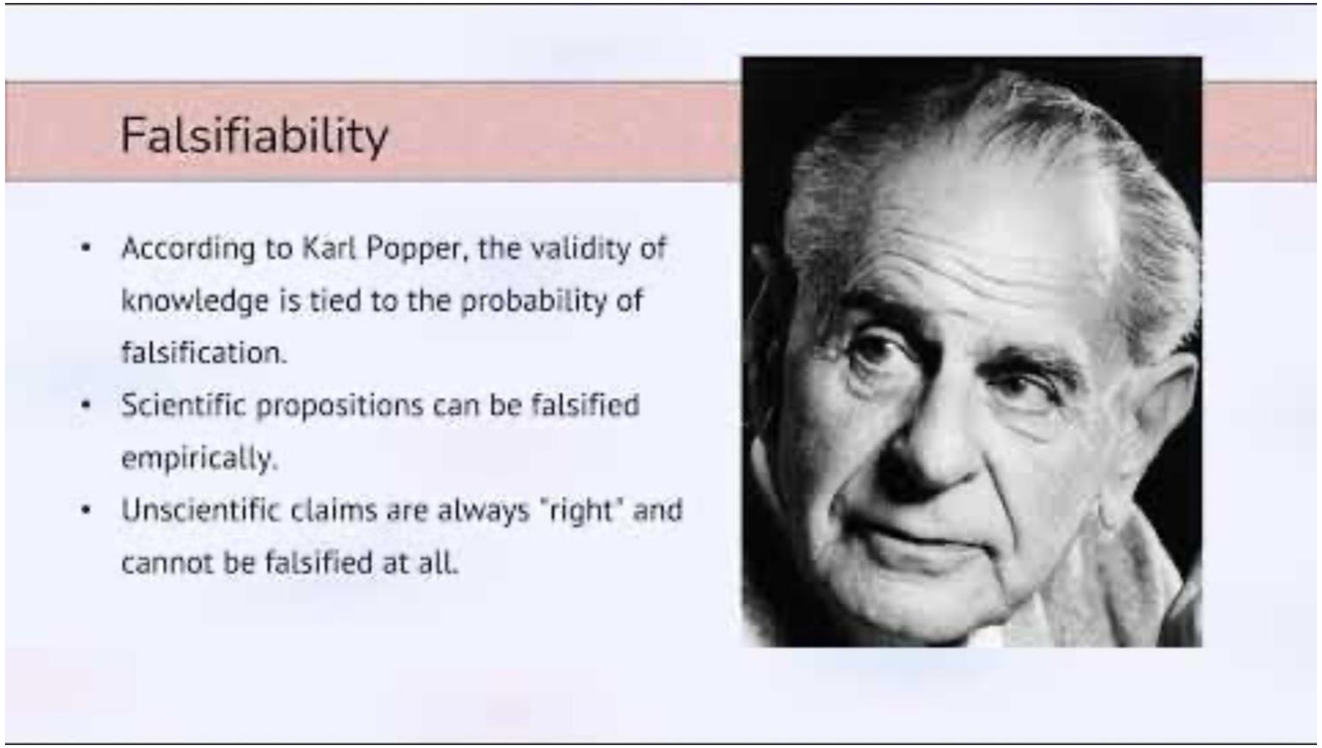
01
Probability in long run

02
Low p-value Insight

03
Hardly happen by chance

Apply to teaching

- Video lectures made by converting PowerPoint sides:
- https://www.youtube.com/watch?v=_GnVvuMnMHI



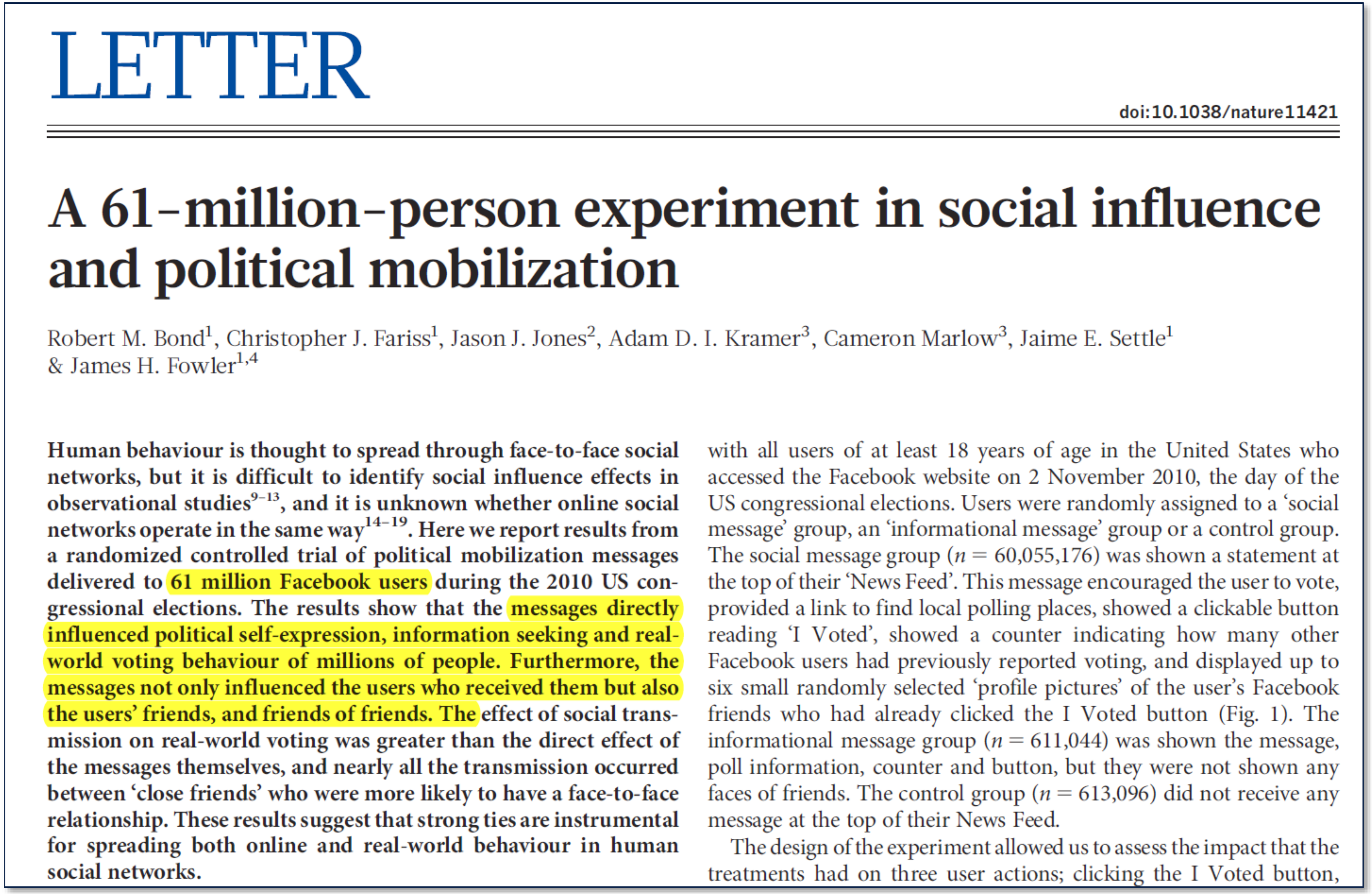
Falsifiability

- According to Karl Popper, the validity of knowledge is tied to the probability of falsification.
- Scientific propositions can be falsified empirically.
- Unscientific claims are always "right" and cannot be falsified at all.

Apply to teaching

- Video lectures can be easily translated into another language (e.g., Chinese, German, French, Spanish…etc.):
- https://www.youtube.com/watch?v=s76Bzc8kwfs

Apply to research: Limitations of classical statistics



LETTER

doi:10.1038/nature11421

A 61-million-person experiment in social influence and political mobilization

Robert M. Bond[1], Christopher J. Fariss[1], Jason J. Jones[2], Adam D. I. Kramer[3], Cameron Marlow[3], Jaime E. Settle[1] & James H. Fowler[1,4]

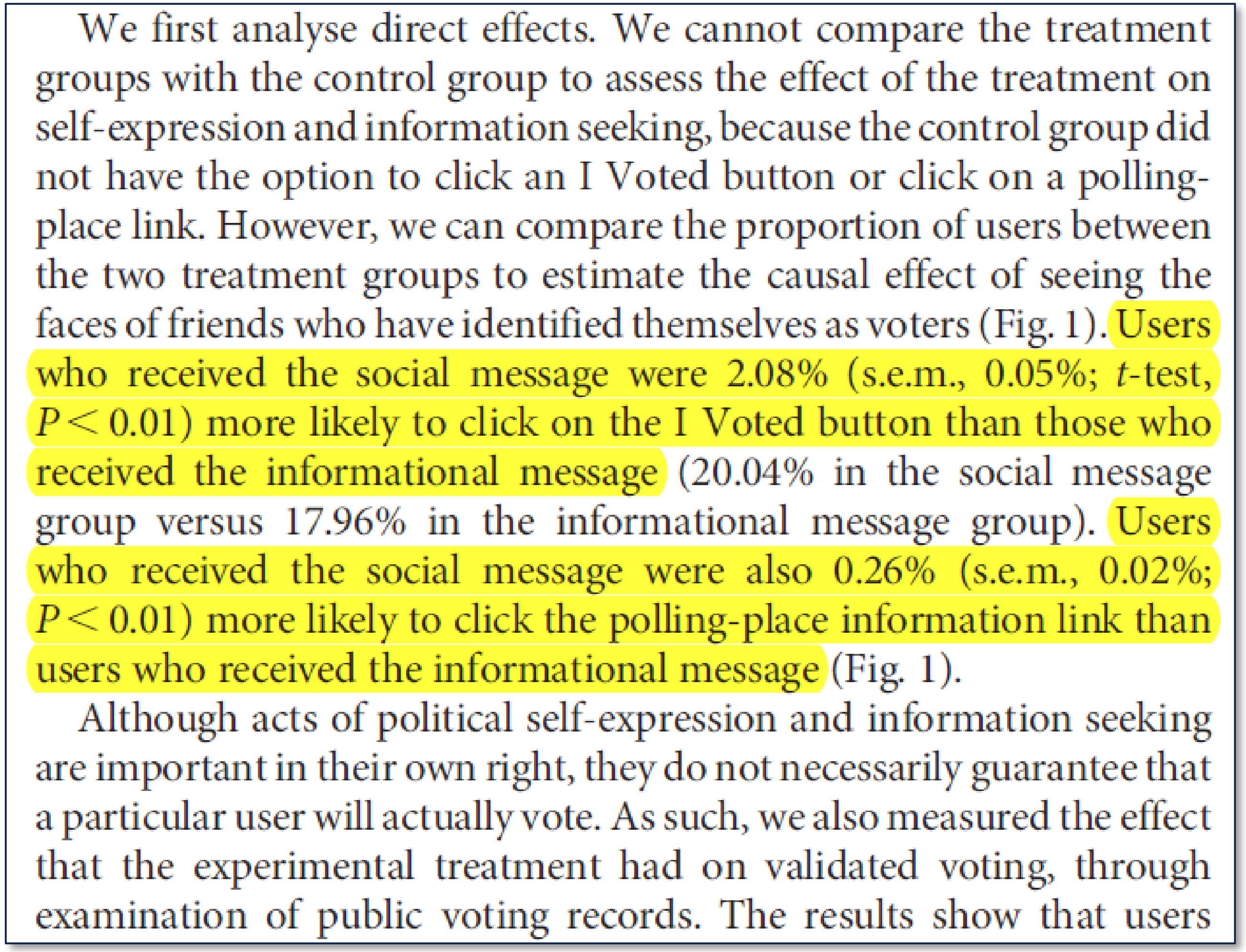
Human behaviour is thought to spread through face-to-face social networks, but it is difficult to identify social influence effects in observational studies[9–13], and it is unknown whether online social networks operate in the same way[14–19]. Here we report results from a randomized controlled trial of political mobilization messages delivered to 61 million Facebook users during the 2010 US congressional elections. The results show that the messages directly influenced political self-expression, information seeking and real-world voting behaviour of millions of people. Furthermore, the messages not only influenced the users who received them but also the users' friends, and friends of friends. The effect of social transmission on real-world voting was greater than the direct effect of the messages themselves, and nearly all the transmission occurred between 'close friends' who were more likely to have a face-to-face relationship. These results suggest that strong ties are instrumental for spreading both online and real-world behaviour in human social networks.

with all users of at least 18 years of age in the United States who accessed the Facebook website on 2 November 2010, the day of the US congressional elections. Users were randomly assigned to a 'social message' group, an 'informational message' group or a control group. The social message group (n = 60,055,176) was shown a statement at the top of their 'News Feed'. This message encouraged the user to vote, provided a link to find local polling places, showed a clickable button reading 'I Voted', showed a counter indicating how many other Facebook users had previously reported voting, and displayed up to six small randomly selected 'profile pictures' of the user's Facebook friends who had already clicked the I Voted button (Fig. 1). The informational message group (n = 611,044) was shown the message, poll information, counter and button, but they were not shown any faces of friends. The control group (n = 613,096) did not receive any message at the top of their News Feed.

The design of the experiment allowed us to assess the impact that the treatments had on three user actions; clicking the I Voted button,

Apply to research: Limitations of classical statistics

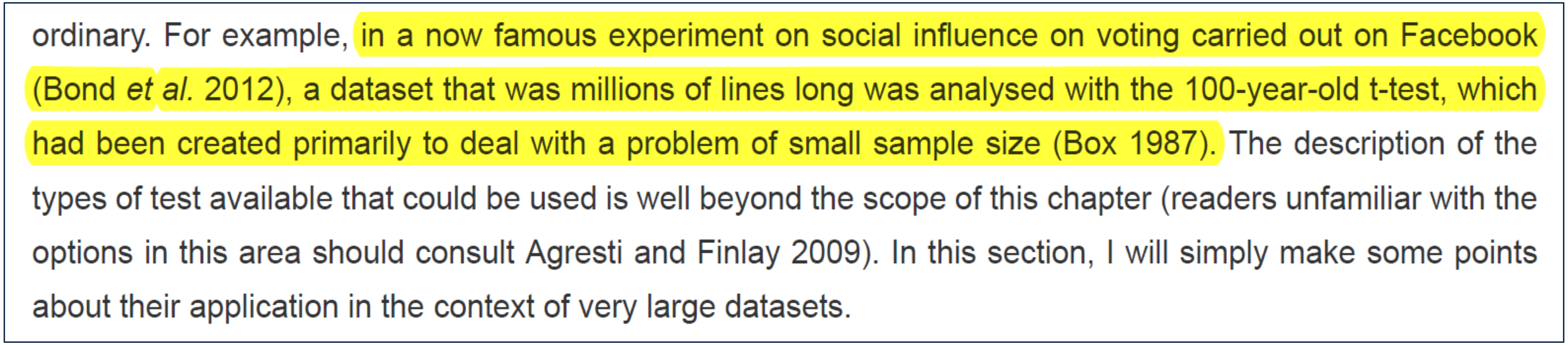
- Despite availability of machine learning, many researchers still use classical statistics.
- Running t-tests with 61M observations.



We first analyse direct effects. We cannot compare the treatment groups with the control group to assess the effect of the treatment on self-expression and information seeking, because the control group did not have the option to click an I Voted button or click on a polling-place link. However, we can compare the proportion of users between the two treatment groups to estimate the causal effect of seeing the faces of friends who have identified themselves as voters (Fig. 1). Users who received the social message were 2.08% (s.e.m., 0.05%; *t*-test, $P < 0.01$) more likely to click on the I Voted button than those who received the informational message (20.04% in the social message group versus 17.96% in the informational message group). Users who received the social message were also 0.26% (s.e.m., 0.02%; $P < 0.01$) more likely to click the polling-place information link than users who received the informational message (Fig. 1).

Although acts of political self-expression and information seeking are important in their own right, they do not necessarily guarantee that a particular user will actually vote. As such, we also measured the effect that the experimental treatment had on validated voting, through examination of public voting records. The results show that users

Apply to research: Limitations of classical statistics

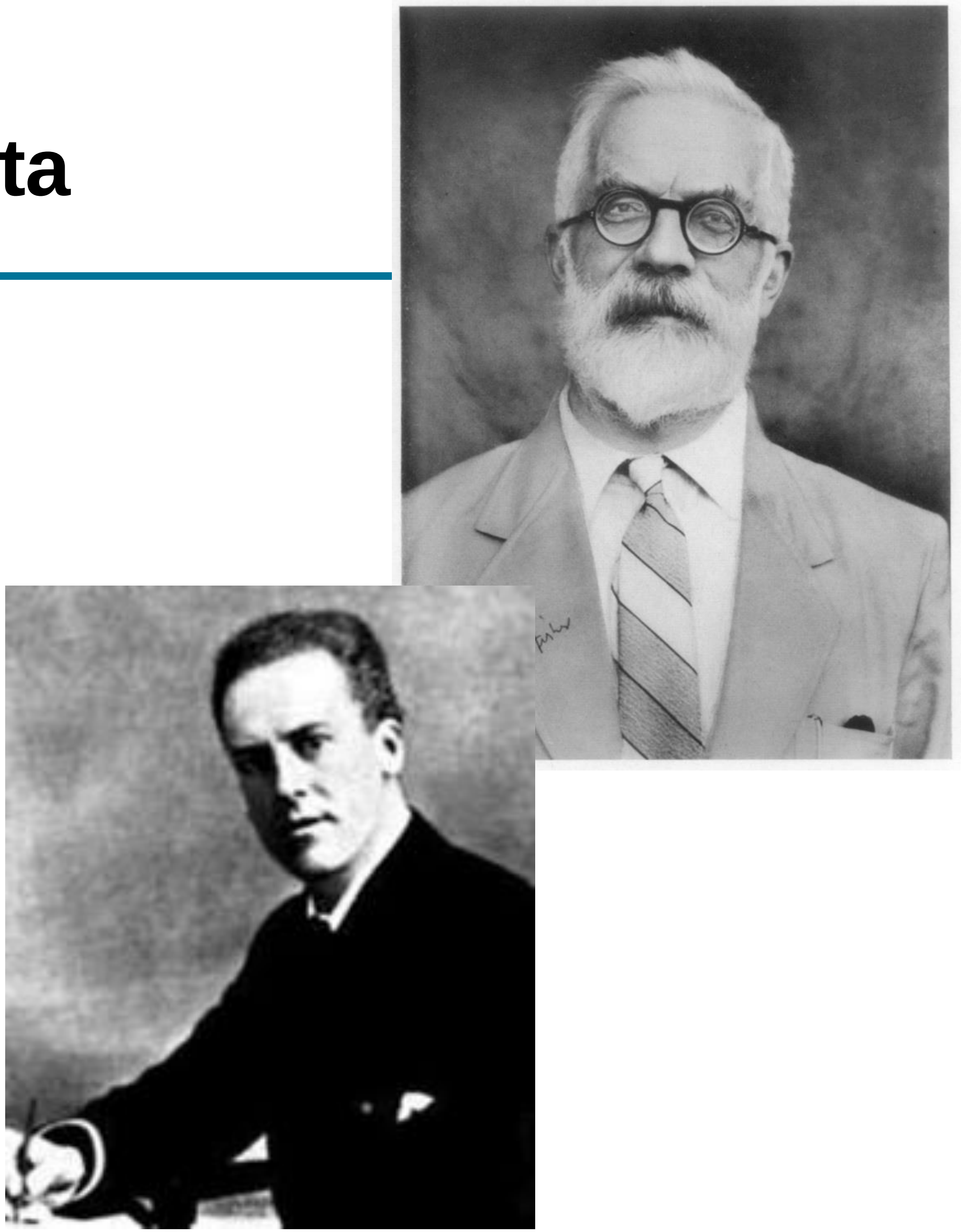


ordinary. For example, in a now famous experiment on social influence on voting carried out on Facebook (Bond *et al.* 2012), a dataset that was millions of lines long was analysed with the 100-year-old t-test, which had been created primarily to deal with a problem of small sample size (Box 1987). The description of the types of test available that could be used is well beyond the scope of this chapter (readers unfamiliar with the options in this area should consult Agresti and Finlay 2009). In this section, I will simply make some points about their application in the context of very large datasets.

- Bright, J. (2017). Big social science: Doing big data in the social sciences. In Nigel G. Fielding, Raymond M. Lee & Grant Blank (Eds.). *The SAGE handbook of online research methods* (pp. 125-139). Sage. DOI: https://dx.doi.org/10.4135/9781473957992

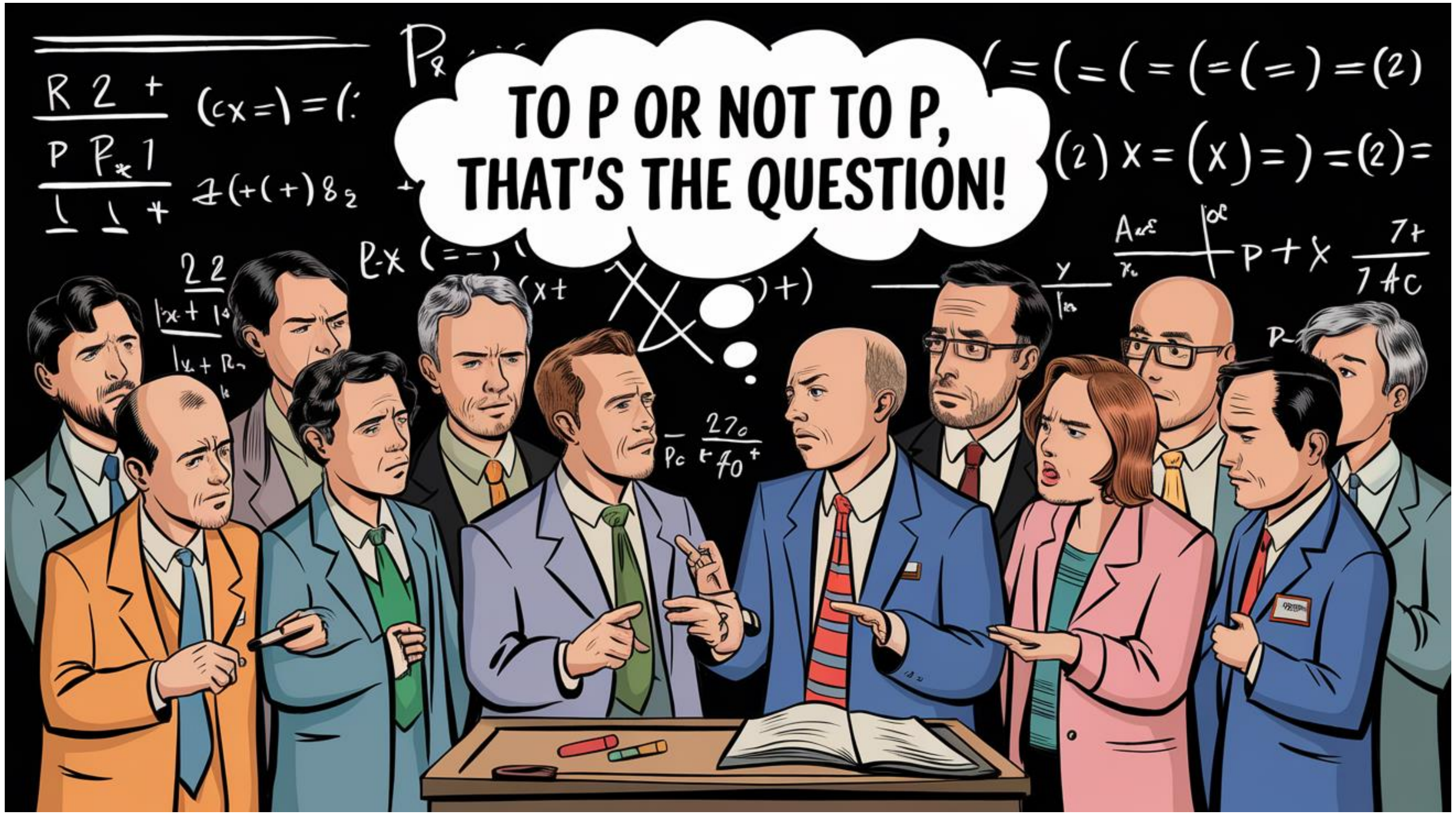
Apply to research: Structured data

- Classical statistics, also known as the frequentist school, was developed in the late 19th and early 20th century for small-sample studies.
- In order to infer from a small sample to a wider population, researchers count on theoretical, hypothetical sampling distributions.
- The decision is made by comparing the test statistics against the sampling distribution.



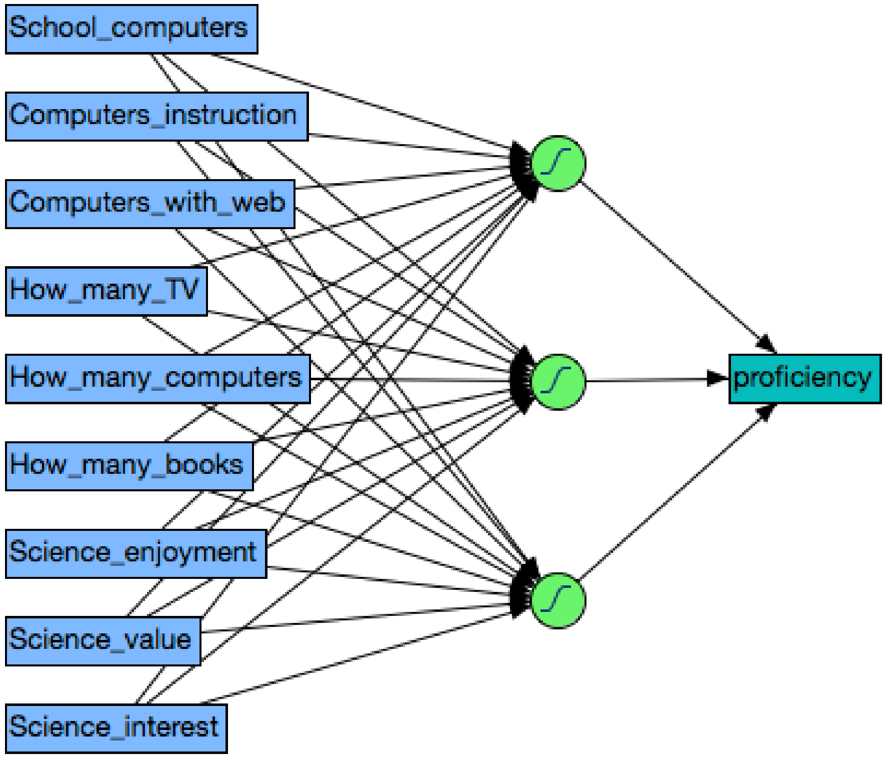
Apply to research: Structured data

- <span>**P value**</span>: If I repeat the same study over and over in the long run, what is the chance that I can observe the test statistics obtained from the sample?



Apply to research: Structured data

- However, is it meaningful to ask this hypothetical question when you have access to **actual, big data**?
- Data analytics utilizing AI and machine learning is a **paradigm shift**. The key is **pattern recognition**: learning the pattern of the data and validating it iteratively.
- Powerful algorithms: Ensemble methods, neural networks. These methods work well with medium and small data, too.



Apply to research: Structured data

- Rethinking data analysis in the era of big data:
- https://www.youtube.com/watch?v=-SSQKpxNrl8

Massive datasets with millions of records and thousands of features (variables) are now widely available.

Apply to research: Unstructured data

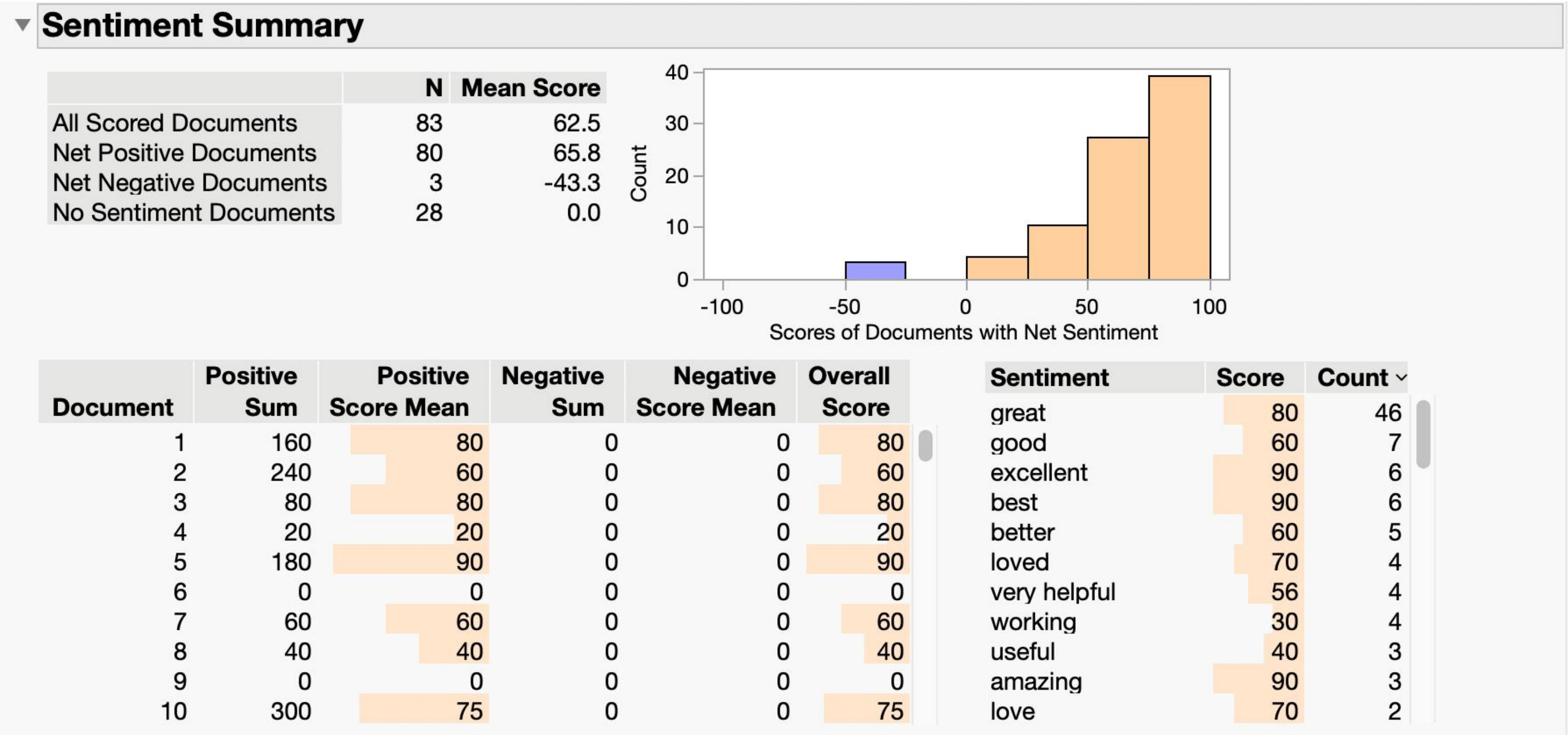
- Unstructured data (e.g., open-ended text, audio, images, video) outnumber structured data (e.g., numbers in a table)
- AI tools, such as text mining, speech recognition, and computer vision apps, are capable of analyzing unstructured data, yet these are underutilized.
- Interpreting open-ended data is subject to **Selective Perception** and **Attentional Bias**, which describe how individuals focus on information that aligns with their beliefs while disregarding conflicting evidence.
- AI approach is a powerful method to counteract such biases.

Apply to research: Unstructured data

- Unstructured data (e.g., open-ended text, audio, images, video) outnumber structured data (e.g., numbers in a table)
- AI tools, such as text mining, speech recognition, and computer vision apps, are capable of analyzing unstructured data, yet these are underutilized.
- Interpreting open-ended data is subject to **Selective Perception** and **Attentional Bias**, which describe how individuals focus on information that aligns with their beliefs while disregarding conflicting evidence.
- AI approach is a powerful method to counteract such biases.

Apply to research: Unstructured data

- Sentiment analysis



Sentiment Summary

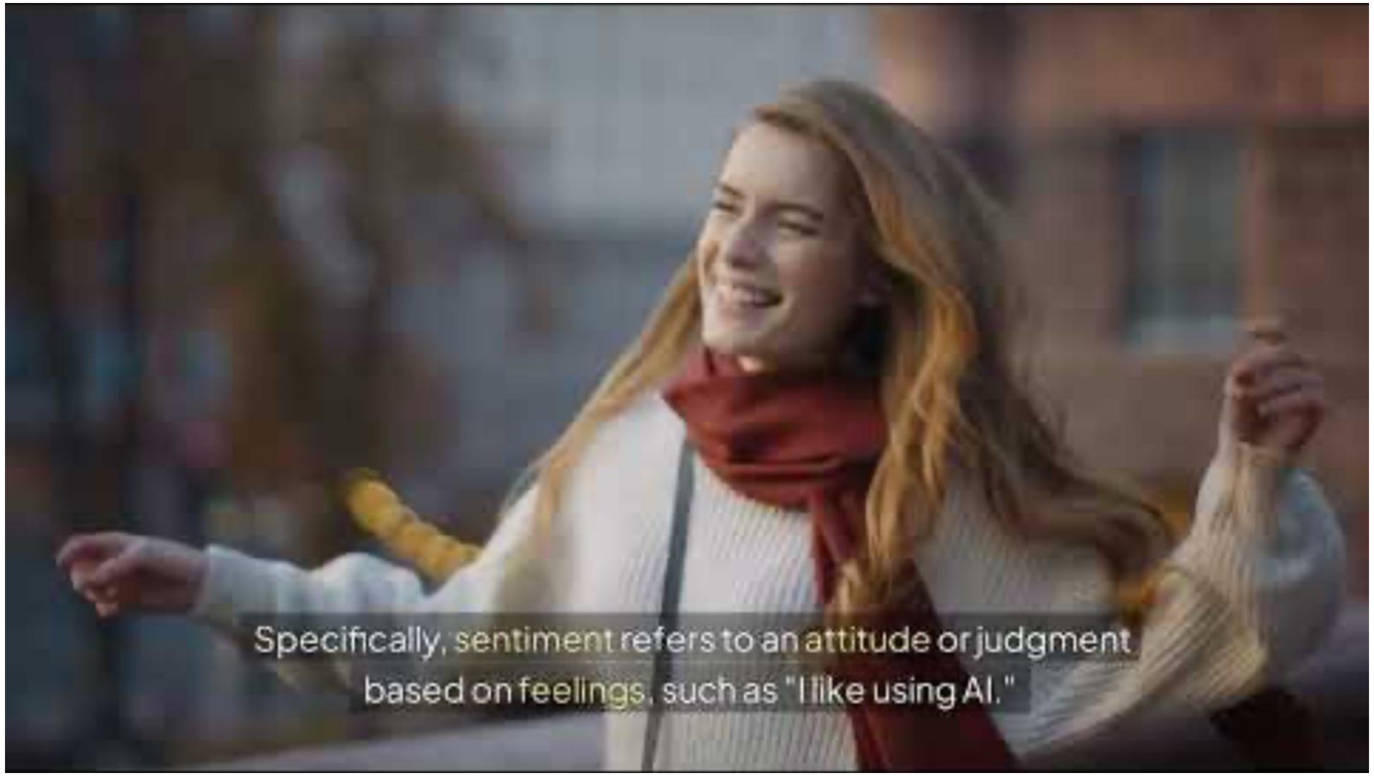
	N	Mean Score
All Scored Documents	83	62.5
Net Positive Documents	80	65.8
Net Negative Documents	3	-43.3
No Sentiment Documents	28	0.0

Document	Positive Sum	Positive Score Mean	Negative Sum	Negative Score Mean	Overall Score
1	160	80	0	0	80
2	240	60	0	0	60
3	80	80	0	0	80
4	20	20	0	0	20
5	180	90	0	0	90
6	0	0	0	0	0
7	60	60	0	0	60
8	40	40	0	0	40
9	0	0	0	0	0
10	300	75	0	0	75

Sentiment	Score	Count
great	80	46
good	60	7
excellent	90	6
best	90	6
better	60	5
loved	70	4
very helpful	56	4
working	30	4
useful	40	3
amazing	90	3
love	70	2

Apply to research: Unstructured data

- Sentiment analysis: https://www.youtube.com/watch?v=FIkE3tkaIT0


Specifically, sentiment refers to an attitude or judgment
based on feelings, such as "I like using AI."

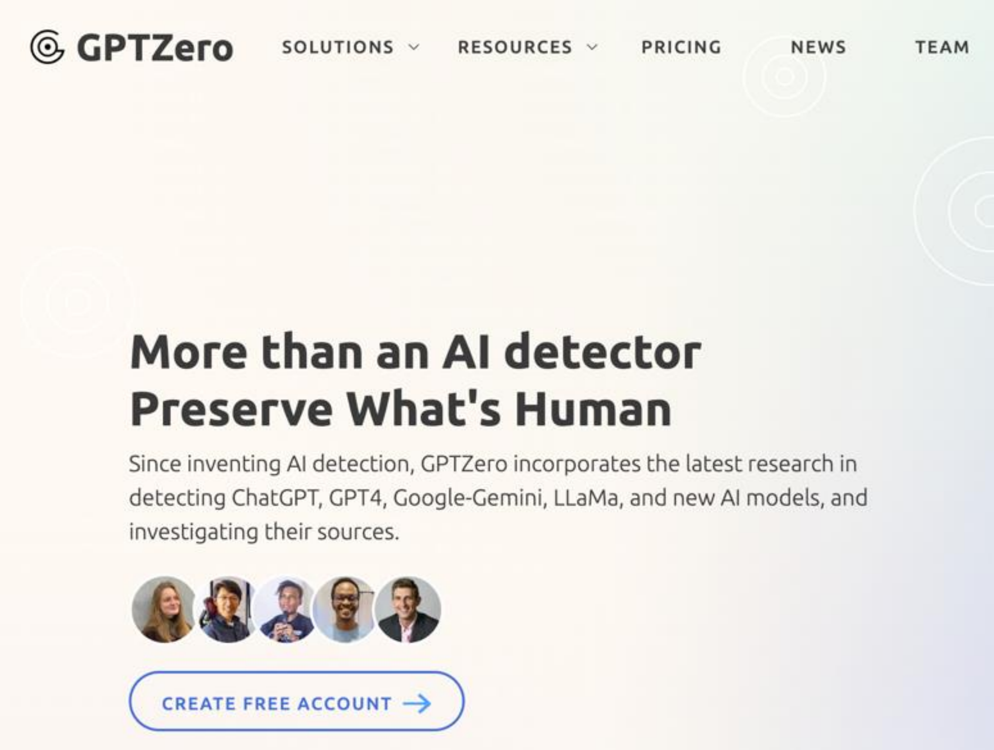
The bad: Laziness and cheating

HAWAIʻI PACIFIC UNIVERSITY

Over-reliance and cheating

- Students may over-rely on AI and become lazy.
- Cheating rates have been high for a long time. According to longitudinal surveys conducted by Stanford, long before ChatGPT was introduced, 60 to 70 percent of students have reported engaging in cheating.
- AI makes cheating easier. Students have been using AI tools to generate essays, solve math problems entirely, and even answer exam questions. A 2023 survey conducted by BestColleges indicates that 56% of university students use AI on assignments and exams. Of 200 million writing assignments reviewed by Turnitin, 3% were generated mostly by AI.

Cheating and academic dishonesty



More than an AI detector
Preserve What's Human

Since inventing AI detection, GPTZero incorporates the latest research in detecting ChatGPT, GPT4, Google-Gemini, LLaMa, and new AI models, and investigating their sources.

- Currently, there are several software applications designed to detect AI-generated text, such as GPTZero and Turnitin.
- However, they are not 100% accurate or reliable. Many AI-generated texts can be highly sophisticated and closely mimic human writing styles, making detection challenging.

Cheating and academic dishonesty

- Jared Mumm, an agricultural sciences and natural resources instructor, copied student essays into ChatGPT and asked the AI if it had generated the content. ChatGPT incorrectly claimed authorship of all the submitted papers.
- Mumm assigned failing grades to the students and temporarily withheld their diplomas.

Cheating and academic dishonesty

- Students were disheartened. Some provided evidence such as Google Docs timestamps to prove their work was original, but Mumm initially disregarded this proof, responding with "I don't grade AI bullshit" in the grading software.
- At the end, Texas A&M University confirmed that no students ultimately failed the class or were banned from graduating due to this issue. The university is developing policies to address the use of AI in coursework.

Countermeasures against academic dishonesty

- Besides using AI detection tools, many universities employ a plethora of countermeasures in order to maintain academic integrity.
- **Writing process documentation**: Require students to submit outlines, drafts, and research notes along with their final papers to demonstrate their thought process and research progression.
- **In-class writing samples**: Collect in-class writing samples to compare with submitted work, looking for significant discrepancies in style or quality.
- **Oral presentations**: Have students present and defend their papers orally, which can reveal if they truly understand the content.

Countermeasures against academic dishonesty

- **Custom, specific assignments**: Design assignments, tests, and exams that require personal experiences, specific class discussions, or unique datasets that would be difficult for AI or paper mills to replicate.
- **Continuous assessment**: Implement a system of continuous assessment rather than relying heavily on final papers or projects.
- **Old-fashioned exams**: Rather than taking online or take-home exams, students must complete exams in person, using pen and paper, within a designated time limit in a classroom setting.

Countermeasures against academic dishonesty

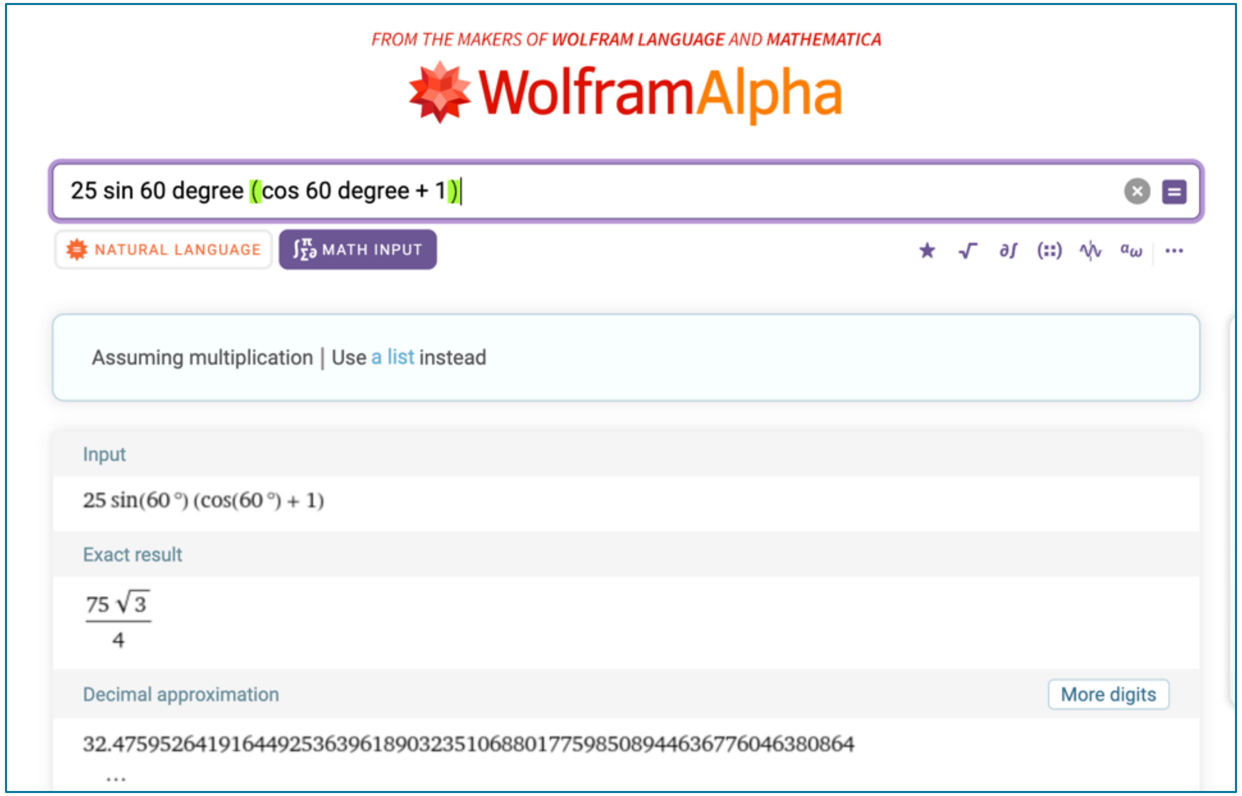
- **Education on academic integrity**: Proactively educate students about the ethical implications and potential consequences of using paper mills or AI-generated content.
- **Provide AI literacy education**: Rather than banning AI tools, teach students how to appropriately use AI for initial research, data analysis, and proofreading. Encourage critical thinking and original analysis beyond AI-generated responses.
- **Establish an AI policy**: Establish a university-wide AI policy to ensure consistency across departments and courses. Emphasize the importance of proper citation and acknowledgment when using AI-generated content.

Countermeasures against academic dishonesty

- AI literacy education and academic integrity policy pertaining to AI must clearly define legitimate use and illegitimate misuse of AI:
- Inputting homework questions into AI and submitting the generated responses as their own work is generally considered academic dishonesty.
- Using AI tools for summarizing articles, collecting information (e.g., ChatGPT, Bing), proofreading (e.g., Grammarly, Wordtune), and generating graphics for presentation (e.g., Midjourney, DALL.E) is fine.

Gray area

- There is some gray area.
- Before AI, there have been software tools for helping students to solve math problems, such as Mathematica, WolframAlpha, SymboLab, and Desmos.
- After entering a math problem, the software app can show both the result and the steps in a second.



25 sin 60 degree (cos 60 degree + 1)

Assuming multiplication | Use a list instead

Input

$25 \sin(60°) (\cos(60°) + 1)$

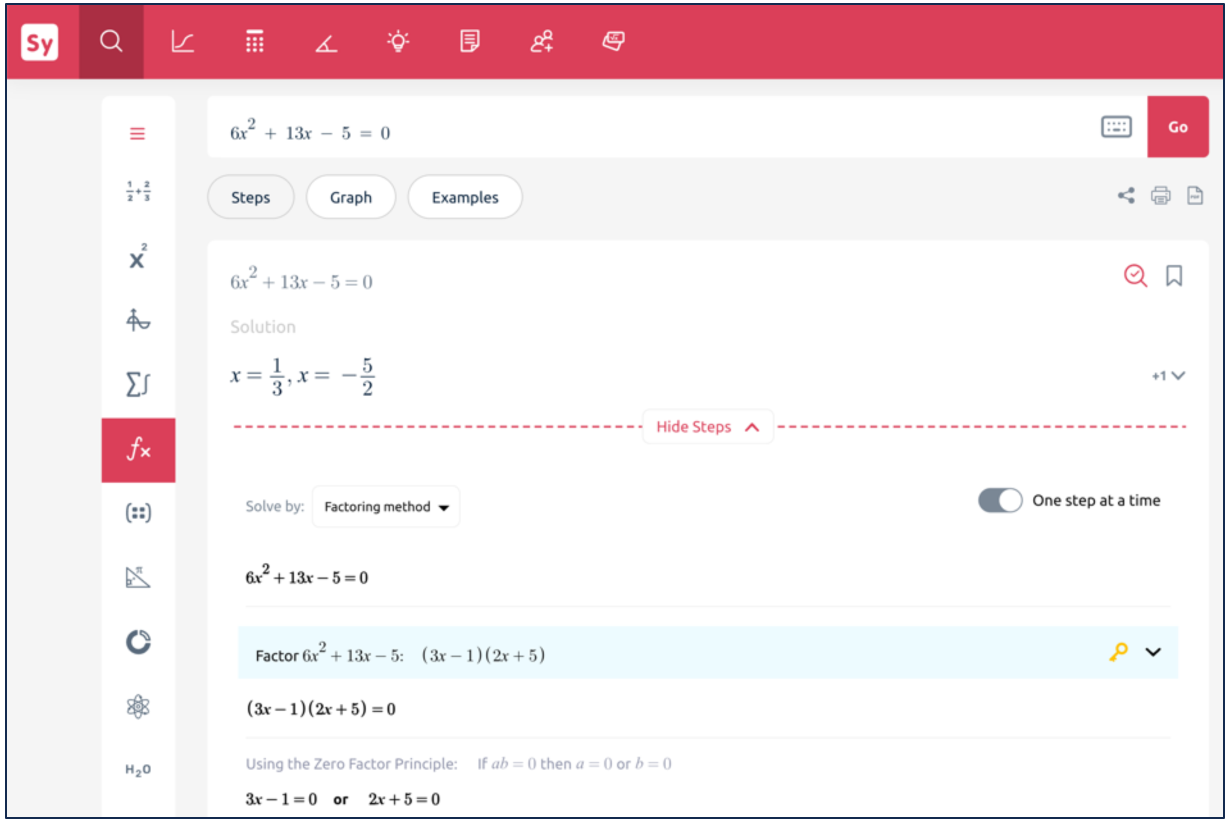
Exact result

$\frac{75\sqrt{3}}{4}$

Decimal approximation

32.475952641916449253639618903235106880177598508944636776046380864
…

Gray area



$6x^2 + 13x - 5 = 0$

$6x^2 + 13x - 5 = 0$

Solution

$x = \frac{1}{3}, x = -\frac{5}{2}$

+1

Solve by: Factoring method

$6x^2 + 13x - 5 = 0$

Factor $6x^2 + 13x - 5$: $(3x-1)(2x+5)$

$(3x-1)(2x+5) = 0$

Using the Zero Factor Principle: If $ab = 0$ then $a = 0$ or $b = 0$

$3x - 1 = 0$ or $2x + 5 = 0$

- Some math professors do accept and even encourage the use of tools like WolframAlpha, SymboLab, and Desmos, as long as Students understand the underlying concepts to interpret and apply the results.
- It is similar to an accountant to use TurboTax.

Gray area

- However, the use of AI for math problem-solving is often viewed differently for several reasons.
  - **Broader capability**: AI can often solve a wider range of problems and explain steps in natural language, potentially replacing more of the student's own thought process.
  - **Lack of transparency**: It's not always clear how AI arrives at its solutions, unlike traditional math software with known algorithms.
  - **Rapid evolution**: The capabilities of AI are expanding quickly, making it challenging for educators to adapt their teaching and assessment methods.

Gray area

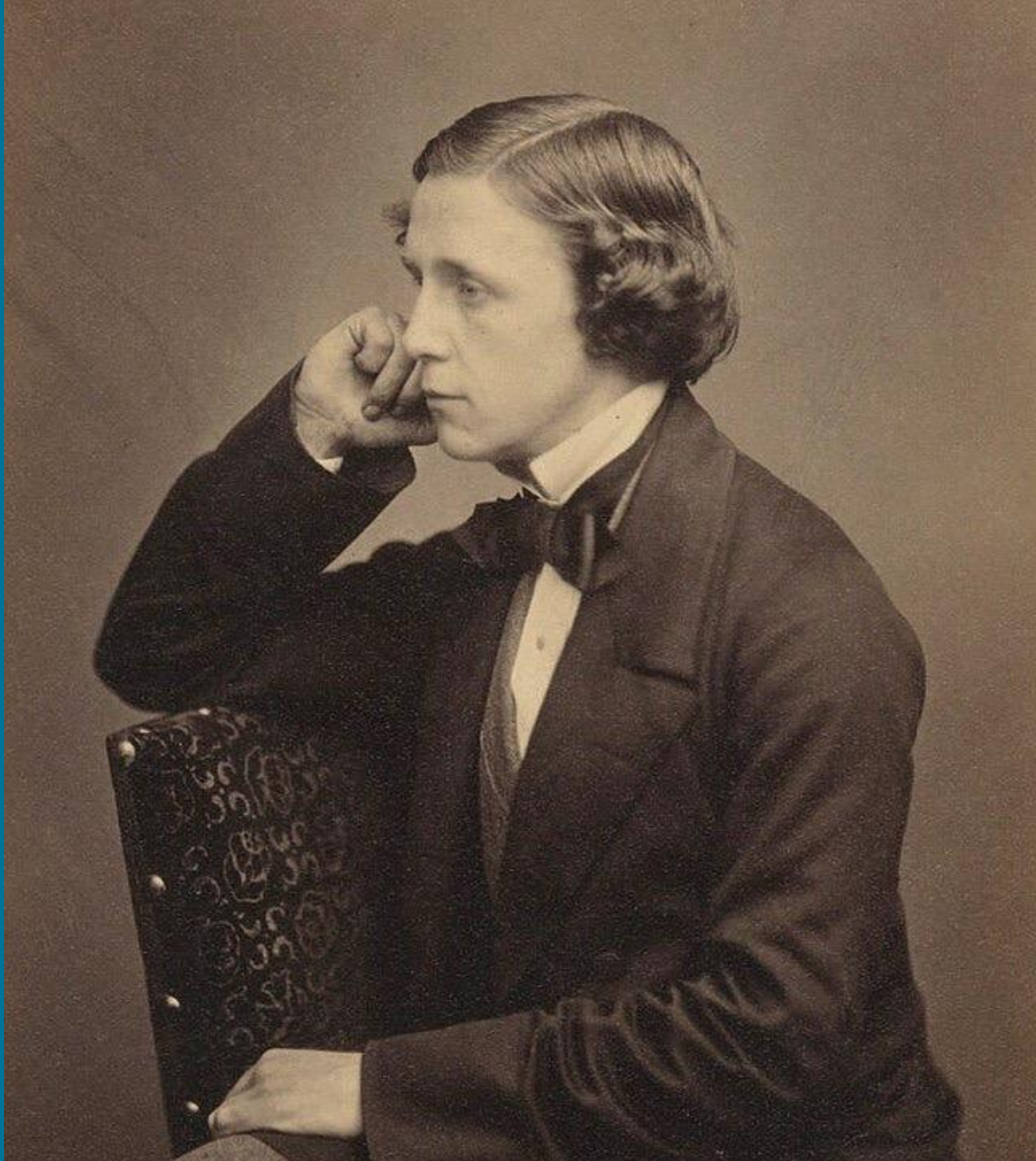
- The line between "acceptable tool" and "cheating" in using AI to solve math problems is not clear-cut. It depends on:
  - The educational level (grade school, high school vs. university)
  - The course objectives (learning concepts vs. learning procedures)
  - How the AI is being used (as a learning assistant vs. a substitute for understanding)

The Ugly: Disinformation and deepfake

HAWAIʻI PACIFIC UNIVERSITY

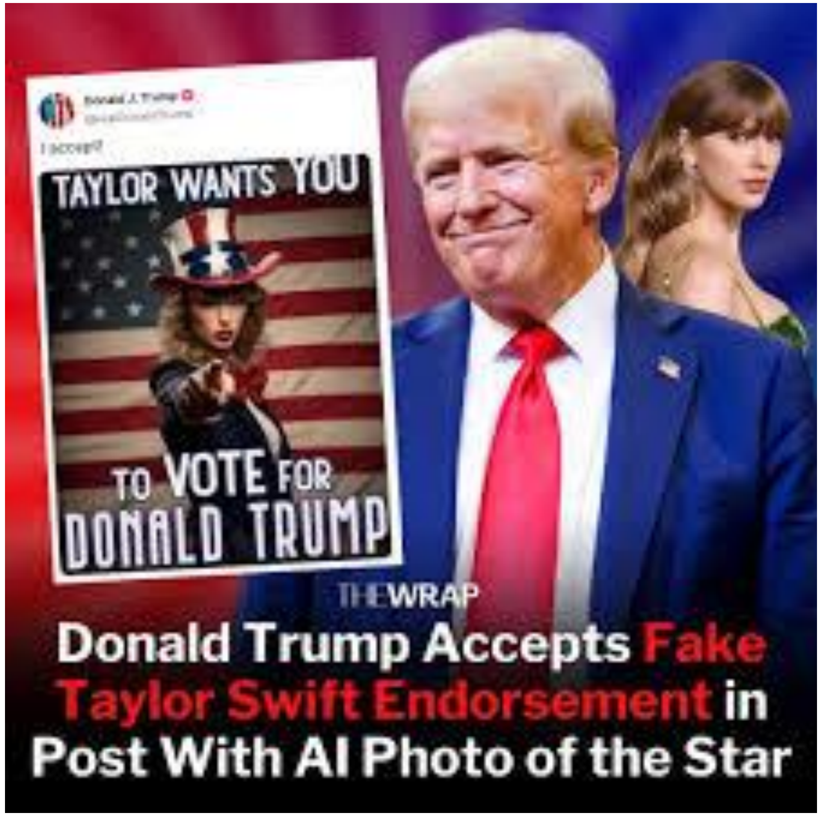
Lewis Carroll (from Alice in Wonderland)

"I am not crazy; my reality is just different from yours."



Alarming situation

- On August 20, 2024, Donald Trump posted on his social media platform *Truth Social*, sharing AI-generated images that falsely suggested Taylor Swift had endorsed him for the 2024 presidential election.
- Trump wrote "I accept!" alongside these images, implying he was accepting Swift's endorsement.
- Taylor Swift has not endorsed any candidate for the 2024 election.



Donald Trump Accepts Fake Taylor Swift Endorsement in Post With AI Photo of the Star

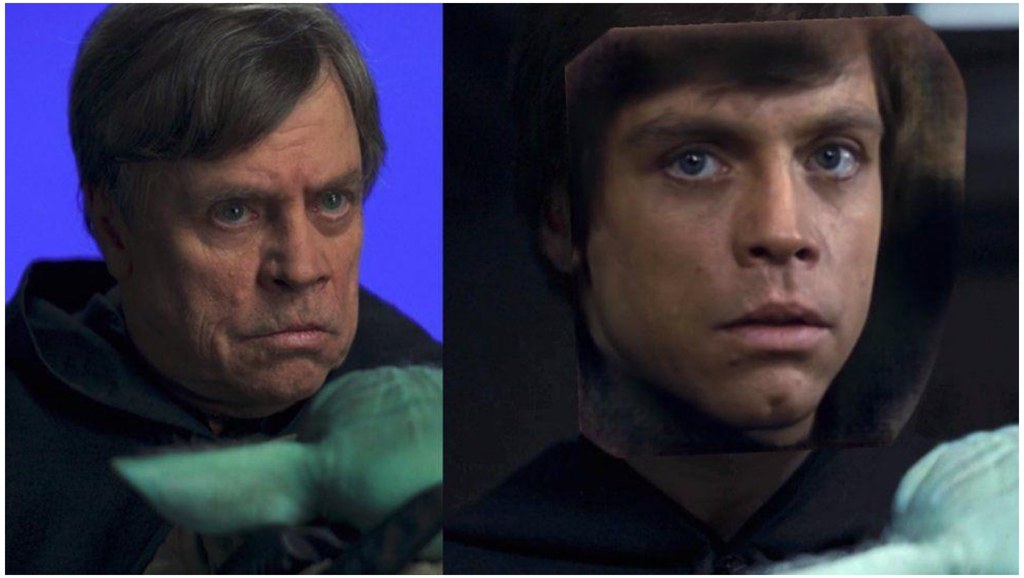
Alarming situation

- In 2024 *WIRED* magazine has launched the *AI Elections Project*, aimed at monitoring the use of generative AI in over 60 countries' elections.
- The project has uncovered alarming instances of AI-generated content being used for political propaganda and disinformation.
- In India and Indonesia, deepfake videos of deceased leaders have surfaced, appearing to endorse their political successors.
- In South Africa, the rapper Eminem's likeness has been used to endorse opposition parties without his consent.
- A deepfake of President Joe Biden has been circulating, urging voters in New Hampshire to stay home on election day.

Alternate reality and parallel universe

- The spread of disinformation and misinformation is detrimental to democracy as people are unable to make informed decisions. As Bill Maher said, “Democracy dies in dumbness.”
- People who are constantly exposed to disinformation and misinformation are trapped in their alternate reality, parallel universe, or fantasy. Probably they may say, “I am not crazy; my reality is just different from yours."

Deepfake in Hollywood



- Hollywood has started using deepfake technology in movie and TV production.

- In some sense, the AI-generated instructional video shown before is a form of deepfake.

Identity theft and impersonation

- In May 2024 an employee in the finance department of a company received a message supposedly from the UK-based CFO about a "confidential transaction”.
- Initially skeptical, the employee was convinced after participating in a **video conference call**. The call included deepfake versions of the CFO and other company executives.
- Over a week, the employee made 15 transfers totaling $25 million to five Hong Kong bank accounts.
- The fraud was discovered when the employee contacted company headquarters, but it was too late.

Generative adversarial Network (GAN)

- DeepNude is a specific form of deepfake technology. DeepNude refers to a particular application or software that was designed to digitally "undress" images of women, creating fake nude images from ordinary photos.
- DeepNude, like many other deepfake technologies, is derived from techniques related to Generative Adversarial Networks (GANs), which were invented by Ian Goodfellow and his colleagues in 2014.

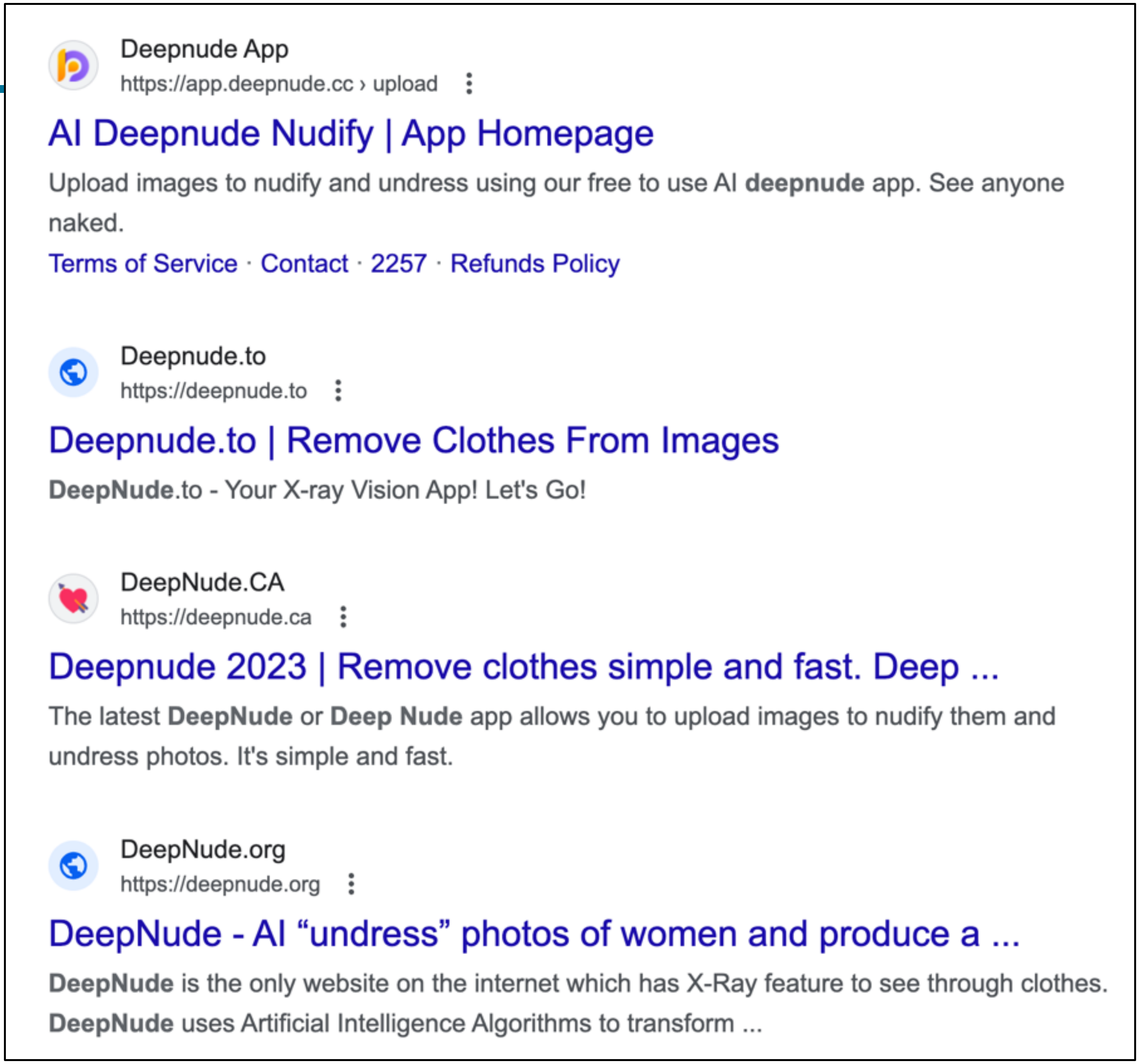
Images generated by GAN

DeepNude

- In June 2019, a software app called DeepNude was released. This application uses GAN to remove women's clothes in images within a split second.
- There are paid and non-paid versions of the app, with the paid version priced at $50.

DeepNude

- Numerous women became victims of DeepNude, causing a widespread outrage.
- Faced with public pressure, the creator of DeepNude decided to remove the app, but it is too late. Many other variations of DeepNude continue to circulate on the Internet.
- It is almost impossible to ban these websites, especially those hosted in foreign countries.



Deepnude App
https://app.deepnude.cc › upload

AI Deepnude Nudify | App Homepage

Upload images to nudify and undress using our free to use AI **deepnude** app. See anyone naked.

Terms of Service · Contact · 2257 · Refunds Policy

Deepnude.to
https://deepnude.to

Deepnude.to | Remove Clothes From Images

DeepNude.to - Your X-ray Vision App! Let's Go!

DeepNude.CA
https://deepnude.ca

Deepnude 2023 | Remove clothes simple and fast. Deep ...

The latest **DeepNude** or **Deep Nude** app allows you to upload images to nudify them and undress photos. It's simple and fast.

DeepNude.org
https://deepnude.org

DeepNude - AI “undress” photos of women and produce a ...

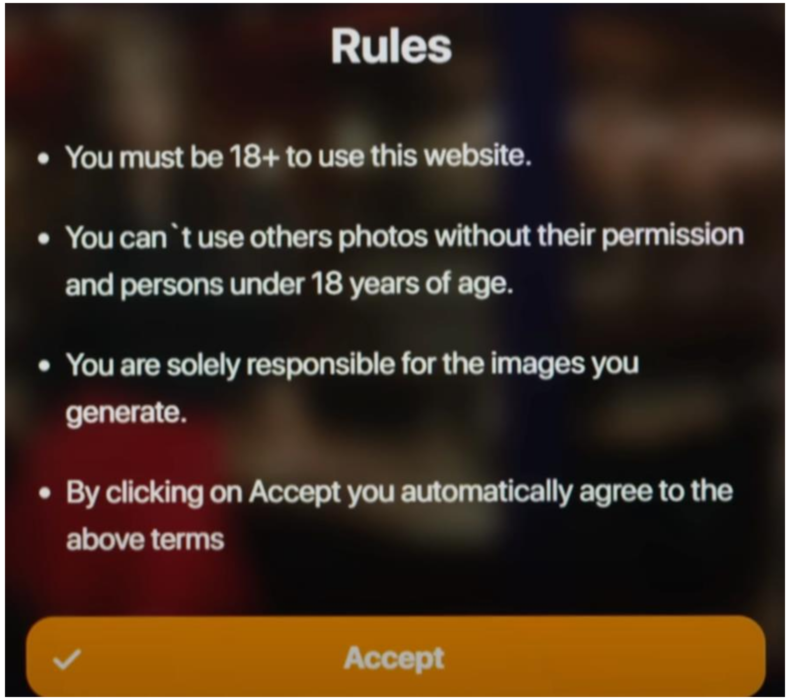
DeepNude is the only website on the internet which has X-Ray feature to see through clothes. **DeepNude** uses Artificial Intelligence Algorithms to transform ...

The curse of open source

- There are many copycats and similar applications to DeepNude, even after the original was pulled.
  - **Open source algorithms**: The core algorithms used in DeepNude, like Generative Adversarial Networks (GANs), are open source and widely available. This makes it relatively easy for others to recreate similar functionality.
  - **Public datasets**: Large datasets of images used to train these models are often publicly available.
  - **Distributed code**: Even though the original DeepNude was taken down, copies of its code were distributed and remain available online.

Westfield Highschool

- In 2023, male students at Westfield High School in Westfield, New Jersey, used DeepNude websites to create explicit, fake images of their female classmates by manipulating innocent photos sourced from social media and school events.
- The Westfield Police Department was notified and launched an investigation into the matter, but no one was arrested.

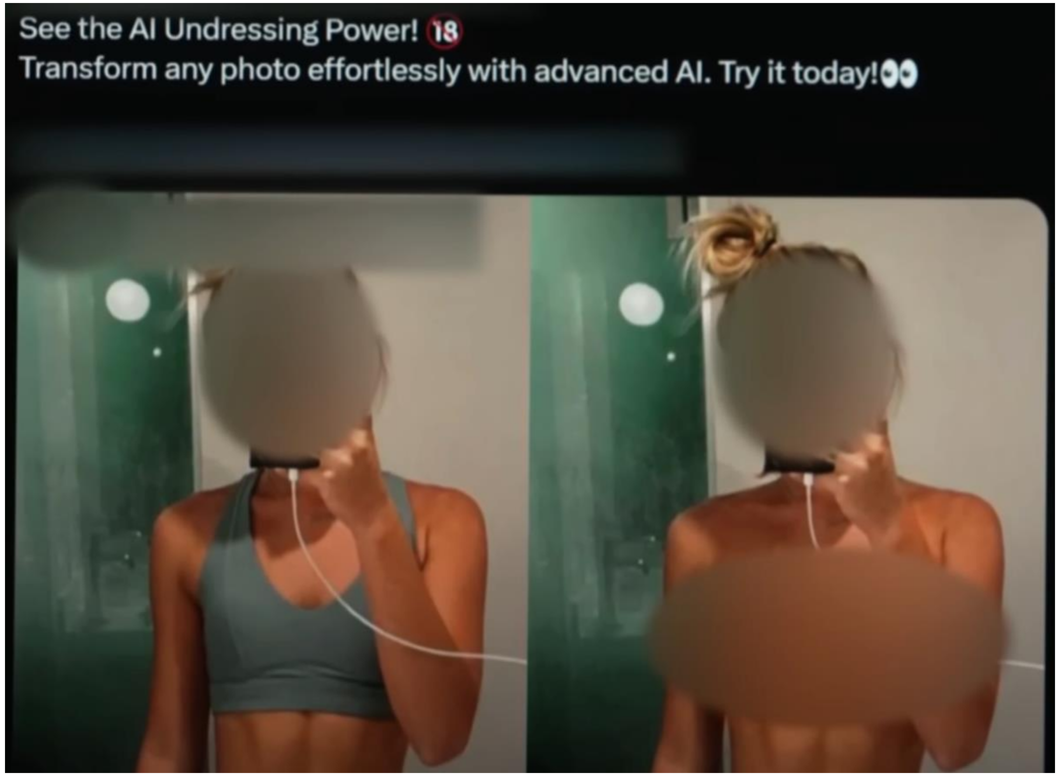


Rules

- You must be 18+ to use this website.
- You can`t use others photos without their permission and persons under 18 years of age.
- You are solely responsible for the images you generate.
- By clicking on Accept you automatically agree to the above terms

Westfield Highschool

- In June 2024, the Westfield Board of Education updated its policies to include language addressing the use of AI in the context of harassment, intimidation, and bullying (HIB).
- https://www.youtube.com/watch?v=JS86nW40Jq4



See the AI Undressing Power! 🔞
Transform any photo effortlessly with advanced AI. Try it today!👀

Concluding remarks

Summary

- The **“good”**: As a powerful tool, AI can improve efficiency and effectiveness in teaching and research.
- AI bot: Non-judgmental, always friendly, personal tutor.
- AI and machine learning, which aims to pattern recognition, can overcome limitations of classical statistics.
- AI can go beyond traditional structured data. AI can mitigate human bias while analyzing unstructured data.

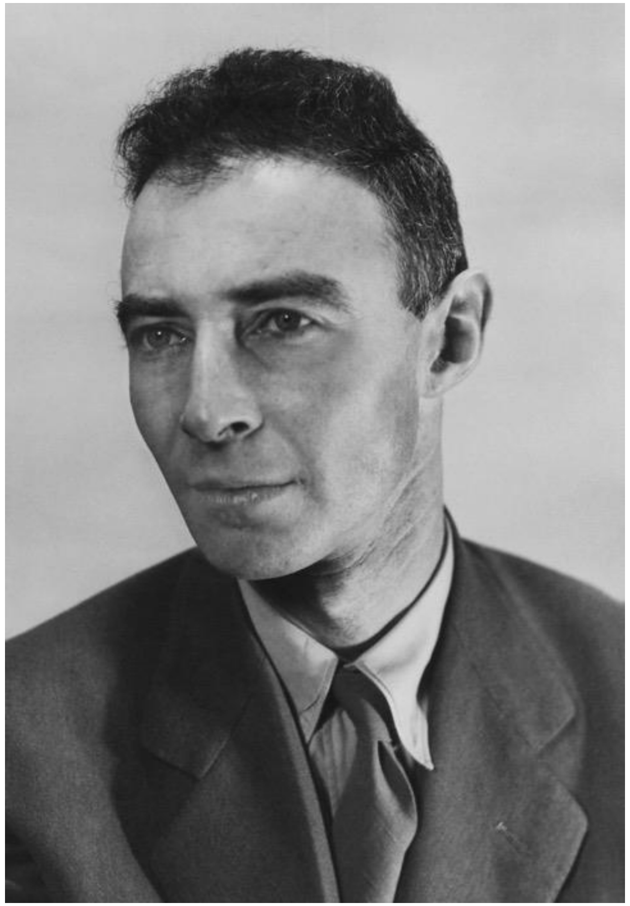
Summary

- The **“bad”** part is a gray area. The outcomes largely depend on how effectively we guide students in utilizing AI responsibly. Without proper guidance, students risk misusing AI and missing valuable learning opportunities. This highlights the need to revamp our current teaching and assessment methods.
- The **“ugly”** side of AI, such as disinformation and DeepNude, poses significant dangers. It is imperative to educate students about these risks to help them discern information and avoid abusing AI technology.

“Now I am become Death, the destroyer of worlds”.

- Robert Oppenheimer

Oppenheimer and A-bomb



- When Oppenheimer developed the atomic bomb, his aim was to bring the war to a swift end. He didn't anticipate the subsequent arms race between the U.S. and the USSR.
- Once he recognized that nuclear weapons could threaten the very survival of civilizations, he opposed the development of the hydrogen bomb. His loyalty was questioned, and his security clearance was ultimately revoked.

Our responsibility?

- This situation mirrors the invention of GANs, where humans struggle to control the negative consequences of our own creation.
- The ethical dilemma is as follows: when we develop and deploy something powerful, are we obligated to consider potential harmful outcomes and implement safety measures to mitigate them?
- As faculty, it is our responsibility to take proactive steps. The future is in our hands!

More info

- Email1: cayu@hpu.edu
- Email 2: chonghoyu@gmail.com
- Blog: https://creative-wisdom.me/blog
- YouTube: https://www.youtube.com/@datafrontiers
- Research1: https://www.creative-wisdom.com/pub/pub.html
- Research2: https://scholar.google.com/citations?user=mdGny3EAAAAJ&hl=en

HAWAIʻI PACIFIC UNIVERSITY

Disclosure: AI tools are utilized for initial research, proofreading, and generating images for the presentation. The key ideas originates from the author.