

M-Lab CARTE AI Workshop 2025

AI Ethics and Safety

Presented by Nakul Upadhyia

Ethical Concerns

Bias

Data Privacy

Crime &
Misuse

Intellectual
Property

Environmental
Impact

Socio-
Economic
Consequences

Where AI bias comes from

- Bias in AI can arise in many different stages of the process, but can be broadly sorted into three categories:

1. Data bias

- Where the information used to train an AI model is unrepresentative or incomplete

2. Algorithmic bias

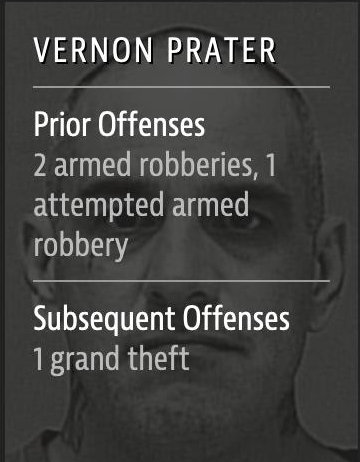

- When the model itself learns incorrect assumptions about the problem being addressed

3. User bias

- When the people using an AI system introduce their own biases

Data Bias

- Data is possibly the most common source of bias in AI
- When given a skewed understanding of the world, the best a model can do is replicate that understanding
- Famous example: COMPAS system

Two Petty Theft Arrests	
	
VERNON PRATER	BRISHA BORDEN
Prior Offenses 2 armed robberies, 1 attempted armed robbery	Prior Offenses 4 juvenile misdemeanors
Subsequent Offenses 1 grand theft	Subsequent Offenses None
LOW RISK 3	HIGH RISK 8

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Algorithmic bias

- Turnitin builds software to identify plagiarism in student-submitted work
- The model is effective, and has been shown to (generally) accurately identify plagiarism
- However, the sophistication of plagiarism is not evenly distributed among students
- Students with the best grasp of English have the best chance at evading detection by the algorithm!
- Even though the training data was not biased towards native English speakers, **the result is an algorithm that is more likely to flag work by non-native speakers**

User Bias

- British National Act Program – a tool created as a *proof of concept* to help evaluate possibility for British citizenship
- Immigration officers began to rely heavily on the prototype in real cases, even as immigration law changed and new practices came into prominence

```
if X is father of Peter
then X is a parent of Peter

if X is a parent of Peter
and X is a British citizen on date (3 May 1983)
then Peter has a parent
    who qualifies under 1.1 on date (3 May 1983)

    Peter was born in the U.K.
    Peter was born on date (3 May 1983)
    (3 May 1983) is after or on commencement, so
if Peter has a parent
    who qualifies under 1.1 on date (3 May 1983)
then Peter acquires British citizenship
    on date (3 May 1983) by sect. 1.1

    Peter is alive on (16 Jan 1984), so
if Peter acquires British citizenship
    on date (3 May 1983) by sect. 1.1
and (16 Jan 1984) is after or on (3 May 1983)
and not[Peter ceases to be a British citizen on date Y
    and Y is between (3 May 1983) and (16 Jan 1984)]
then Peter is a British citizen on date (16 Jan 1984) by sect 1.1
```

“The British Nationality Act as a Logic Program”
Sergot et al. 1986

Privacy

- Data privacy in the era of Big Data is already a concern.
- AI can supercharge privacy concern:
 - Data leaks from big players.
 - AI Agents leaking sensitive information.
 - AI powered surveillance.

ARTIFICIAL INTELLIGENCE | OPENAI

POISONED INVITE

It's Staggeringly Easy for Hackers to Trick ChatGPT Into Leaking Your Most Personal Data

"This is very, very bad."

By [Victor Tangemann](#) / Published Aug 7, 2025 10:30 AM EDT



ShadowLeak Zero-Click Flaw Leaks Gmail Data via OpenAI ChatGPT Deep Research Agent

Sep 20, 2025 · Ravie Lakshmanan

Artificial Intelligence / Cloud Security

— Trending News

Stealit Malware Abuses Node.js Single

Google Gemini vulnerability enables hidden phishing attacks

News

Jul 15, 2025 · 3 mins

Generative AI

Network Security

Zero-Day Vulnerabilities

PUBLIC SAFETY

Austin drops AI surveillance cameras from consideration as residents raise privacy concerns

BY LUZ MORENO-LOZANO, KUT · SEPTEMBER 25, 2025

[Facebook](#) [Twitter](#) [Email](#) [More](#)

The city of Austin is no longer considering using artificial intelligence to help catch people breaking into cars and committing other crimes at parks and greenbelts, at

Crime and Misuse

- AI models can be jailbroken and misused.
- Crime and Misuse of AI is extremely dangerous and becoming easier every day.
- How do we define and regulate this?

Manitoba

'Definitely my son's voice': Manitoba woman targeted by AI phone scam

Fraudsters using AI to mimic voices of loved ones in 'very targeted' scams: investigator



Mike Arsenault · CBC News · Posted: May 14, 2025 6:00 AM EDT | Last Updated: May 14



Record numbers of women in Scotland are victims of fake image 'revenge porn'

The Revenge Porn Helpline said that almost 30,000 women in the country are targeted every year and warned this was only the 'tip of the iceberg'

TECHNOLOGY

President Trump signs Take It Down Act, addressing nonconsensual deepfakes. What is it?




TECH

California just passed new AI and social media laws. Here's what they mean for Big Tech

PUBLISHED TUE, OCT 14 2025 7:00 AM EDT | UPDATED TUE, OCT 14 2025 7:54 AM EDT



Samantha Subin
@SAMANTHA_SUBIN

SHARE    

Intellectual Property

- **Training:** *GenAI was trained using unlicensed data.*
 - How do individuals get compensated? How much?
- **Generation:** Who owns a generated piece of work?

Business

Anthropic agrees to pay \$1.5B US to settle author class action over AI training

AI company downloaded books from pirating sites to train its chatbot, Claude

Thomson Reuters - Posted: Sep 05, 2025 5:55 PM EDT | Last Updated: September 5



from a group of authors who
rain its AI chatbot, Claude.

LIVES IN A METH LAB UNDER THE SEA

OpenAI's Sora 2 Is Generating Video of SpongeBob Cooking Meth, Highlighting Copyright Concerns

"We don't call it 'stuff,' Patrick. It's Blue Barnacle."

By [Victor Tzeng](#) / Published Oct 6, 2025 6:00 AM EDT



Vox_Ocult via X

RED HOT

What Is an "Author"?-Copyright Authorship of AI Art Through a Philosophical Lens

[Mackenzie Caldwell](#)

Copyright Law

AI Art

Mackenzie Caldwell

Big AI's Dirty Secret

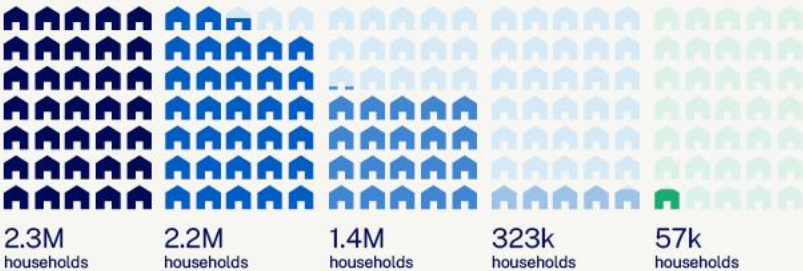
Power party!

Who uses the most electricity?

Electricity consumption in gigawatt hours



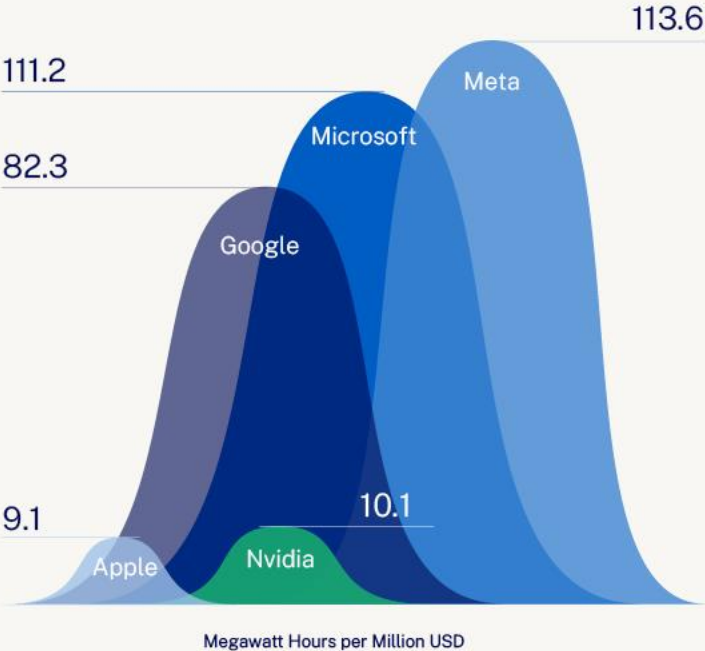
The average U.S. household consumes approximately 10,791 (kWh) *



Microsoft's annual electricity use could power 48 Disneyland Paris parks for an entire year! That's a lot of 'it's a small world' rides! **

Bang for buck!

How much energy is used per million dollars earned?



Apple makes \$1M in sales with the same electricity needed for 152 full EV charges!

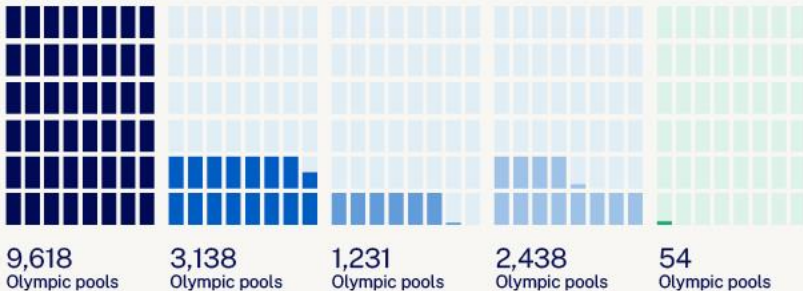
Splash zone!

How much water do they use?

Water consumption in cubic meters



An Olympic-sized swimming pool requires 2.5 million liters of water to fill, based on its standard dimensions.



Google's annual water consumption could fill over 120 million bathtubs — enough for almost everyone in Japan to take a bath!

Fun facts

All figures are based on the latest available annual data report

Google: 2023 Environmental report
Meta: 2024 Sustainability report
Apple: 2024 Environmental progress report



Google's annual energy could brew over 1 trillion Nespresso cups

<https://www.nytimes.com/wirecutter/reviews/best-nespresso-machine/>

Microsoft: 2024 Environmental sustainability report data fact sheet
Nvidia: 2024 Sustainability report



Microsoft's annual water consumption could fill 9 thousand Boeing 747-400 jets

https://www.boeing-747.com/fun_facts_from_boeing.php

* <https://www.eia.gov/tools/faqs/faq.php?id=97&t=3>
** <https://www.pv-magazine.com/2022/04/22/first-milestone-for-solar-carport-at-disneyland-in-paris/>



These five major tech companies consume ~1.7% of U.S. electricity

<https://www.eia.gov/totalenergy/data/monthly/>

IMD / TONOMUS Global Center for Digital and AI Transformation

'I can't drink the water' - life next to a US data centre

10 July 2025

Michelle Fleury & Nathalie Jimenez North America business correspondent & Business reporter, Georgia

Share ↗

ARTIFICIAL INTELLIGENCE | ETHICS

POWER PLAY

AI Data Centers Are Skyrocketing Regular People's Energy Bills

How is this fair?

By **Rae Witte** / Published Oct 4, 2025 12:00 PM EDT



1995 2025

AI IS DRAINING WATER FROM AREAS THAT NEED IT MOST

By [Leonardo Nicoletti](#), [Michelle Ma](#) and [Dina Bass](#)
for **Bloomberg Technology + Green**
May 8, 2025

Socio-Economic Consequences: Jobs

TECHNOLOGY

Job Interviews Are Broken

People are sneaking answers from AI, and who can blame them?

By Ian Bogost

The biggest fear was that AI would steal our jobs. In reality, it's replacing those of people who are learning how to work.

13 October 2025 | By Dr. Rosalia Neve

Socio-Economic Consequences: Trust

Consumer trust and data sharing in the age of Generative AI: survey insights

Our Global AI Perception Survey explores how the proliferation of artificial intelligence (AI) and Generative AI (GenAI) is impacting consumer trust and willingness to share personal health data with Life & Health insurers.



EMERGING TECHNOLOGIES

Trust is the new currency in the AI agent economy

Jul 25, 2025

Conclusion

- **Design responsibly:** Build transparency, fairness, and privacy into AI from the start.
- **Educate and empower:** Promote AI literacy for all so users understand both the power and limits of these tools.
- **Regulate:** Support policies that encourage innovation while protecting individuals and communities.
- **Implement wisely:** Think of whether AI is really needed for your tasks.