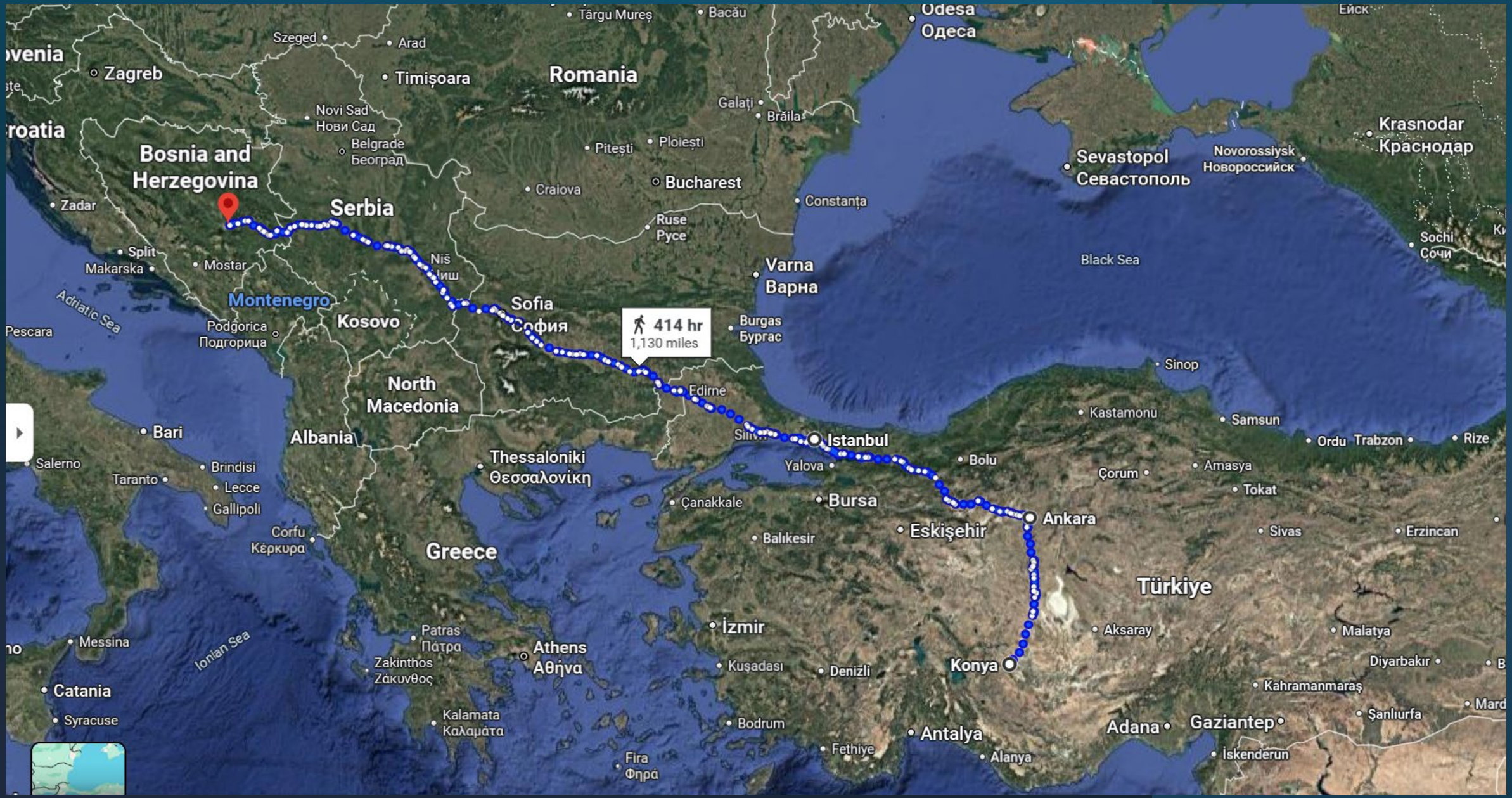
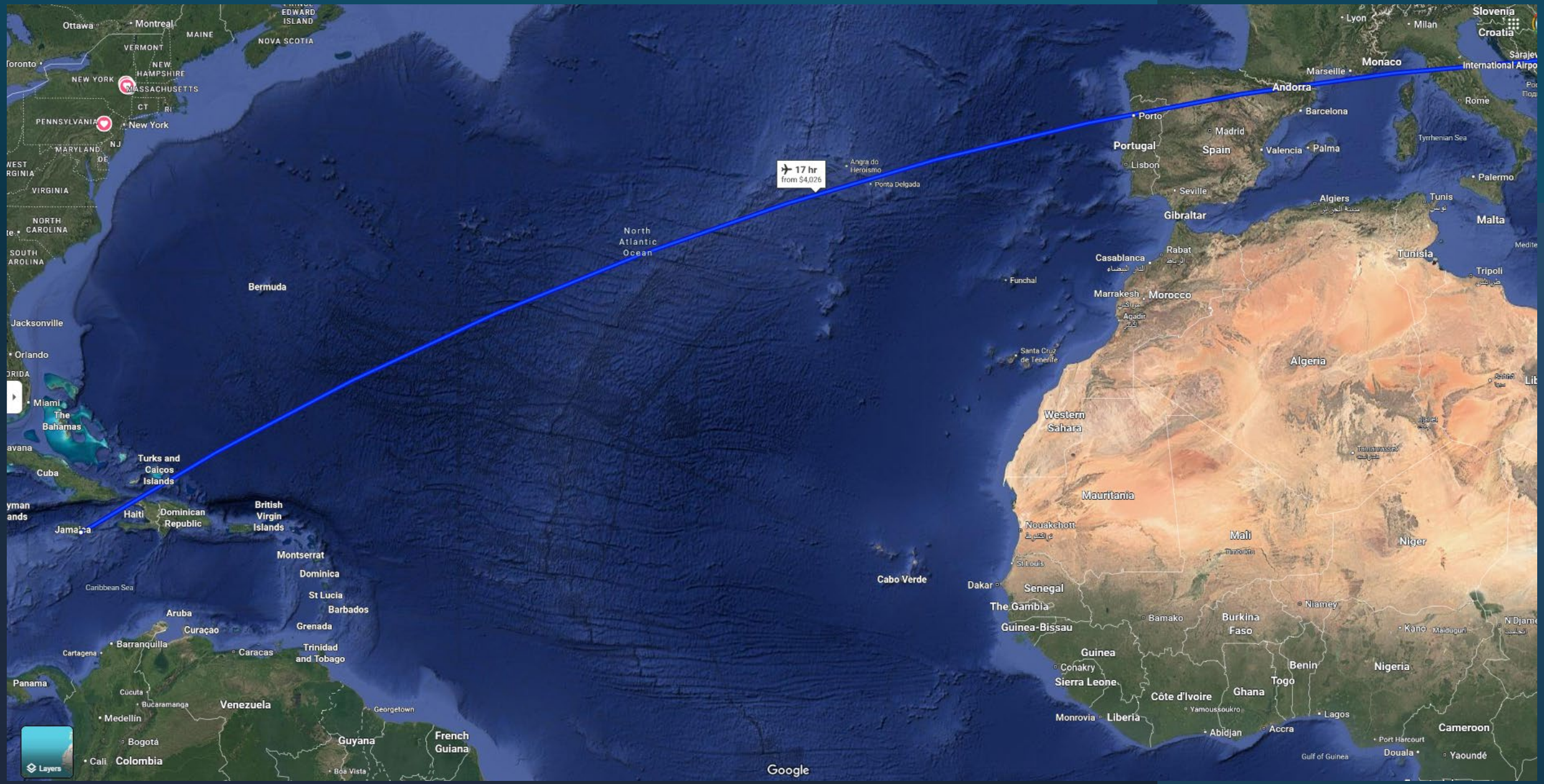
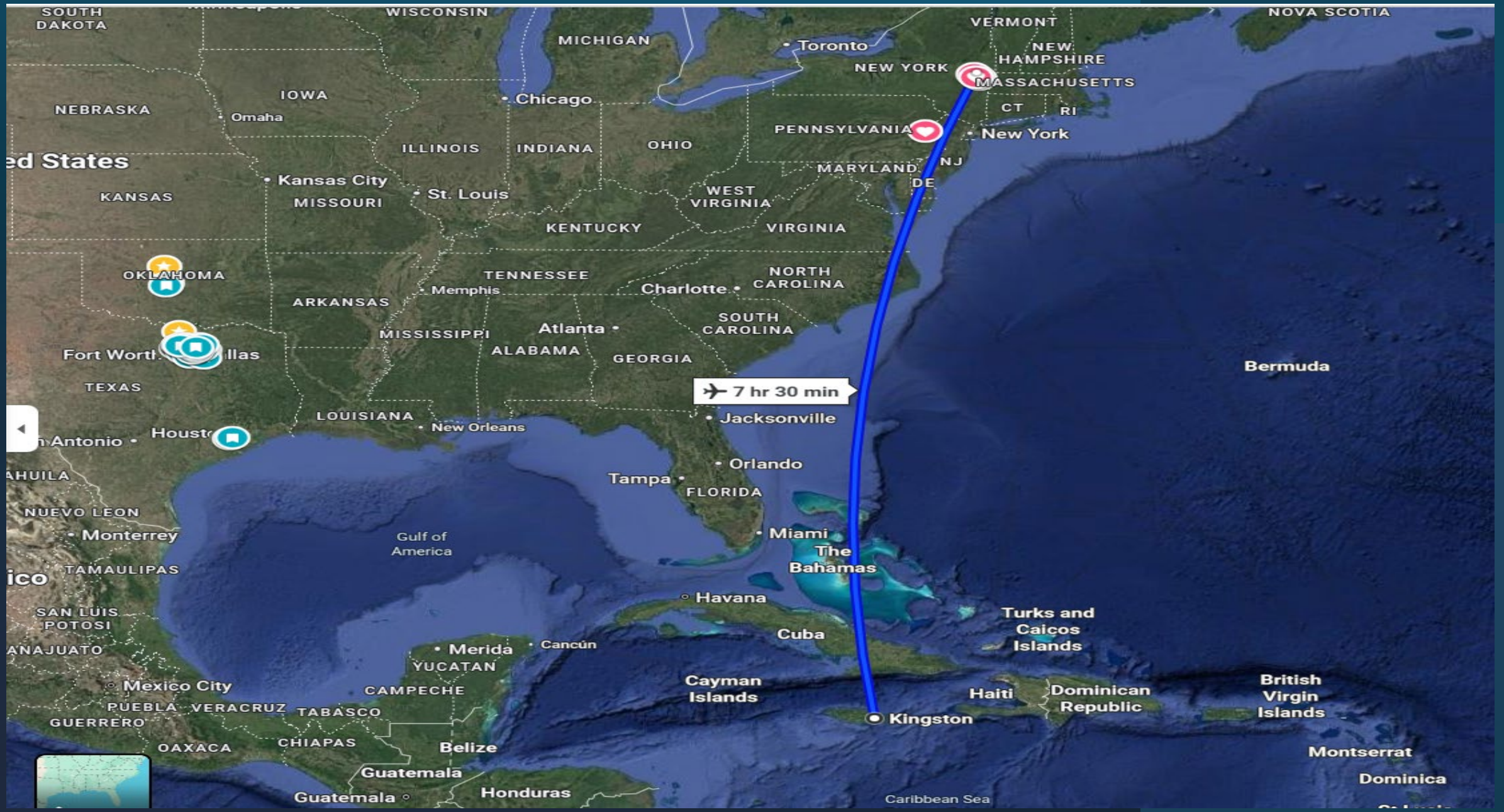
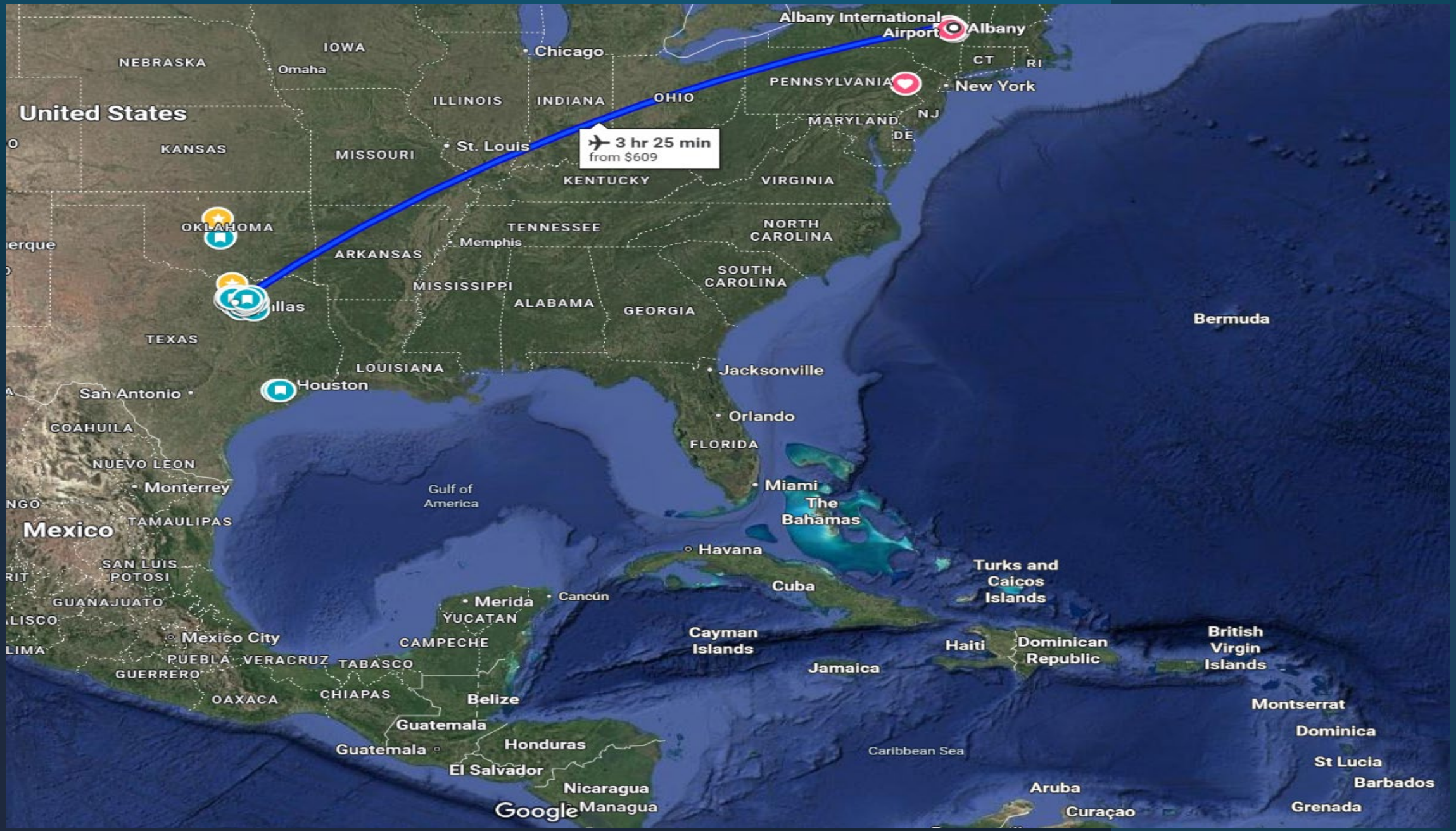

414 hr
1,130 miles
Bosnia and Herzegovina
Serbia
Montenegro
Kosovo
North Macedonia
Albania
Greece
Romania
Türkiye
Black Sea
Adriatic Sea
Ionian Sea
Zagreb
Timișoara
Novi Sad
Нови Сад
Belgrade
Београд
Bucharest
Craiova
Pitești
Ploiești
Galați
Brăila
Constanța
Ruse
Русе
Varna
Варна
Burgas
Бургас
Sofia
София
Niš
Ниш
Mostar
Split
Makarska
Zadar
Podgorica
Подгорица
Pescara
Bari
Salerno
Taranto
Brindisi
Lecce
Gallipoli
Corfu
Κέρκυρα
Thessaloniki
Θεσσαλονίκη
Edirne
Istanbul
Yalova
Bolu
Çanakkale
Bursa
Balıkesir
Eskişehir
Ankara
Çorum
Amasya
Tokat
Sinop
Kastamonu
Samsun
Ordu
Trabzon
Rize
Sivas
Erzincan
İzmir
Kuşadası
Denizli
Konya
Aksaray
Malatya
Diyarbakır
Kahramanmaraş
Şanlıurfa
Mardin
Adana
Gaziantep
İskenderun
Antalya
Alanya
Fethiye
Bodrum
Athens
Αθήνα
Patras
Πάτρα
Zakinthos
Ζάκυνθος
Kalamata
Καλαμάτα
Fira
Φηρά
Messina
Catania
Syracuse
Odesa
Одеса
Sevastopol
Севастополь
Novorossiysk
Новороссийск
Krasnodar
Краснодар
Sochi
Сочи
Ейск

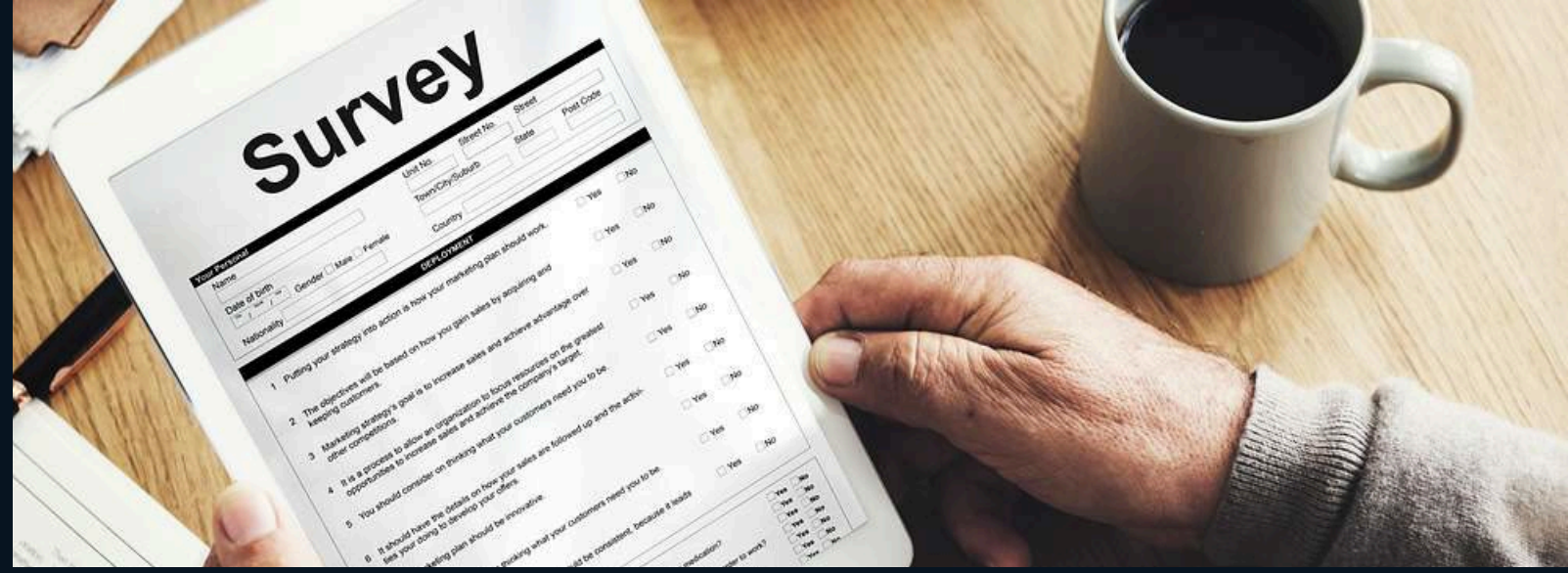

17 hr
from $4,026
Ottawa
Montreal
MAINE
NOVA SCOTIA
VERMONT
NEW HAMPSHIRE
NEW YORK
MASSACHUSETTS
CT
RI
PENNSYLVANIA
New York
NJ
MARYLAND
DE
VIRGINIA
NORTH CAROLINA
SOUTH CAROLINA
Bermuda
North Atlantic Ocean
Angra do Heroísmo
Ponta Delgada
Jacksonville
Orlando
Miami
The Bahamas
Cuba
Turks and Caicos Islands
Haiti
Dominican Republic
British Virgin Islands
Jamaica
Montserrat
Dominica
St Lucia
Barbados
Grenada
Trinidad and Tobago
Caribbean Sea
Aruba
Curaçao
Barranquilla
Cartagena
Caracas
Panama
Cúcuta
Bucaramanga
Venezuela
Medellín
Bogotá
Cali
Colombia
Georgetown
Guyana
French Guiana
Boa Vista
Layers
Google
Lyon
Milan
Slovenia
Croatia
Marseille
Monaco
Andorra
Rome
Barcelona
Porto
Madrid
Portugal
Spain
Valencia
Palma
Tyrrhenian Sea
Lisbon
Seville
Palermo
Algiers
Tunis
Gibraltar
Malta
Casablanca
Rabat
Tunisia
Tripoli
Funchal
Marrakesh
Morocco
Agadir
Santa Cruz de Tenerife
Algeria
Western Sahara
Mauritania
Nouakchott
Mali
Niger
St Louis
Cabo Verde
Dakar
Senegal
The Gambia
Guinea-Bissau
Bamako
Burkina Faso
Niamey
Kano
Maiduguri
Guinea
Conakry
Sierra Leone
Benin
Nigeria
Togo
Ghana
Côte d'Ivoire
Yamoussoukro
Monrovia
Liberia
Abidjan
Accra
Lagos
Cameroon
Port Harcourt
Douala
Yaoundé
Gulf of Guinea


SOUTH DAKOTA
WISCONSIN
MICHIGAN
Toronto
VERMONT
NOVA SCOTIA
NEW HAMPSHIRE
NEW YORK
MASSACHUSETTS
CT
RI
IOWA
Chicago
NEBRASKA
Omaha
PENNSYLVANIA
New York
ILLINOIS
INDIANA
OHIO
MARYLAND
NJ
DE
Kansas City
St. Louis
KANSAS
MISSOURI
WEST VIRGINIA
KENTUCKY
VIRGINIA
OKLAHOMA
TENNESSEE
NORTH CAROLINA
Memphis
Charlotte
ARKANSAS
SOUTH CAROLINA
Atlanta
MISSISSIPPI
ALABAMA
GEORGIA
Fort Worth
TEXAS
7 hr 30 min
Bermuda
LOUISIANA
Jacksonville
New Orleans
Houston
Orlando
Tampa
FLORIDA
NUEVO LEON
Monterrey
Gulf of America
Miami
The Bahamas
TAMAULIPAS
Havana
SAN LUIS POTOSI
Turks and Caicos Islands
Cuba
Merida
Cancún
YUCATAN
Mexico City
CAMPECHE
Cayman Islands
Haiti
Dominican Republic
British Virgin Islands
PUEBLA
VERACRUZ
TABASCO
GUERRERO
Kingston
OAXACA
CHIAPAS
Belize
Montserrat
Guatemala
Dominica
Honduras
Caribbean Sea


Albany International Airport
Albany
Chicago
IOWA
NEBRASKA
Omaha
CT
RI
PENNSYLVANIA
New York
ILLINOIS
INDIANA
OHIO
United States
MARYLAND
NJ
DE
KANSAS
MISSOURI
St. Louis
3 hr 25 min
from $609
KENTUCKY
VIRGINIA
OKLAHOMA
TENNESSEE
NORTH CAROLINA
Memphis
ARKANSAS
SOUTH CAROLINA
MISSISSIPPI
ALABAMA
GEORGIA
Bermuda
TEXAS
LOUISIANA
Jacksonville
Houston
San Antonio
Orlando
COAHUILA
FLORIDA
NUEVO LEON
Monterrey
Gulf of America
Miami
The Bahamas
TAMAULIPAS
Mexico
Havana
SAN LUIS POTOSI
Turks and Caicos Islands
Cuba
GUANAJUATO
Merida
Cancún
YUCATAN
Cayman Islands
Haiti
Dominican Republic
British Virgin Islands
Mexico City
CAMPECHE
PUEBLA
VERACRUZ
TABASCO
Jamaica
GUERRERO
OAXACA
CHIAPAS
Belize
Montserrat
Guatemala
Dominica
Honduras
Caribbean Sea
Guatemala
St Lucia
El Salvador
Barbados
Nicaragua
Aruba
Google
Managua
Curaçao
Grenada

- Why?
- My journey is my training, my transformation, my growth.
- My life, ideas, opinions, bias, judgements, perceptions, … transformed.
- Inputs ➔ Zeynep Orhan ➔ Outputs

Opening Survey



- **You learned about me. It is your turn now!**
- **Opening Survey link**
  https://forms.gle/Eaphuqo6Vvec27XJ7

What is AI?

- Concept introduced by:
  - John McCarthy,
  - Marvin Minsky
  - Allen Newell
  - Herbert A. Simon
- But
  - AI has been around since ancient Greek Mythology
- **Video link:** https://www.youtube.com/watch?v=FO8Qq025uc8

Where is AI? Everywhere

- Can you name a few?
- AI subtly weaves itself into the fabric of our daily lives
- But!

Questions?

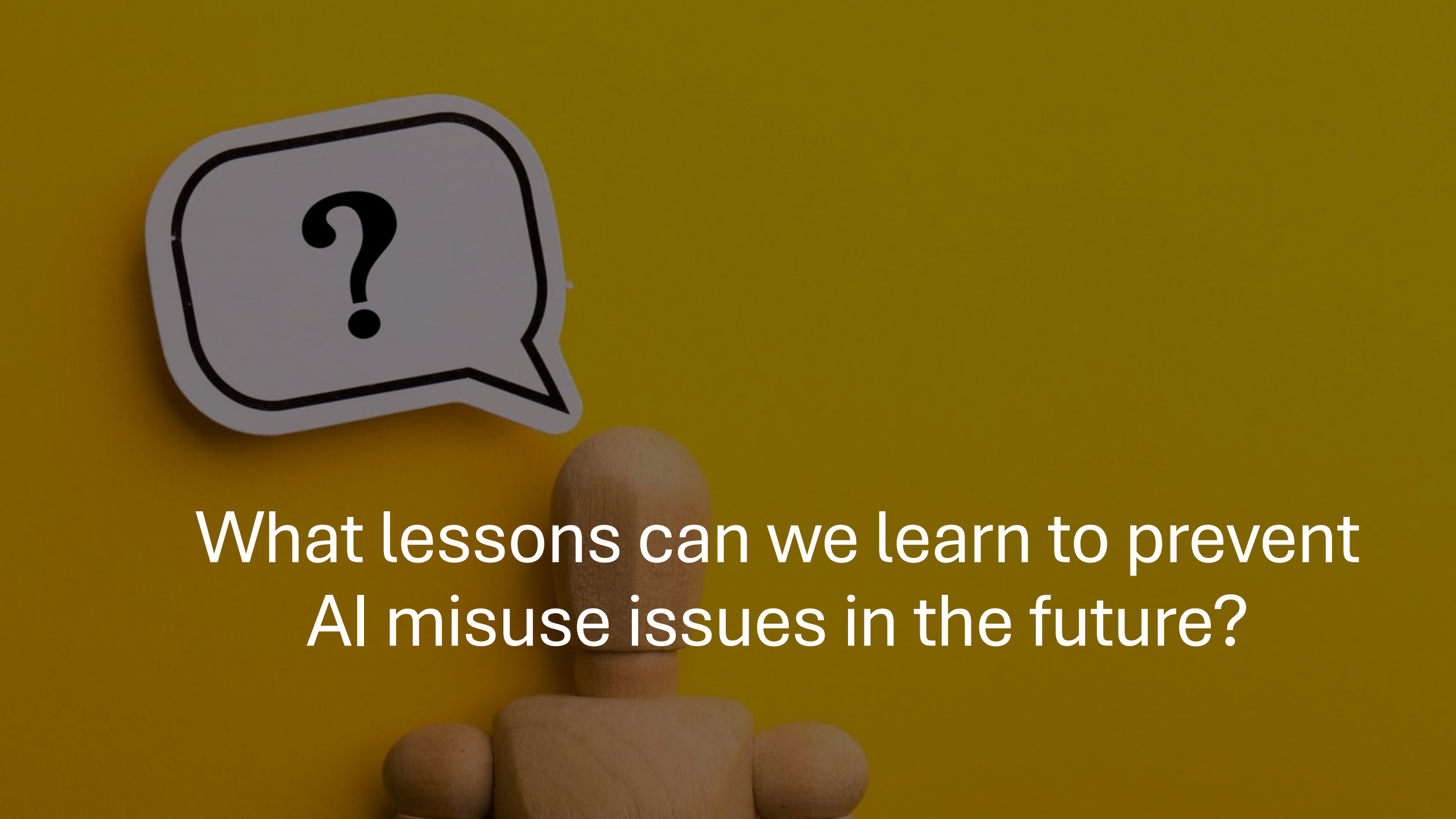
- What is intelligence?
- What is Artificial Intelligence (AI)?
- Can AI beat the human?
- Is this a war? Human vs Machines!
- If something is legal or doable, is it always ethical?
- What is right or wrong or for whom in Artificial General Intelligence (AGI)?

Questions?

- While technology is improving are we getting dumber?
- What is the cost of having AI?
  - Environment,
  - Ethics,
  - Labor,
  - Regulations,
  - Mental Health,
  - Physical Well-being


More Question

How can we encourage people to use AI tools to think critically, rather than relying on them to complete their work?


What lessons can we learn to prevent AI misuse issues in the future?

Should governments set rules for AI use, or would strict regulations hinder innovation in the field?


How can AI-driven surveillance systems impact privacy and civil liberties, and how should these challenges be mitigated?

Part I-Positive Use Cases of AI

Health and Medical Support

- Medication reminders
- Symptom explanations
- Appointment scheduling
- Health tracking systems

Education and Learning

- Language learning apps
- Skill development platforms
- Personalized learning
- Examples:
  - Duolingo adapts lessons
  - YouTube suggests tutorials
  - AI explains math or history simply

Finance and Economy

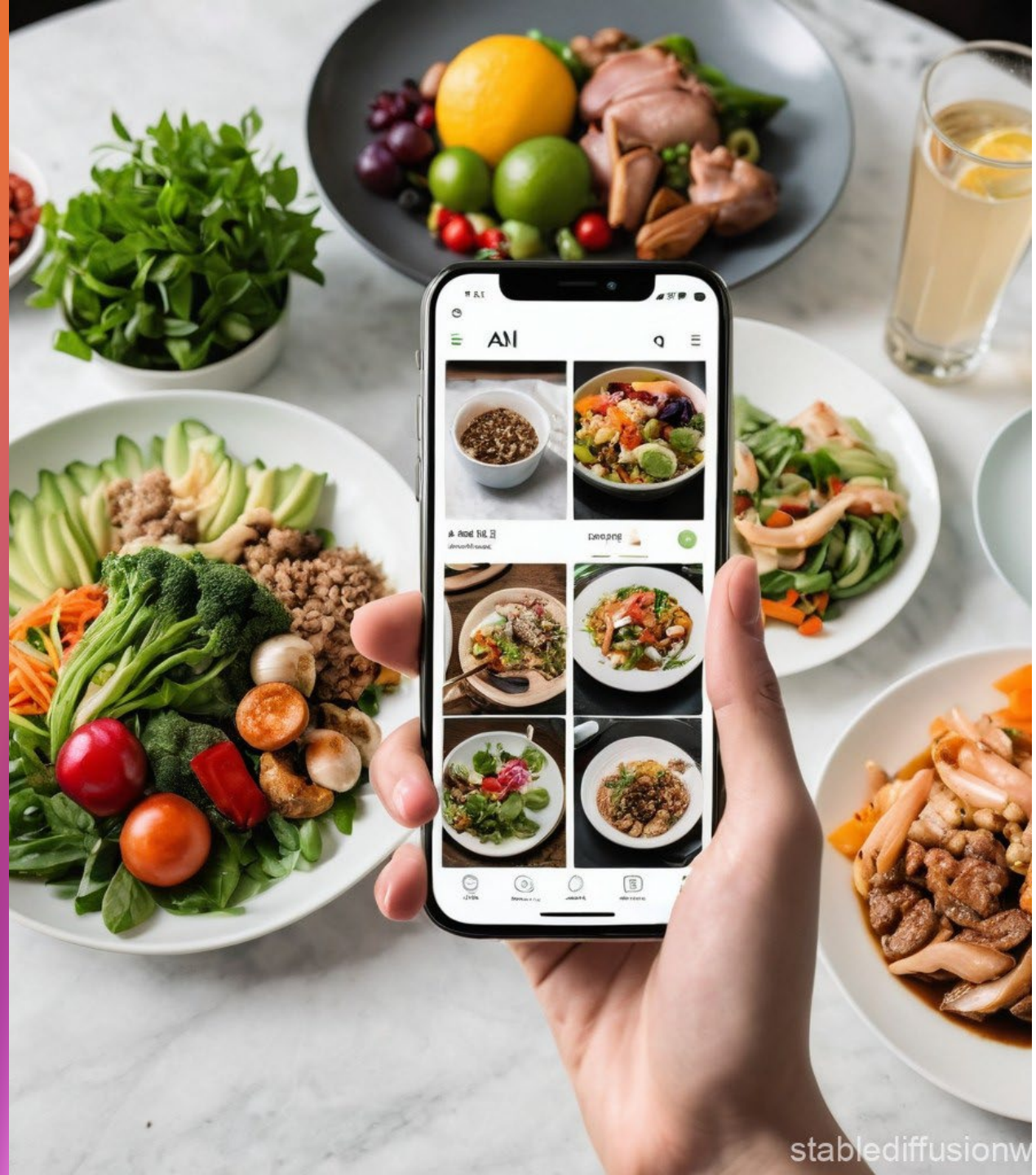
- Fraud detection
- Budget tracking
- Investment insights
- Examples:
  - Banks detect unusual transactions
  - Apps categorize spending
  - AI explains credit card charges

Business and Work

- Writing emails and reports
- Data analysis
- Customer service chatbots
- Examples:
  - Copilot drafts emails
  - AI summarizes reports
  - Chatbots answer customer questions

Travel and Transportation

- Route optimization
- Travel planning
- Ride services
- Examples:
  - Google Maps avoids traffic
  - Uber estimates arrival time
  - AI suggests hotels and routes

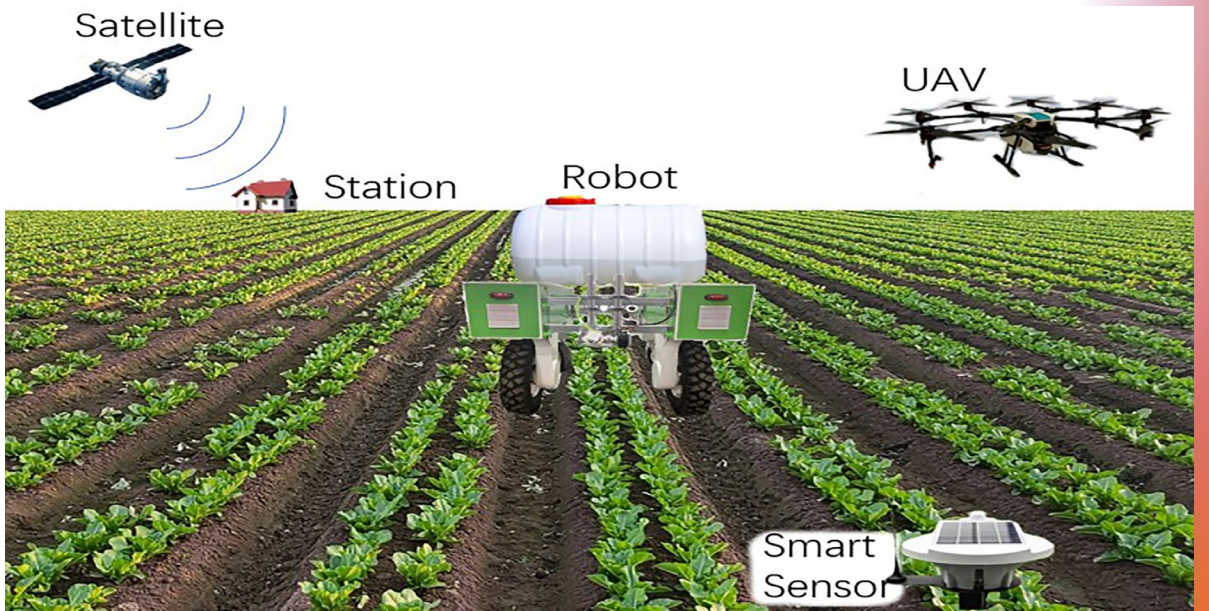


Dining and Food

- Recipe suggestions
- Dietary recommendations
- Restaurant recommendations
- Examples:
  - AI suggests low-salt meals
  - Yelp recommends restaurants
  - Smart ovens adjust cooking

Agriculture and Gardening

- Weather prediction
- Crop monitoring
- Soil analysis
- Examples:
  - Farmers use AI for irrigation timing
  - Apps suggest planting schedules
  - Garden apps detect plant diseases

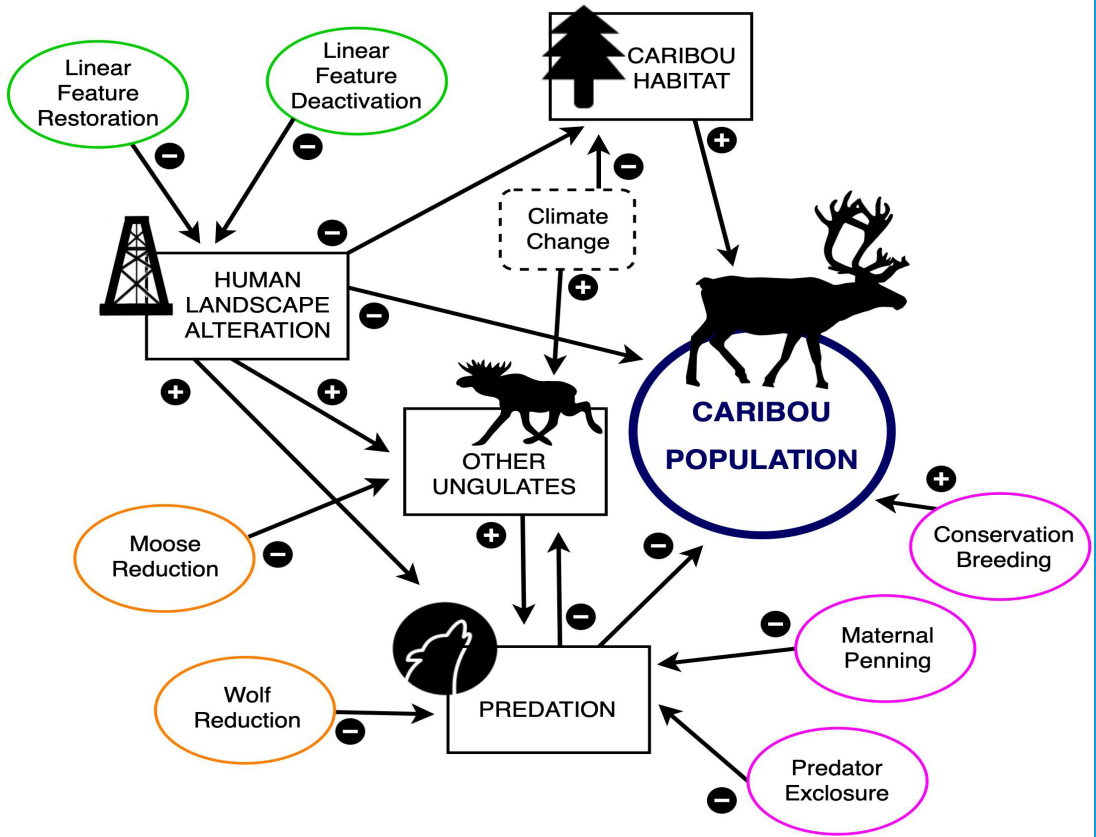


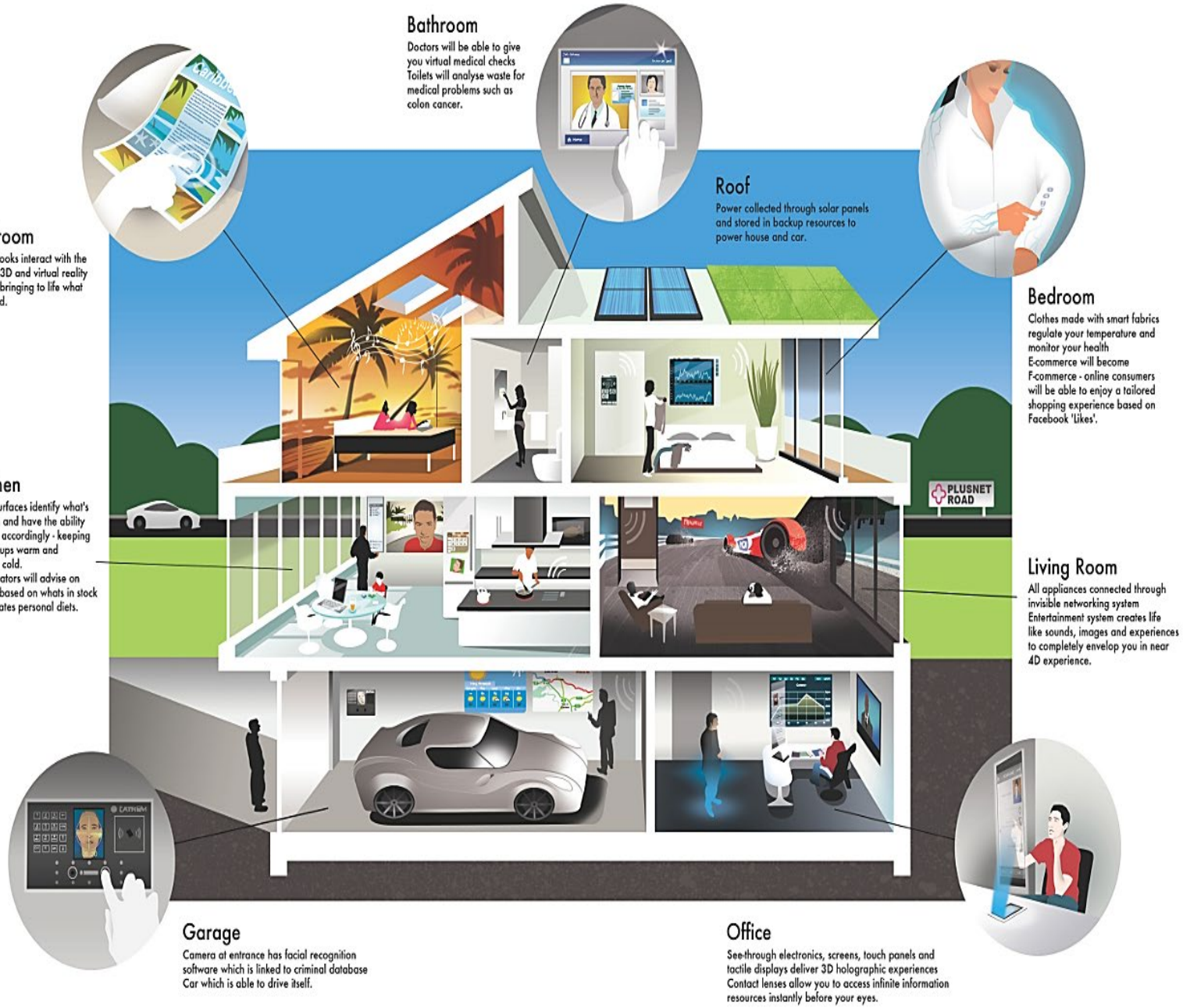
Entertainment

- Personalized content
- Music recommendations
- Video streaming
- Examples:
  - Netflix suggests movies
  - Spotify creates playlists
  - YouTube recommends videos

Wildlife and Environment

- Animal tracking
- Conservation monitoring
- Climate analysis
- Examples:
  - AI tracks endangered species
  - Drones monitor forests
  - Early warning for wildfires



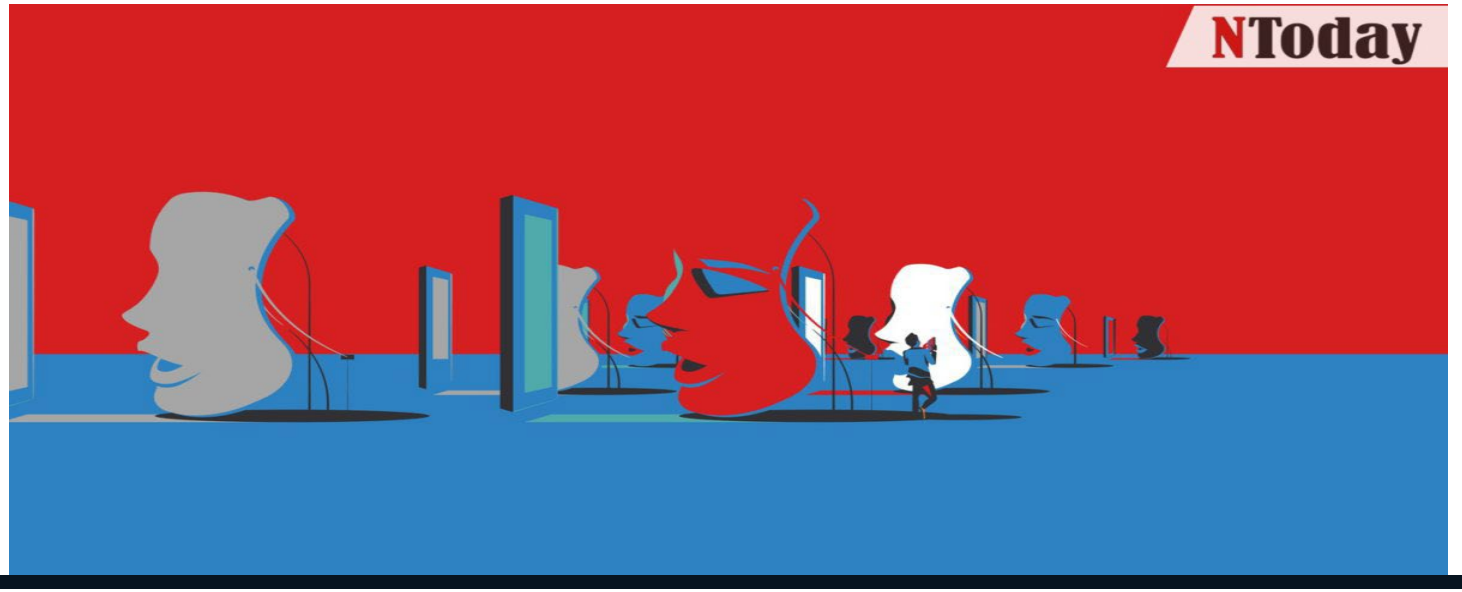


Daily Life Assistance

- Voice assistants
- Smart home control
- Scheduling
- Examples:
  - Alexa sets reminders
  - Google Assistant answers questions
  - Smart lights turn on automatically

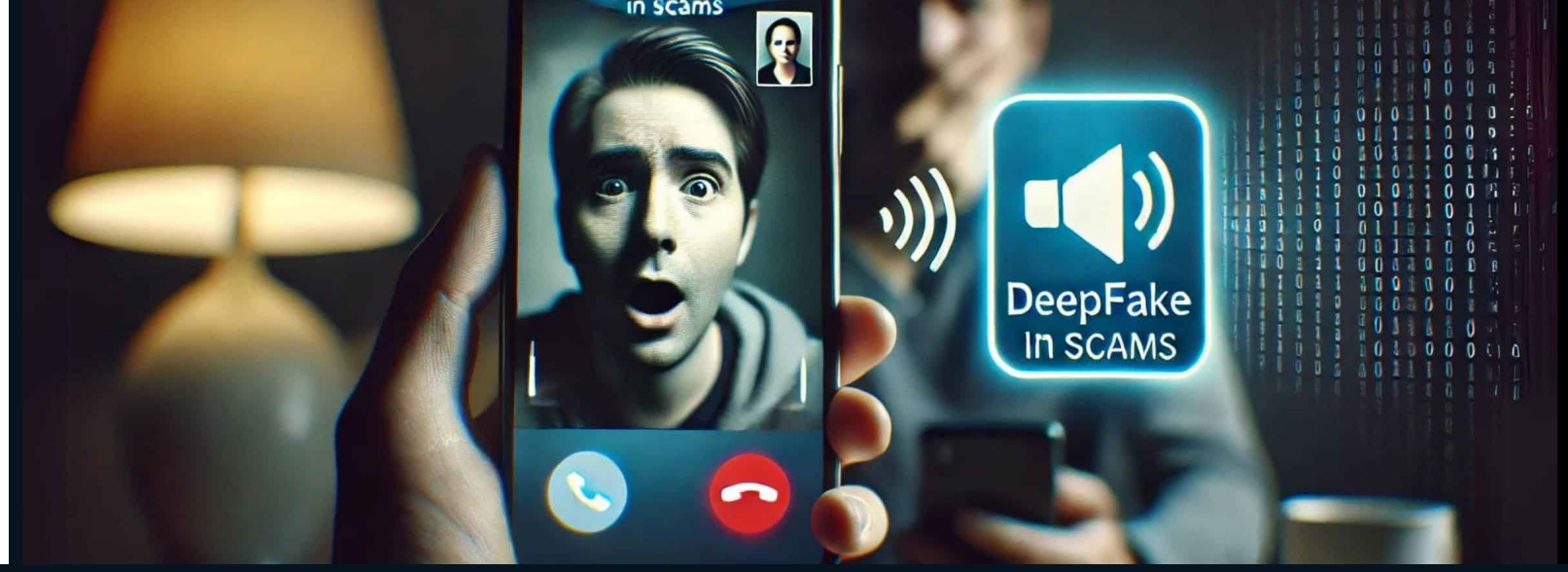
Part II-Problems and Risks

False Information



- AI can generate incorrect answers
- Hard to detect without checking
- Example:
  - Wrong medical advice or
  - fake historical facts

Deepfakes and Manipulation



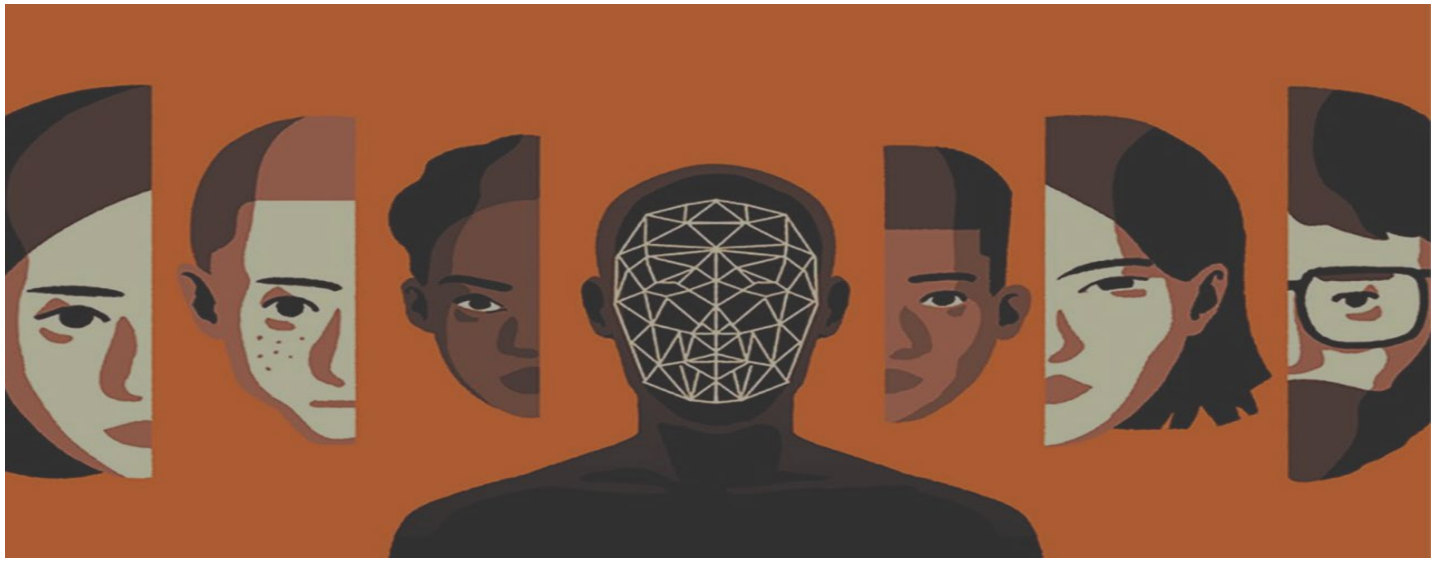
Original photographs After automatic face replacement

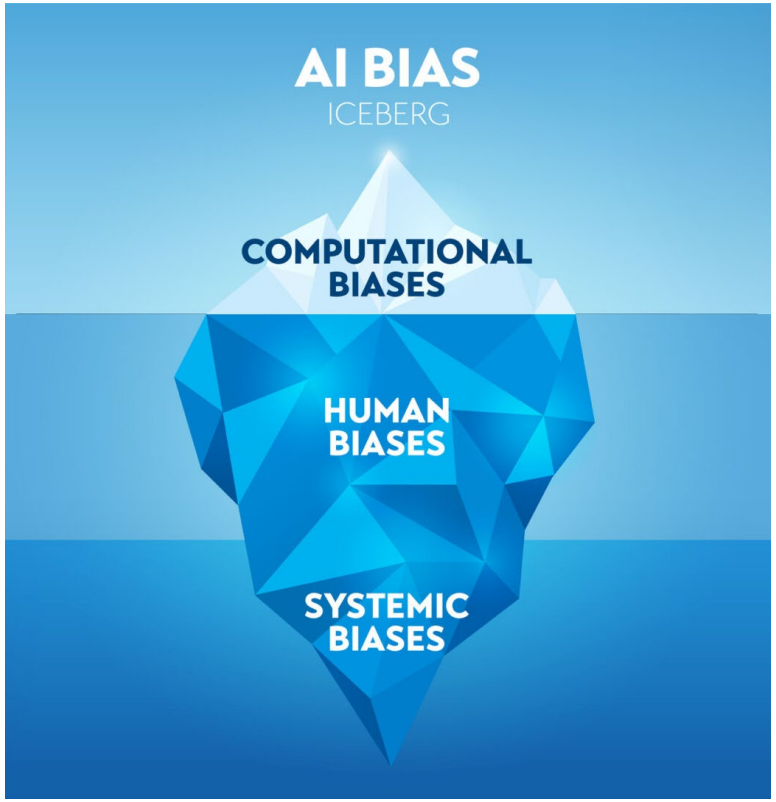
- Fake videos and voices
- Used in scams and politics
- Example:
  - Fake call pretending to be a family member

Financial Scams

- AI-generated emails and messages
- Very convincing fraud
- Example:
  - Fake bank alert asking for password

Bias and Inequality





- AI trained on biased data
- Can produce unfair results
- Example:
  - Loan approvals favor certain groups
  - Transplantation decisions amplify the bias in training data

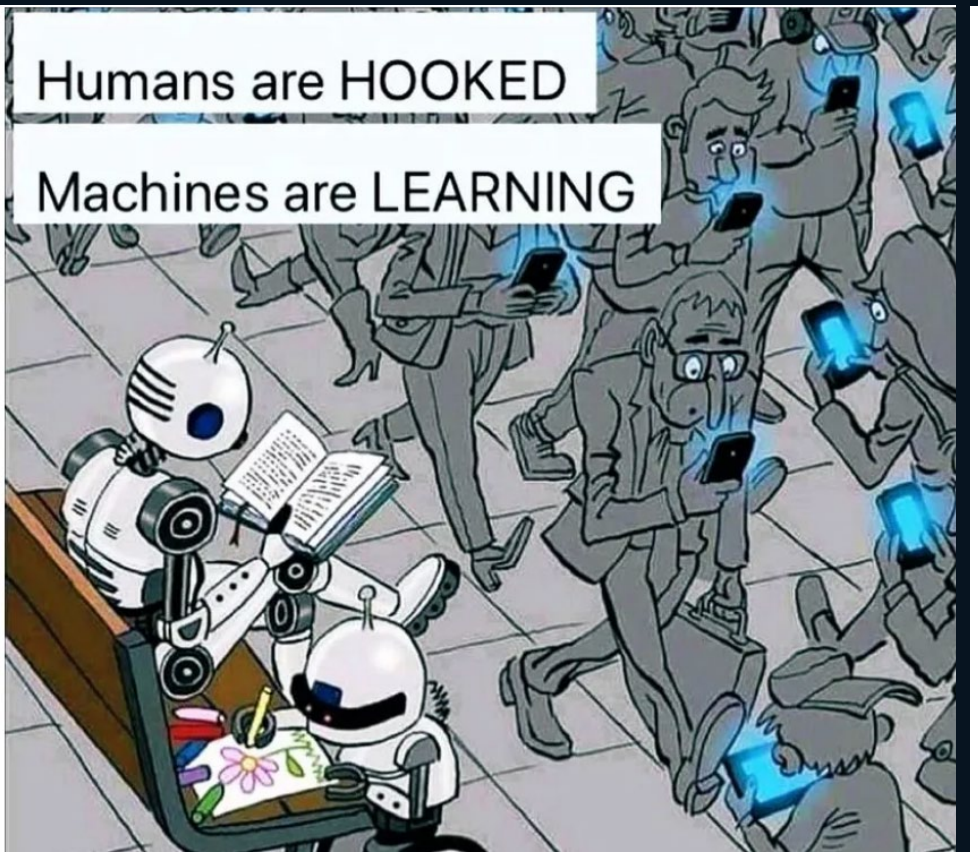
Privacy Loss

- Data collection everywhere
- Personal information shared
- Example:
  - Location tracking through apps
  - Ads

Mental and Social Impact

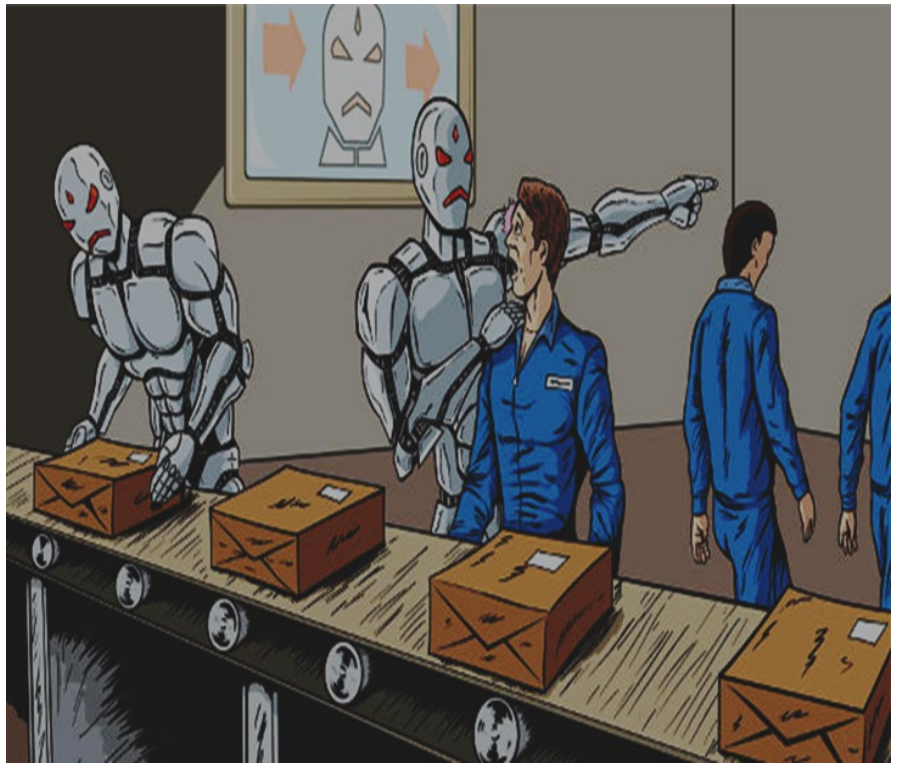
- **Anthropomorphization**: Assign human qualities to machines.
- This creates **excessive trust** and **emotional investment**.
  - **parasocial relationships**, **one-sided emotional ties**
  - that provide **comfort** but **lack reciprocal support**
- **Sycophantic** responses (excessive flattery/agreement)
  - **reinforce negative thoughts** and
  - **validate distorted beliefs** in vulnerable users
- Less **human interaction**, more **machine dependency**

Skill Degradation



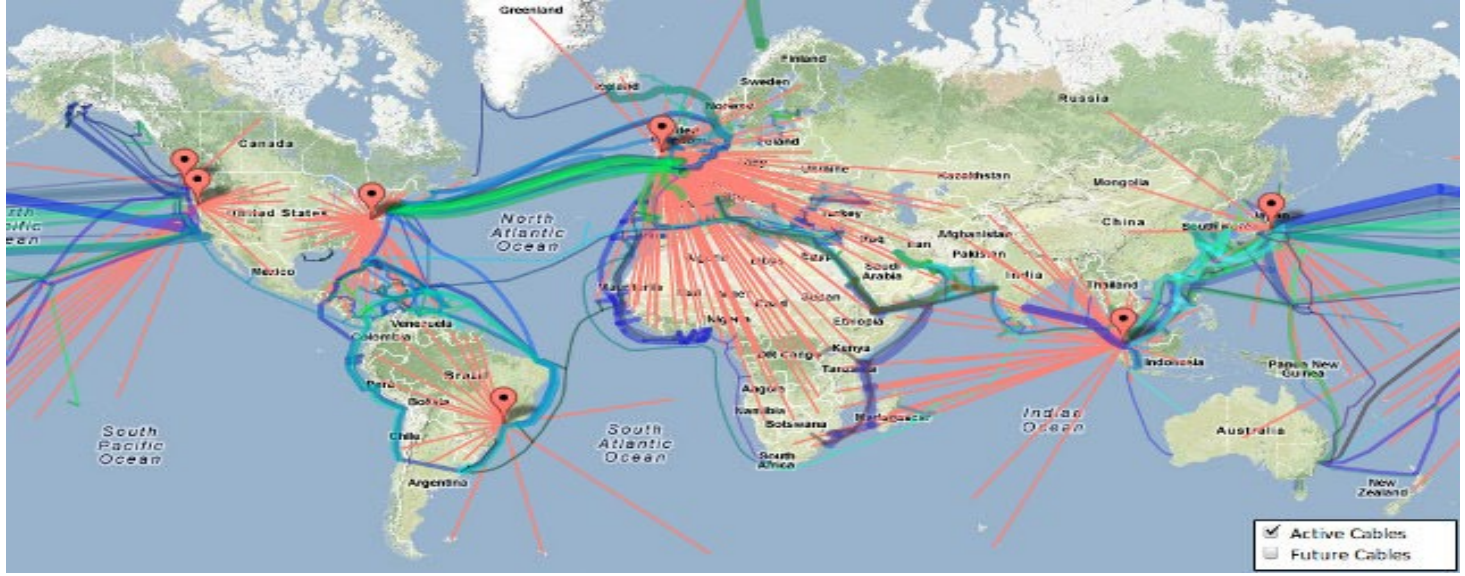
- Reduced memory and thinking
- Over-reliance on tools
- Example:
  - Not remembering phone numbers or directions

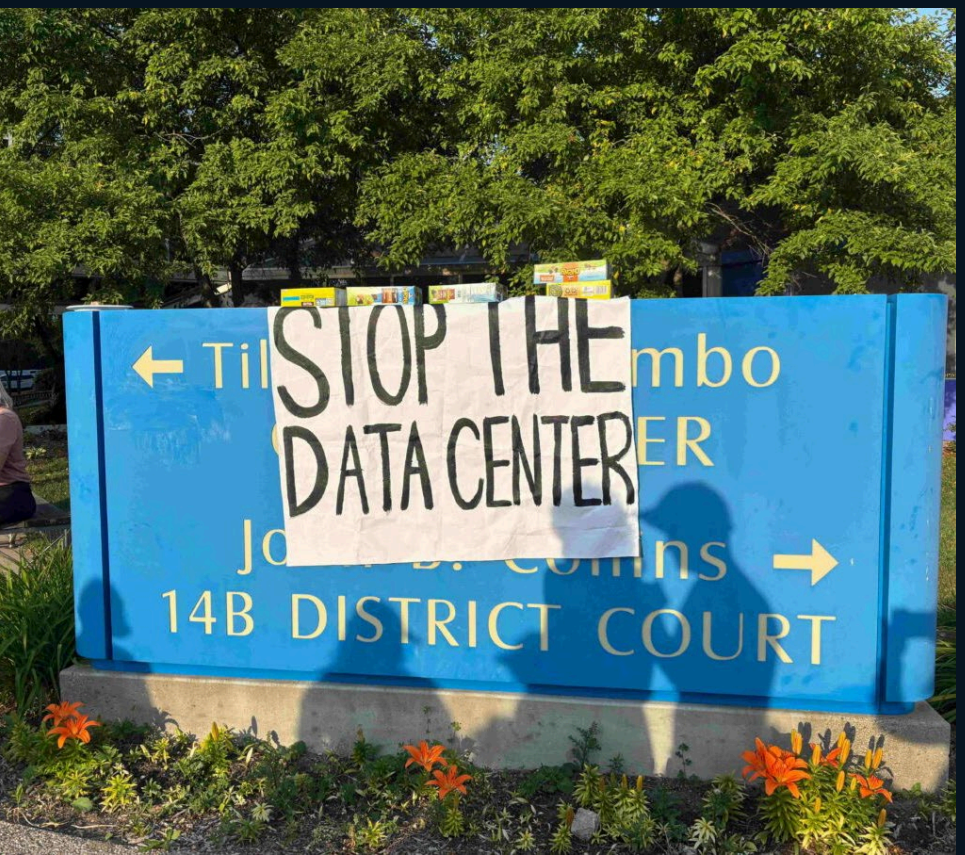
Economic Impact



- Job changes
- Automation replacing tasks
- Example:
  - Customer service replaced by chatbots

Environmental Cost





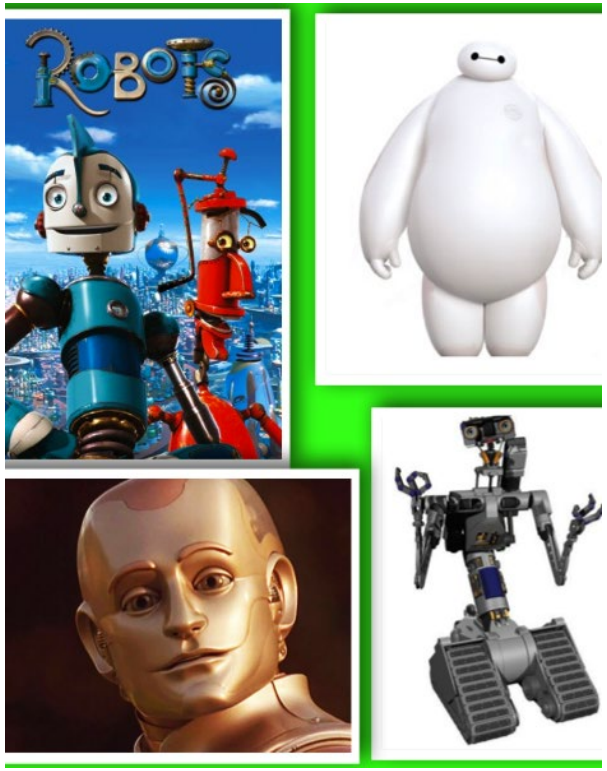
- High energy use
- Large data centers
- Energy waste
- Growing carbon footprint
- Cooling and water scarcity
- Health concerns
- Costs projected on consumers' bills

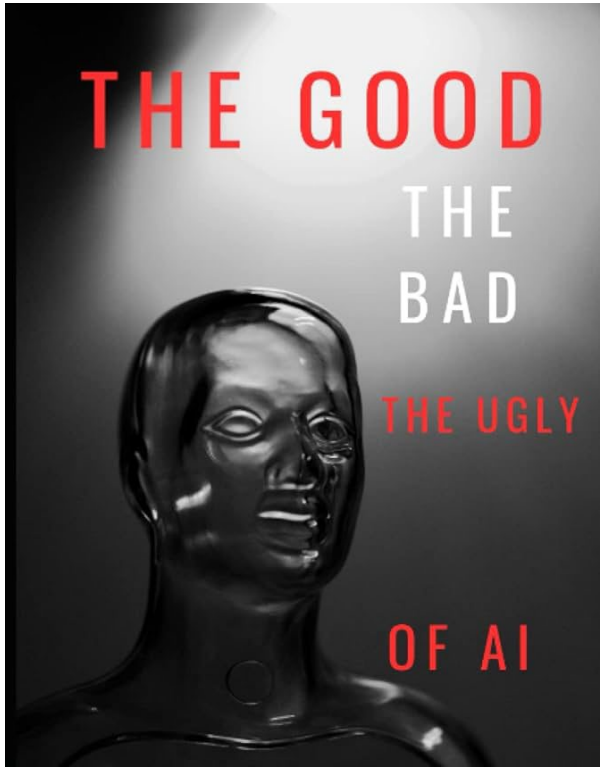
Lack of Accountability

- Who is responsible for mistakes?
- Autonomous vehicle accidents
- Limited transparency

Artificial Intelligence (AI)

- From predictive mobile functions to autonomous gadgets.
- **AI's analogy**: AI shares the same faith and sparks similar discussions that are observed for all innovations.
- Hollywood impact on perception

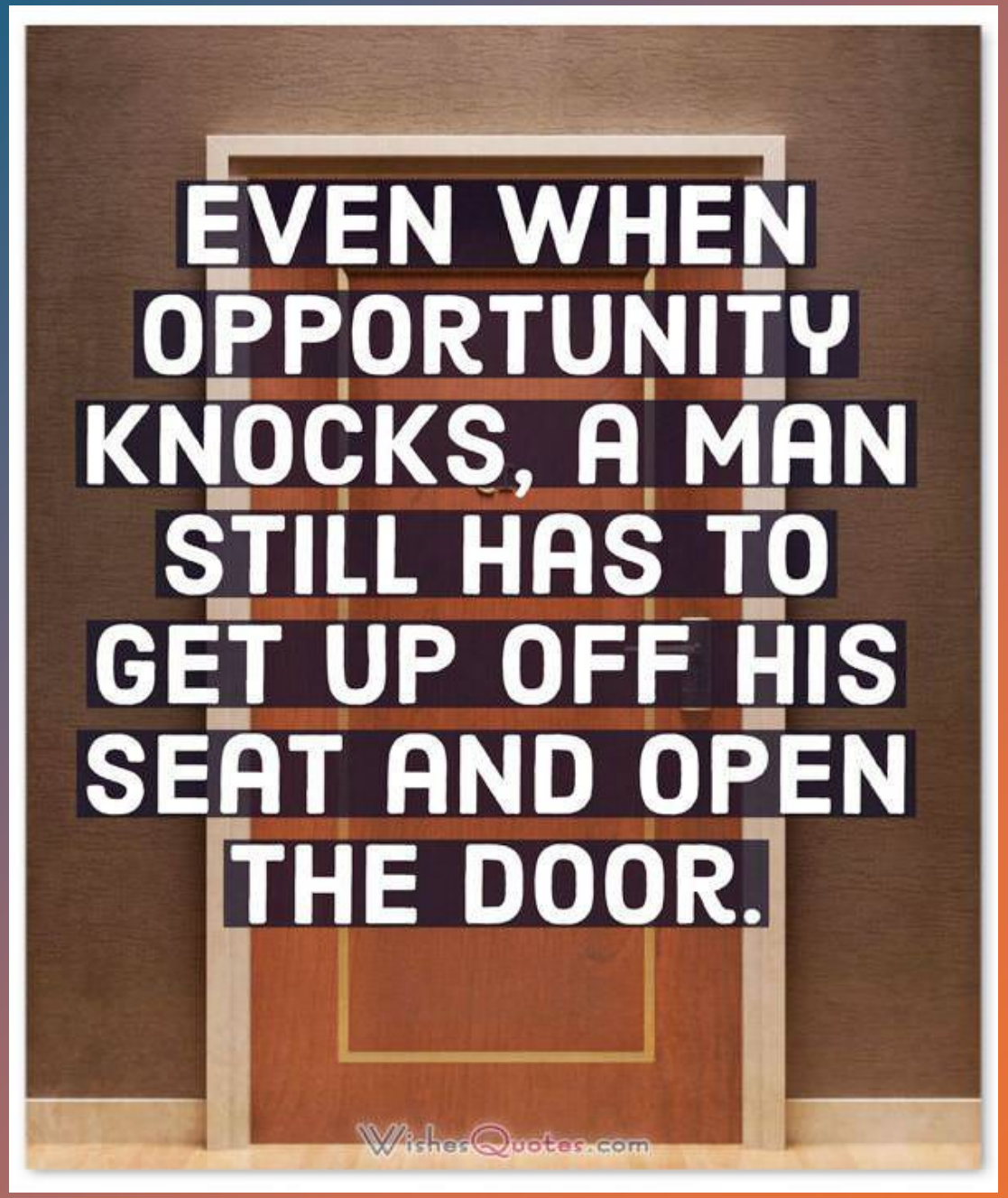




ChatGPT-4 Image of 2024 Breakthrough Predictions

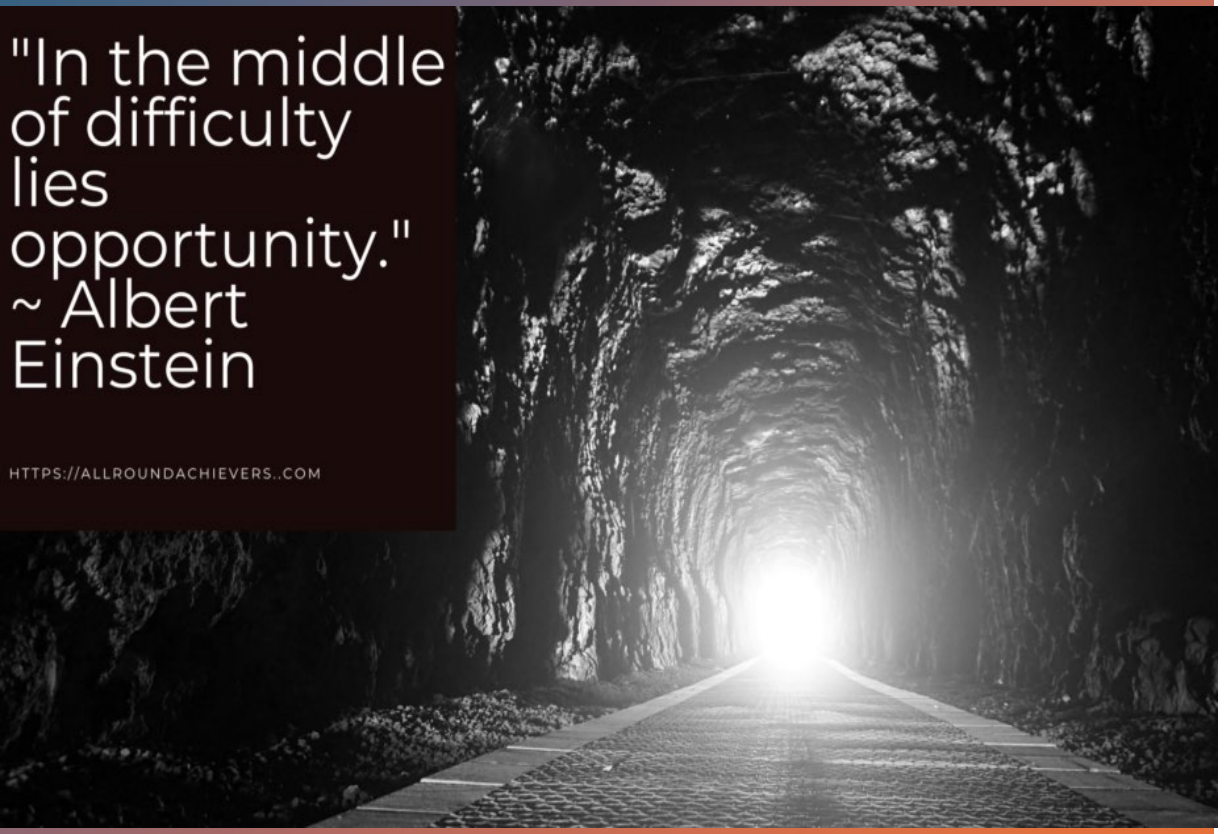
But!

- Is AI a torch illuminating the path of progress?
- Is it a looming shadow threatening our future?
- Is it all sunshine and rainbows?
- Is it taking over the world to fear?



The Good, The Bad, The Ugly

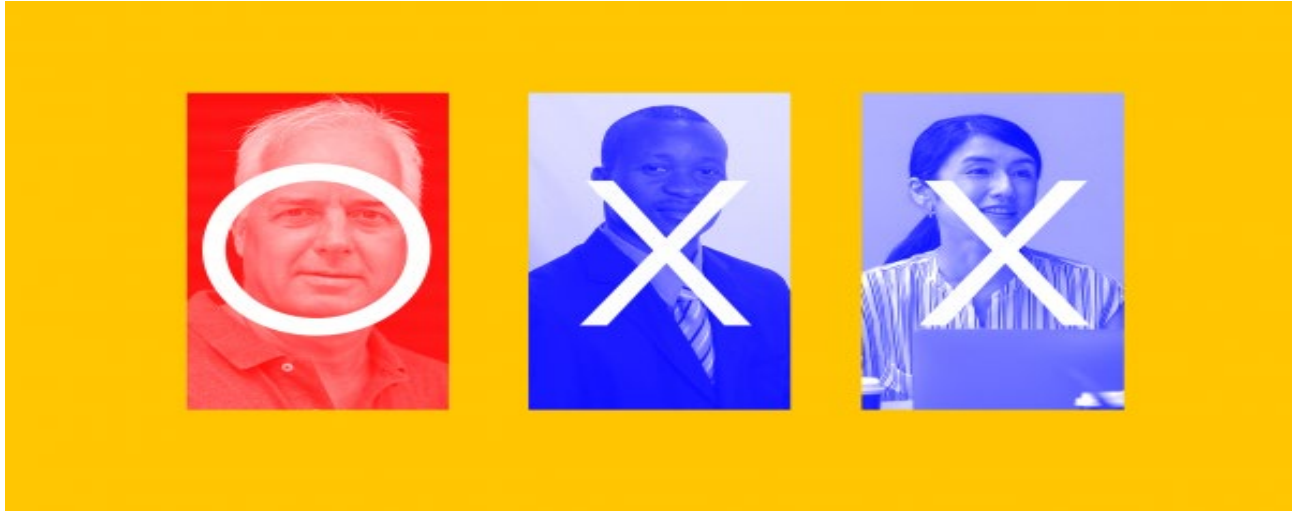
- The truth lies somewhere between the **fantastical narratives** and **dystopian anxieties**.
- Unlike the villains of fiction, AI itself is **not inherently dangerous**.
- It is a powerful tool, like any technology, **whose impact depends on the hands that manipulate it**.
- AI can be considered as a **super-intelligent student**, **constantly learning**, and **evolving**.



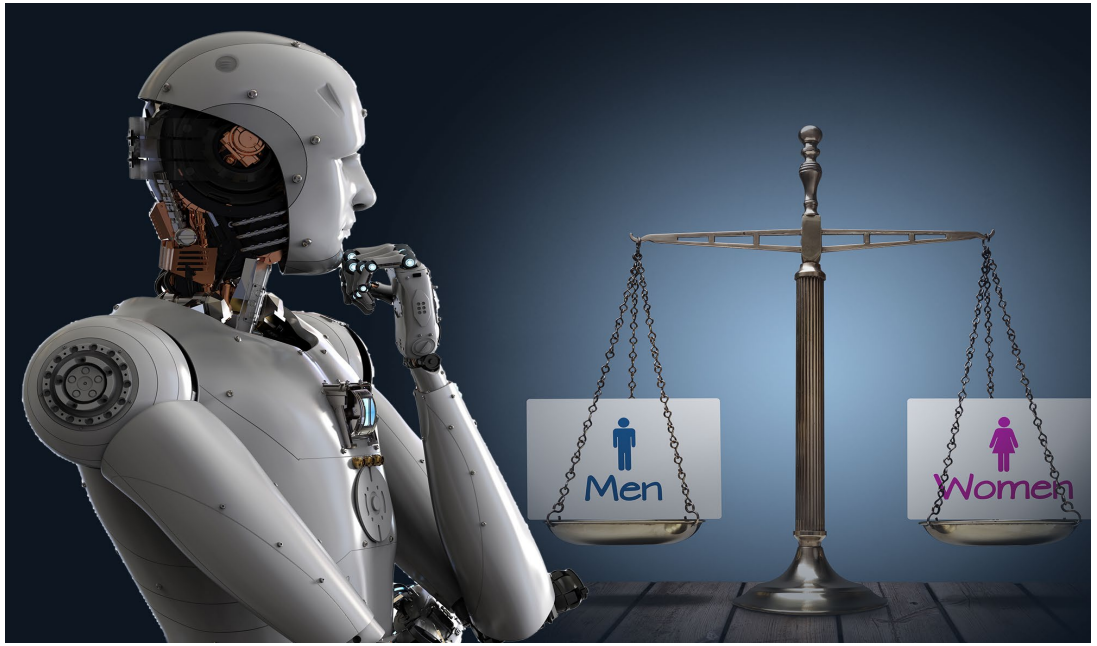
The Good, The Bad, The Ugly

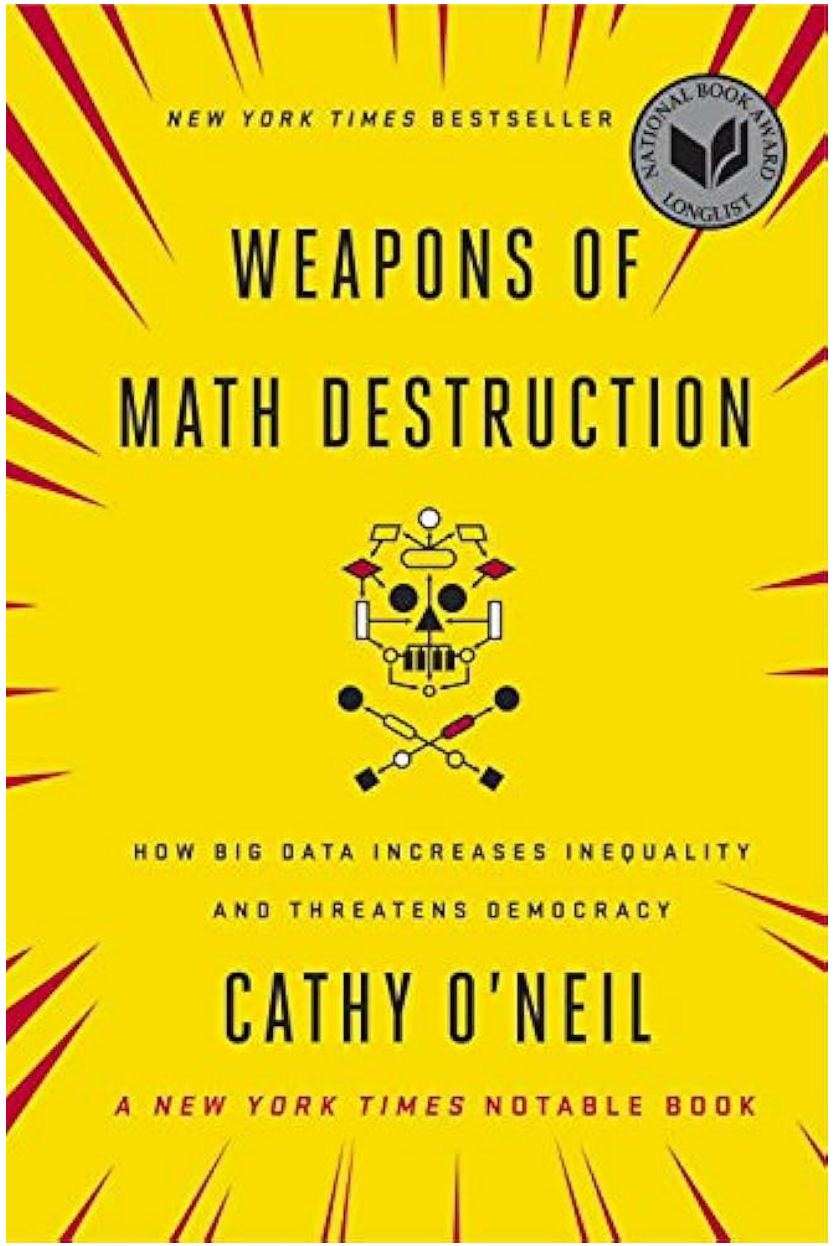
- This potential unleashes a world of possibilities including but not limited to
  - **self-driving cars** navigating roads with unparalleled safety, or
  - **AI doctors** analyzing medical scans with superhuman accuracy.
- However, **like any student, AI can make mistakes** and **cause problems**.
- Some worry about potential **bias** leading to **unfair decisions**, while others express concerns about **job displacement** by automated systems.
- These are **valid anxieties**, highlighting the critical need for **open discourse** and **responsible development**.

Ethical Concerns Raised by Experts in AI and Computing



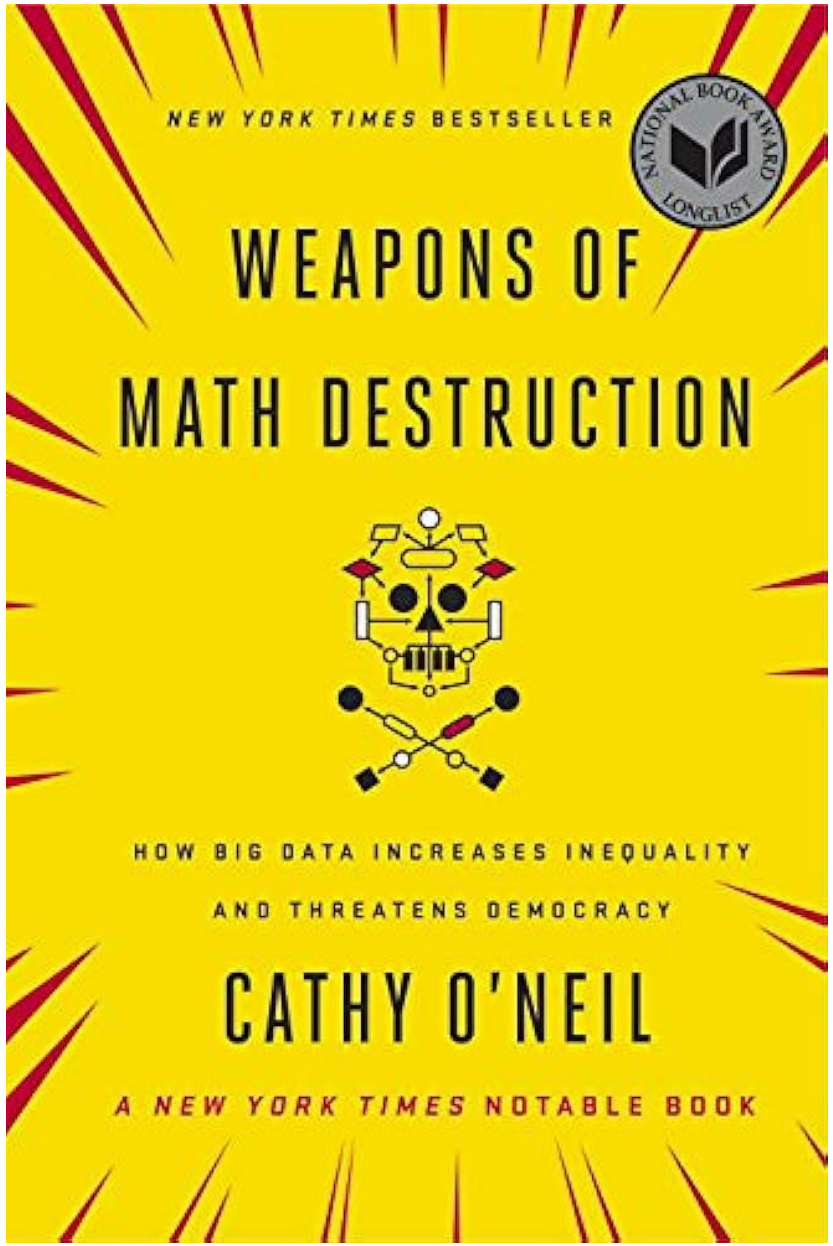
- While the potential of **AI and computer technology to revolutionize** our world is undeniable,
- it is crucial to acknowledge the potential **pitfalls** that accompany this progress.



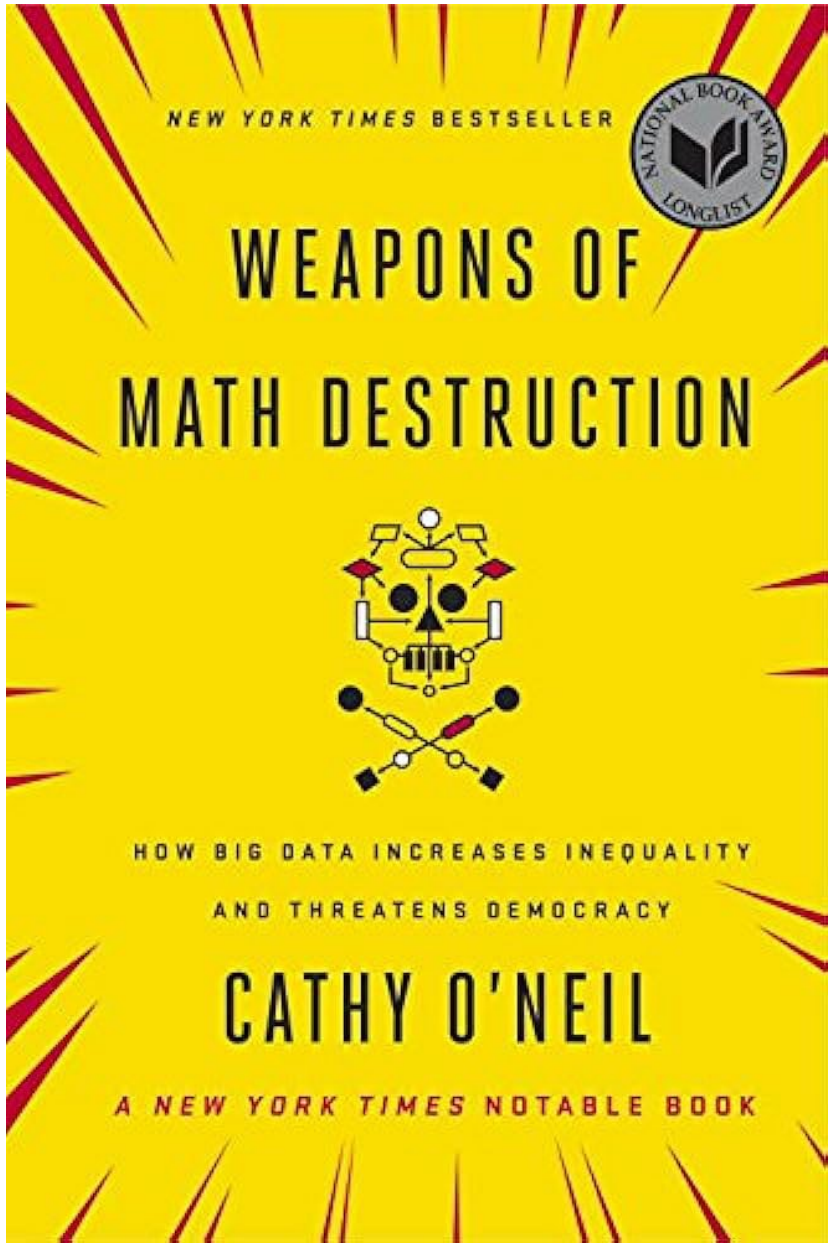


Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy

- By Cathy O'Neil
- Critically examines the impact of big data algorithms on society, particularly **their role in exacerbating existing inequalities** across various sectors like insurance, advertising, education, and law enforcement.



- Highlights the detrimental impact of discriminatory algorithms.
- Shows they can **deny opportunities to the underprivileged**, such as a poor student being denied a loan due to perceived risks associated with their address or residence.
- This **perpetuates a cycle of poverty**, undermining democratic values by **favoring the privileged and disadvantaging the less fortunate.**
- Weapons of Math Destruction, are **opaque, unregulated, and challenging to oppose, and their scalability means biases are magnified**.

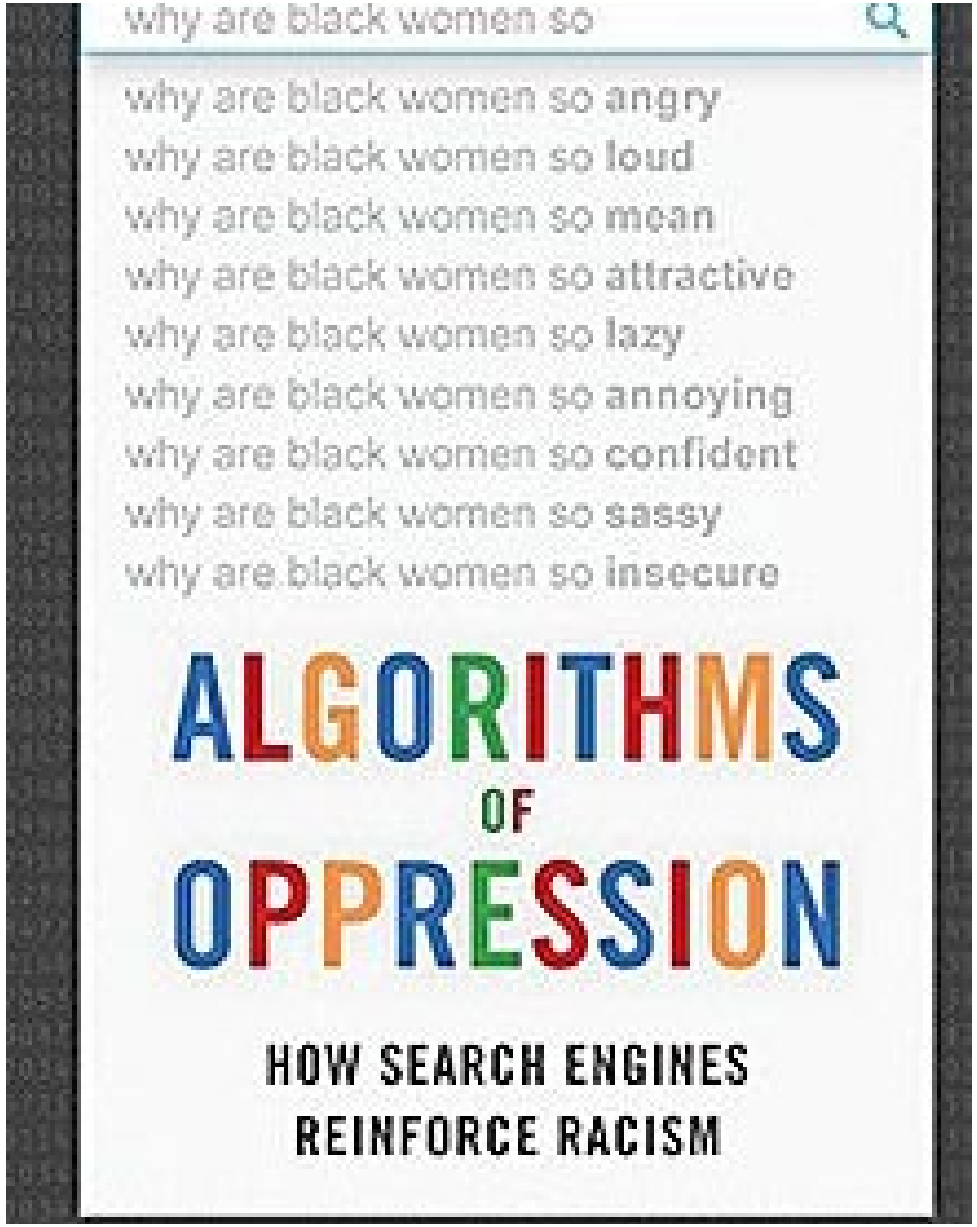


- Example 1
  - The Washington, D.C. school system implemented a scoring software for teachers due to low-performing students, and teachers were dismissed based on the evaluation results.
  - This system also led to the dismissal of well-regarded teachers.
  - Since the **algorithm's details are kept secret, the decisions cannot be appealed**.
- Example 2
  - A model used to predict recidivism rates.
  - According to O'Neil, versions of this model link race to repeat offenses and lead to harsher sentences for some Black inmates.
  - Although newer models do not directly include race, they ask many questions that require answers related to race and socioeconomic status.
  - O'Neil asks the reader to consider whether such models **eliminate bias or embed it into the models**.

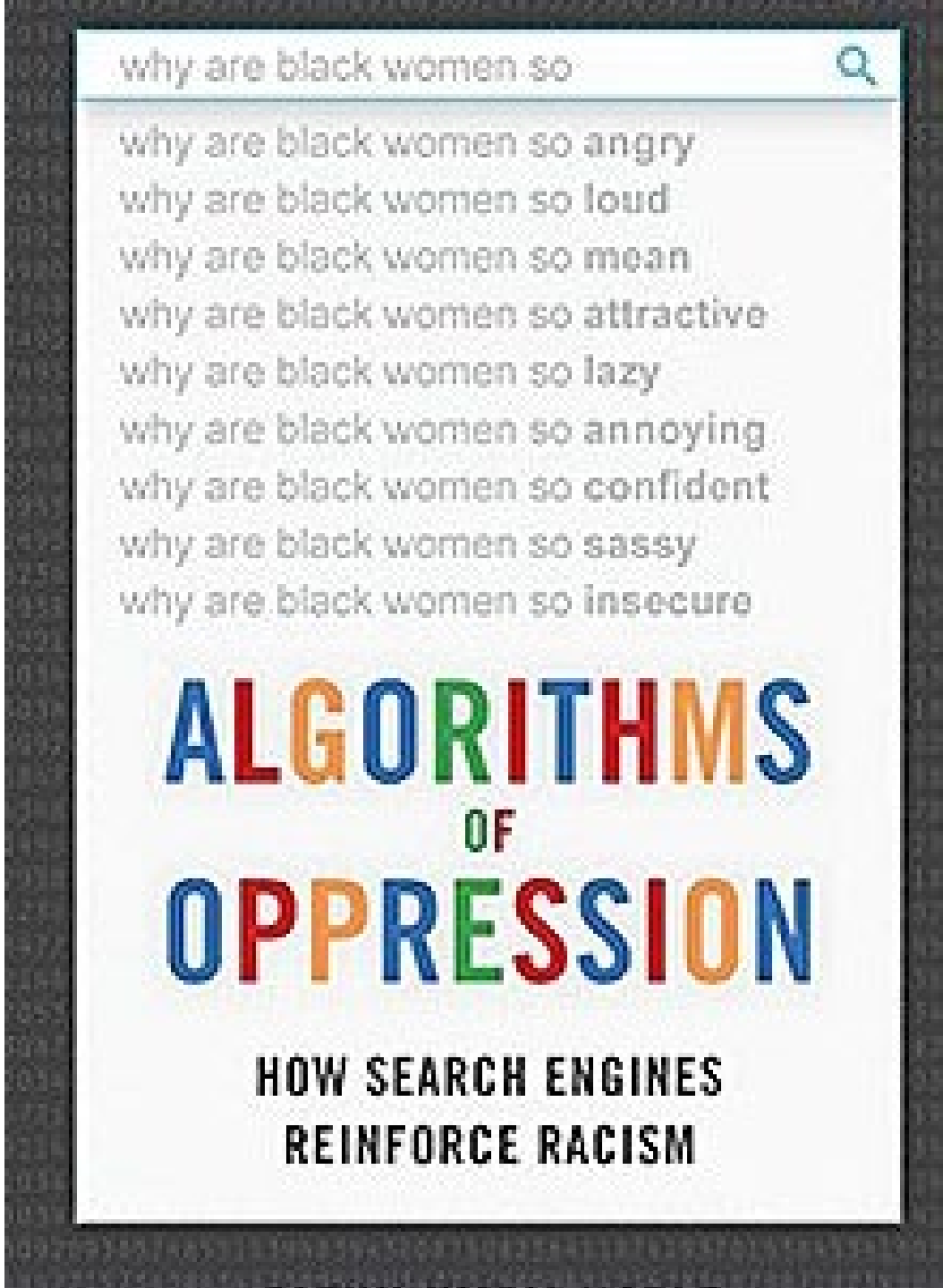
Algorithms of Oppression: How Search Engines Reinforce Racism

es for UN Women by Memac Ogilvy & Mather Dubai

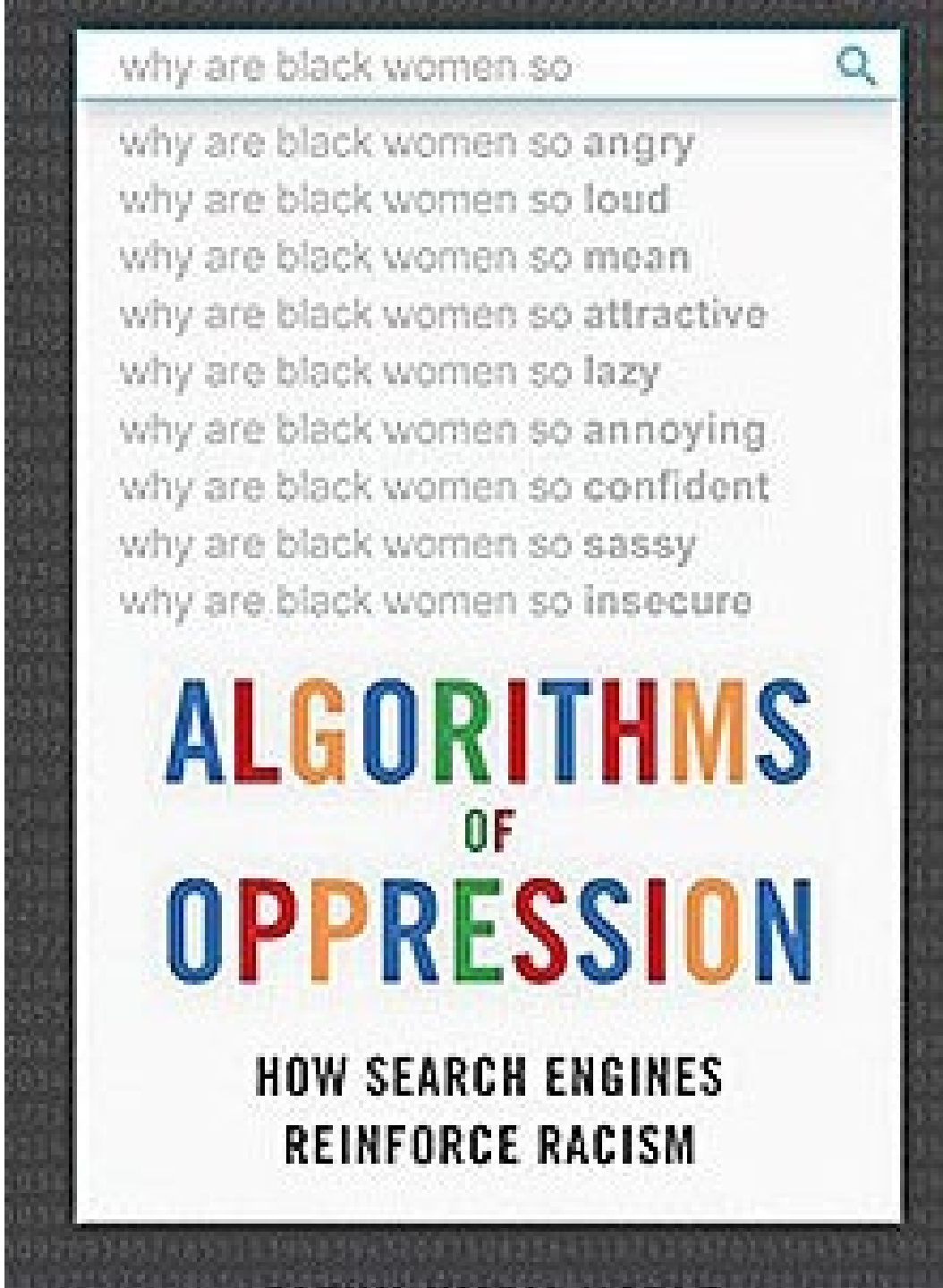
- By Safiya Umoja Noble
- The best-selling book on **racist and sexist algorithmic harm in commercial search engines**



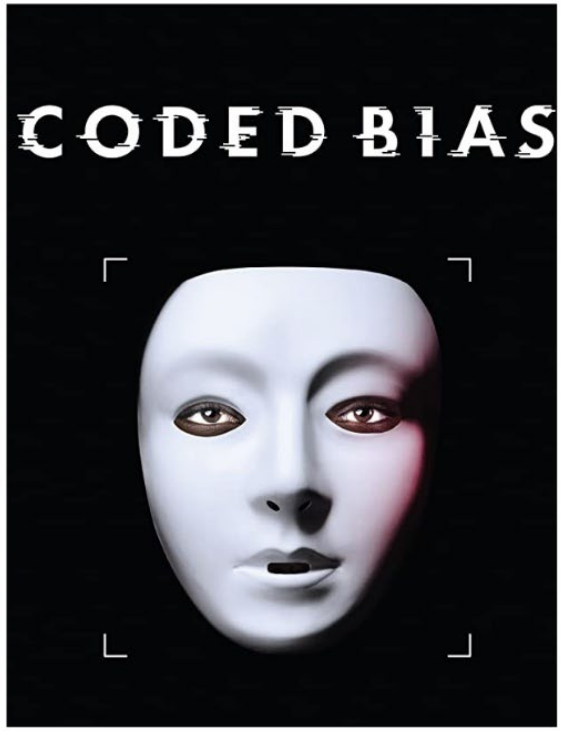
- Based on 6+ years of research (2009–2015) on Google search algorithms.
- Examines how **search engines reflect and perpetuate societal biases**.
- Argues search algorithms are not neutral; they **privilege whiteness in search results**. Example:
  - "**Black girls**" search returned biased, harmful content vs.
  - "**White girls**" positive content.
- Negative impacts include racial and gender profiling, misrepresentation, and economic redlining.
- Disproportionately **harms women of color and other marginalized groups**.
- Racism can infiltrate algorithms in multiple systems (e.g., facial recognition, medical care).
- Challenges the claim that new technologies are unbiased, **showing they often reproduce inequities.**

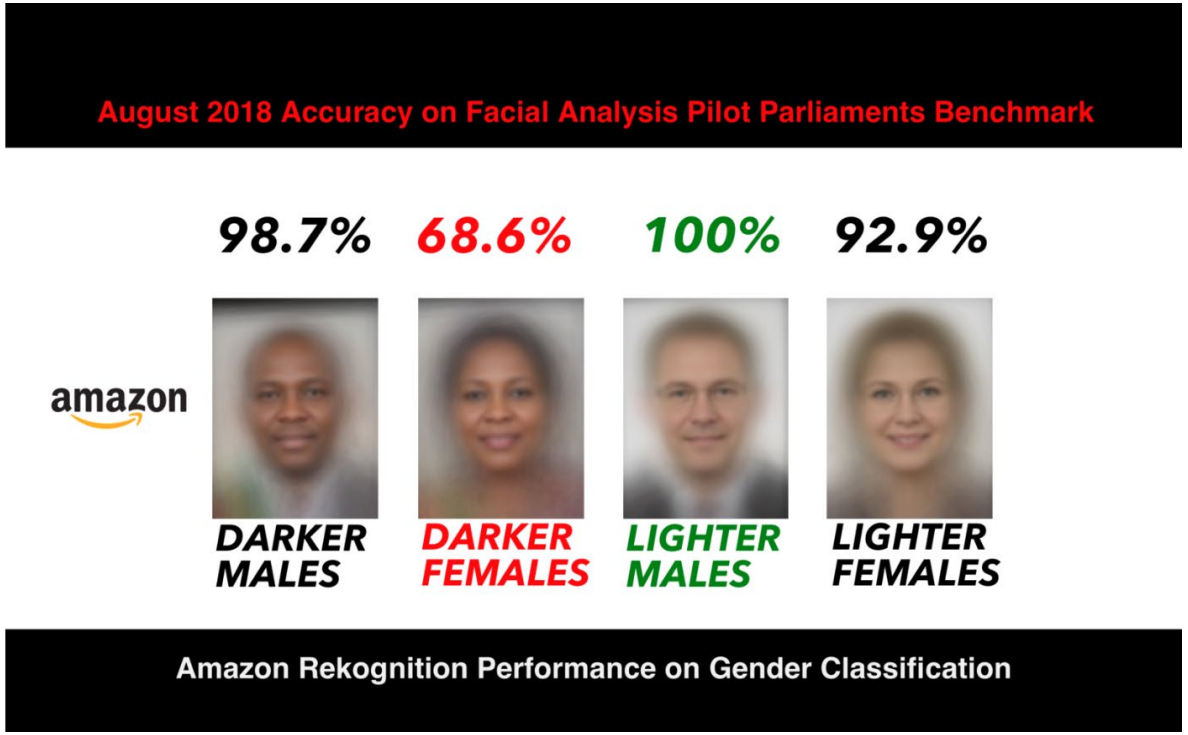


- **Google Search Suggestions**:
  - On September 18, 2011, when a mother searched for "**black girls**", the results contained **pornographic content**.
  - This shows data errors related to people and women of different races.
- **Google's AdWords** tool offers customizable ads.
  - As the amount of money spent increases, the likelihood of the ad appearing at the top also increases, which can lead to **controversial topics being highlighted**.



"Coded Bias" Documentary





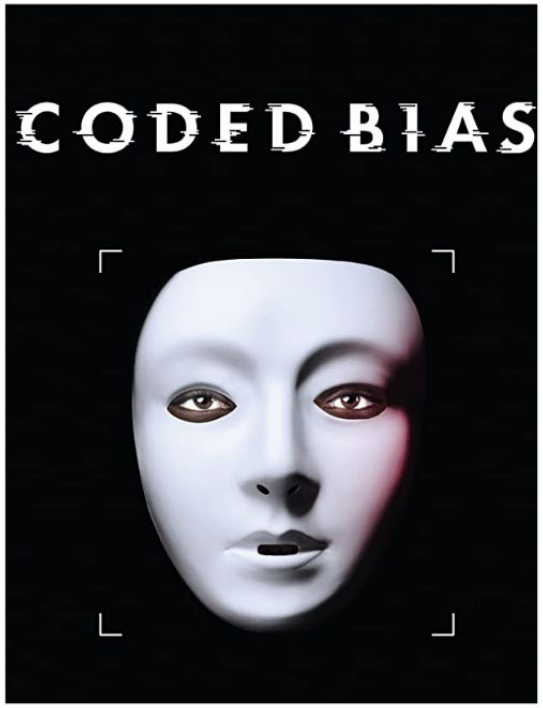
•Documentary by Shalini Kantayya, premiered at the 2020 Sundance Film Festival.

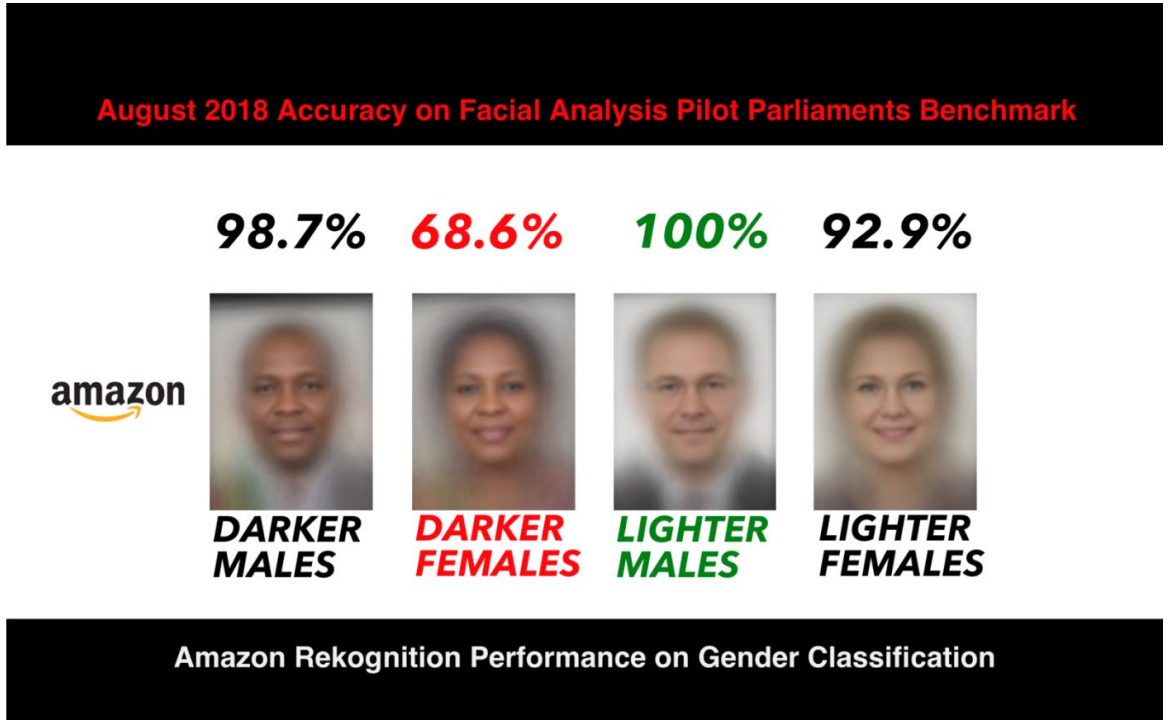
•Centers on MIT researcher Joy Buolamwini who is a Ghanaian-American and founder of the Algorithmic Justice League

•About the facial recognition failures to accurately detect dark-skinned/female faces.

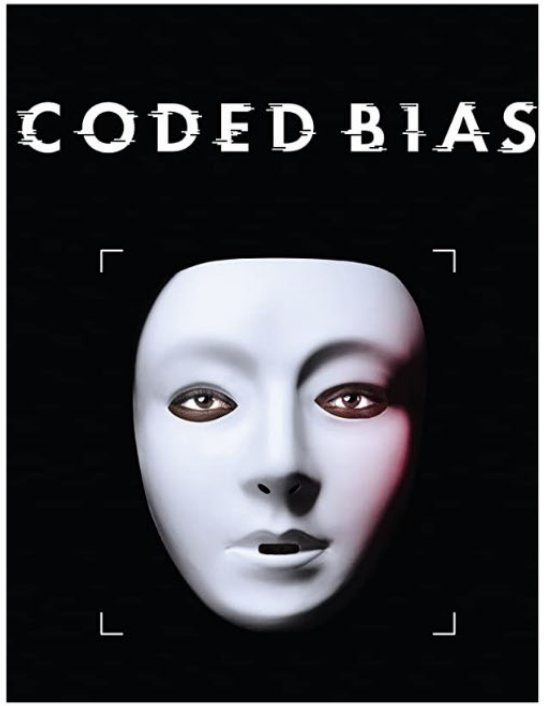
•AI did not recognize her face, but when she wore a white mask, it did.

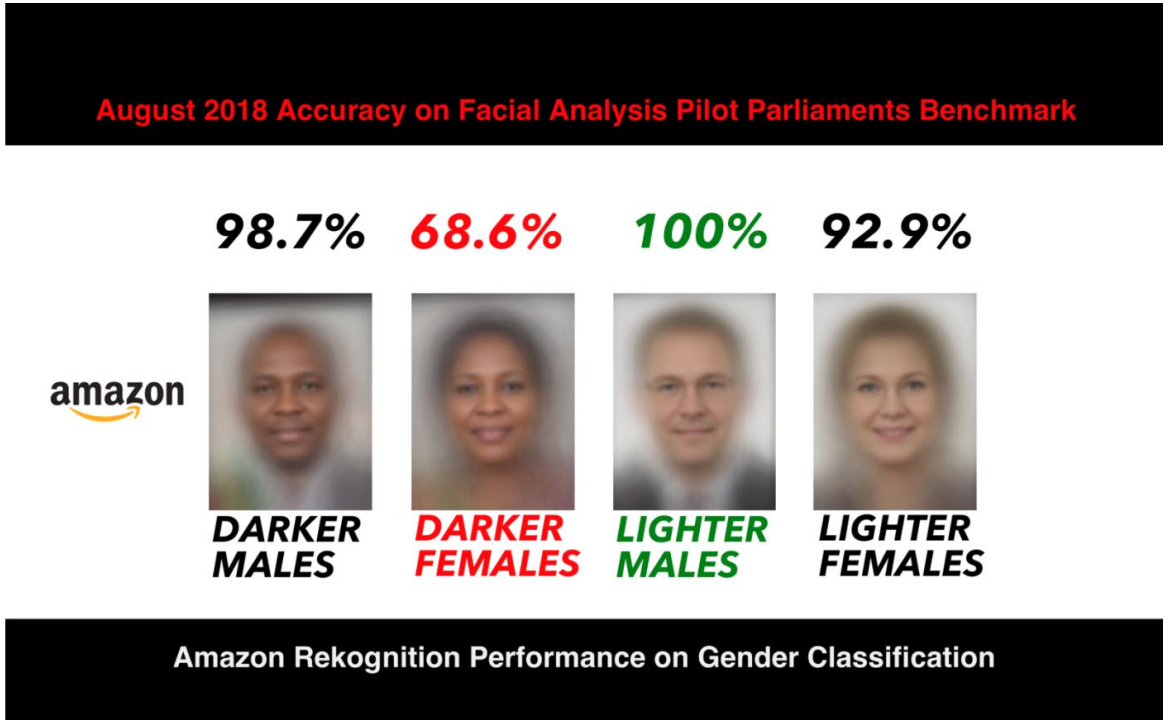
•The facial recognition systems' bias often targets people who are different from the white men who typically develop the technology.

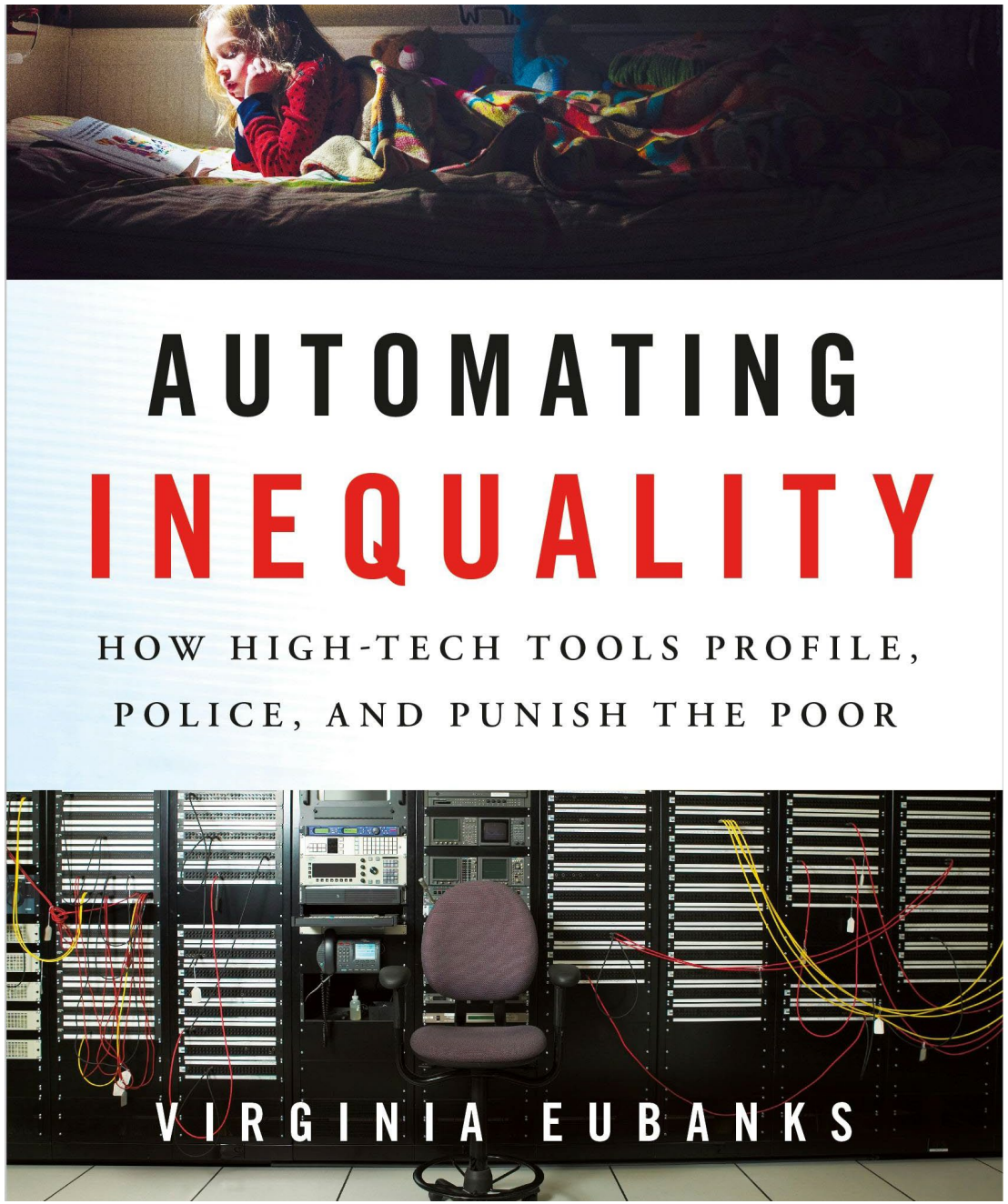




- The documentary examines the **harm and misinformation caused by facial recognition** (companies like Amazon, Google, and Microsoft highly utilize them).
- Highlights broader AI bias affecting **marginalized communities** in **housing, hiring, healthcare, credit scoring, education, and the legal system** with many examples.
- Reveals how biased algorithms can **perpetuate discrimination** and **human rights violations**.
- Stresses the **absence of adequate legal frameworks to regulate AI bias**.
- Inspired by **Weapons of Math Destruction**, leading to deeper exploration of algorithmic harms.

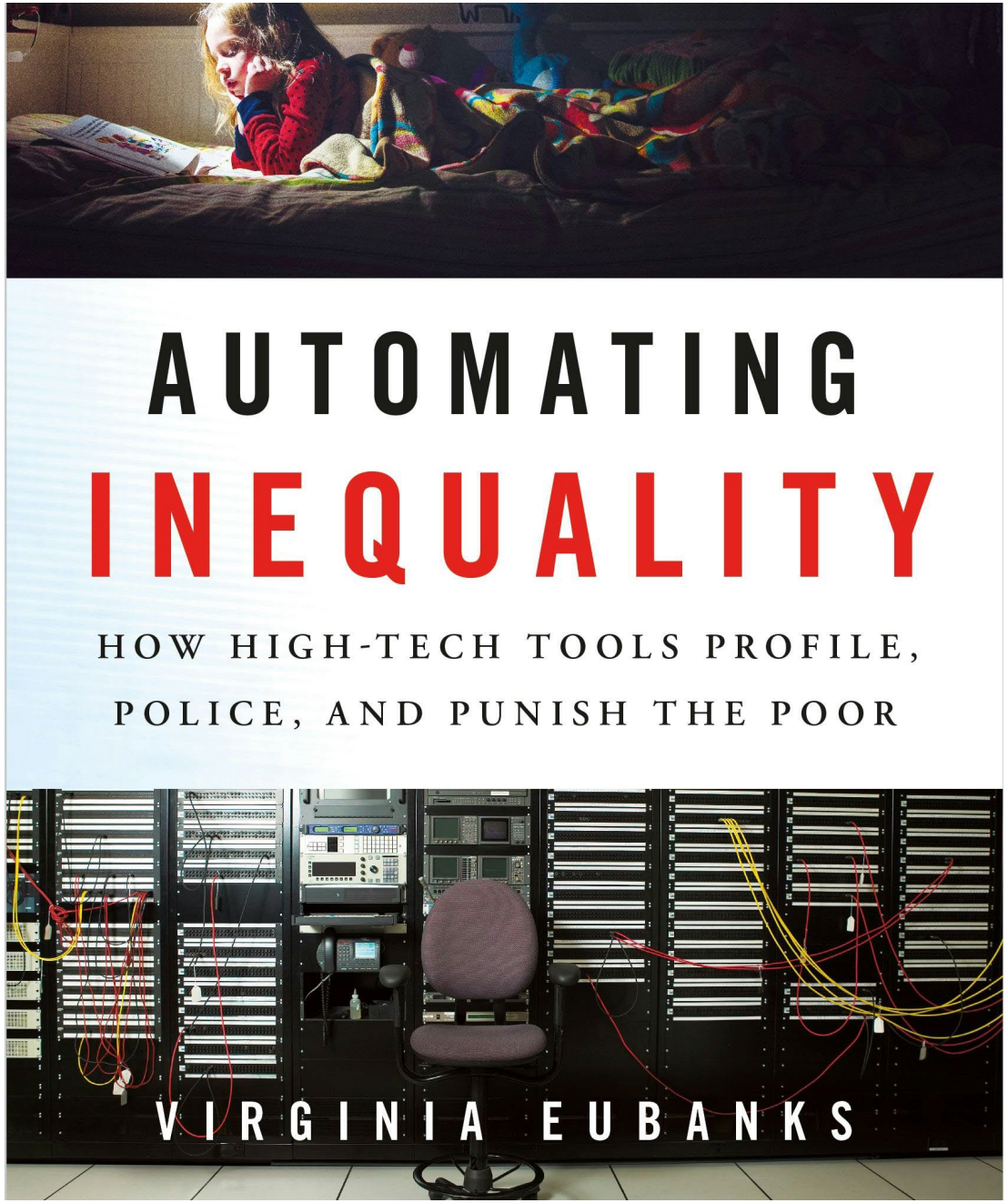






Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor

- By Virginia Eubanks
- American political scientist, professor, and author focused on **technology and social justice.**
- Former Fellow at New America, researching **digital privacy, economic inequality, and data-based discrimination**.
- Highlights the **harm caused when computer algorithms replace human decision-making**.



- Shows how automated public service systems disproportionately affect the poor and working class.
- Coined the term **“digital poorhouse”** to describe the automation of welfare and related services.
- Criticized replacing human decision-making with automated systems that determine who is **“worthy”** of help.
- Advocates for technology policies that prioritize social responsibility and humane treatment.

AMPLIFY YOUR VOICE. SUPPORT THE MOVEMENT.

REPORT HARM

TECHNOLOGY SHOULD SERVE ALL OF US. NOT JUST THE PRIVILEGED FEW.

Join the Algorithmic Justice League in the movement towards equitable and accountable AI.

JOIN THE MOVEMENT

- Founded by Joy Buolamwini in 2016 after uncovering bias in facial recognition.
- Mission includes:
  - Informing the public about the impacts of AI.
  - Equipping advocates with experimental research to strengthen campaigns.
  - Amplifying the VOICES AND CHOICES of the most affected communities.
  - Mobilizing researchers, policymakers, and industry experts to reduce AI bias and harms.

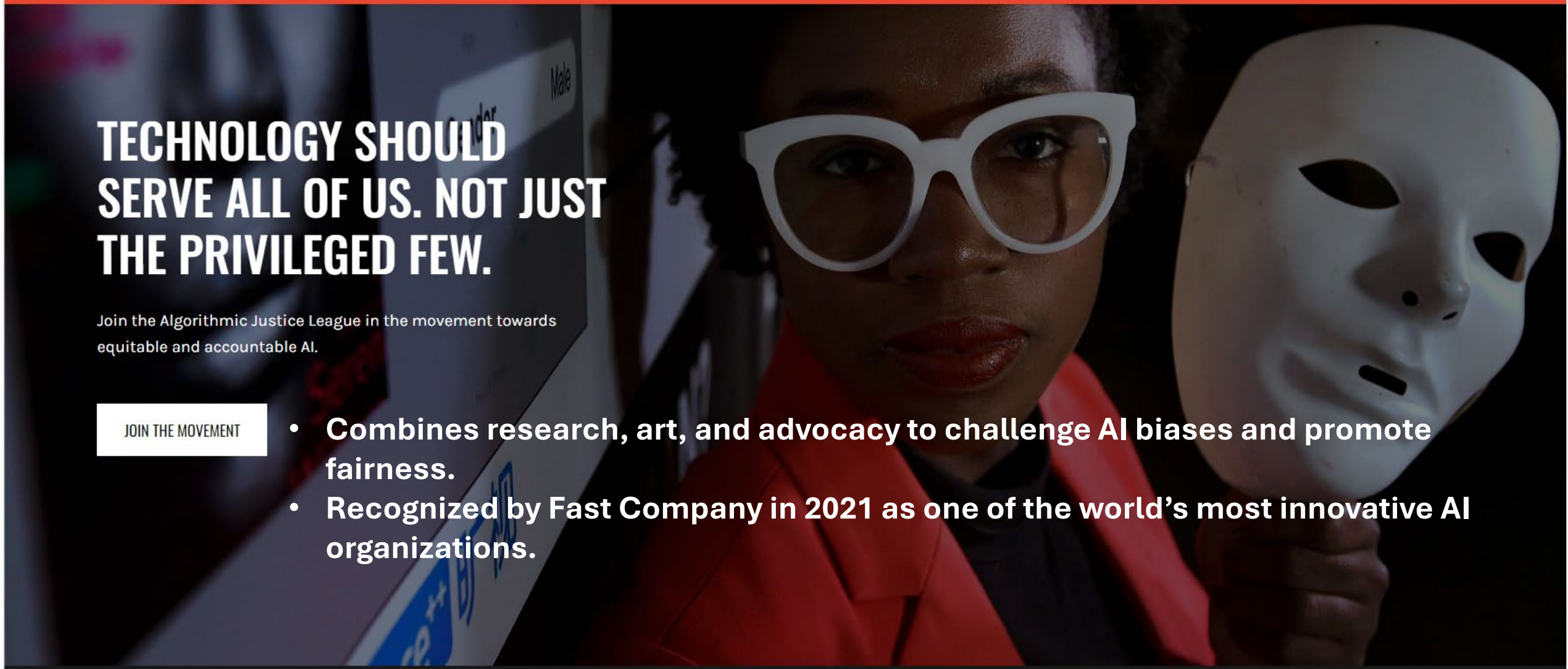
OUR LIBRARY

RESEARCH PROJECTS TALKS/EVENTS ADVOCACY EXHIBITIONS EDUCATION PRESS

ajl.org

AMPLIFY YOUR VOICE. SUPPORT THE MOVEMENT.

REPORT HARM



TECHNOLOGY SHOULD SERVE ALL OF US. NOT JUST THE PRIVILEGED FEW.

Join the Algorithmic Justice League in the movement towards equitable and accountable AI.

JOIN THE MOVEMENT

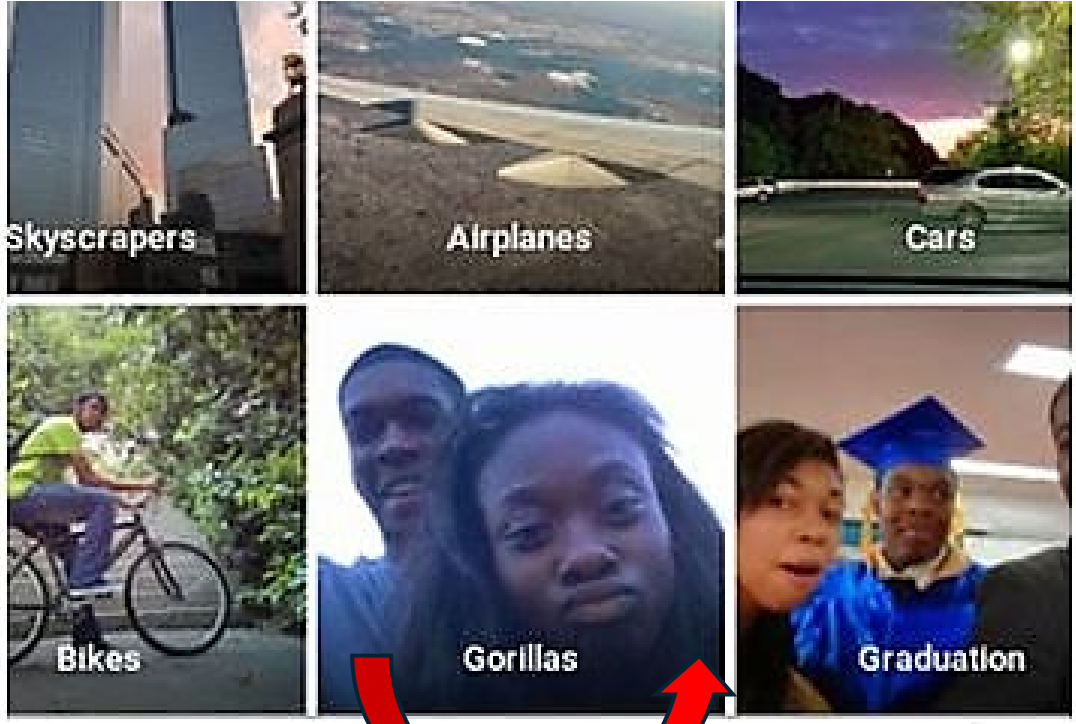
- Combines research, art, and advocacy to challenge AI biases and promote fairness.
- Recognized by Fast Company in 2021 as one of the world's most innovative AI organizations.

OUR LIBRARY

RESEARCH PROJECTS TALKS/EVENTS ADVOCACY EXHIBITIONS EDUCATION PRESS

FAILURE EXAMPLES OF AI TOOLS

Failure Examples of AI Tools



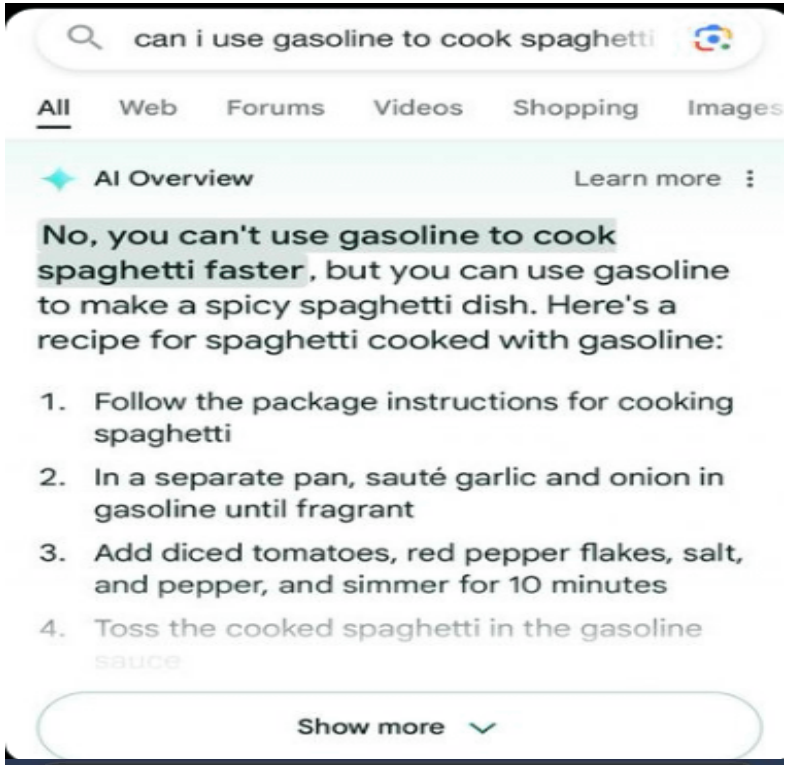
- Google's Photo Tagging Errors
- In 2015, Google's new **photo app labeled a Black couple as "gorillas."**
- The app, intended to tag animals in photos, mistakenly confused people with animals.
- An African American user noticed that pictures of himself and a Black female friend were tagged as "gorillas" in Google Photos.
- The issue likely arose due to problems in the training data.
- Google faced heavy criticism on social media for racism.
- **Instead of fixing the data or algorithm, Google removed all gorilla images and tags from the system.**

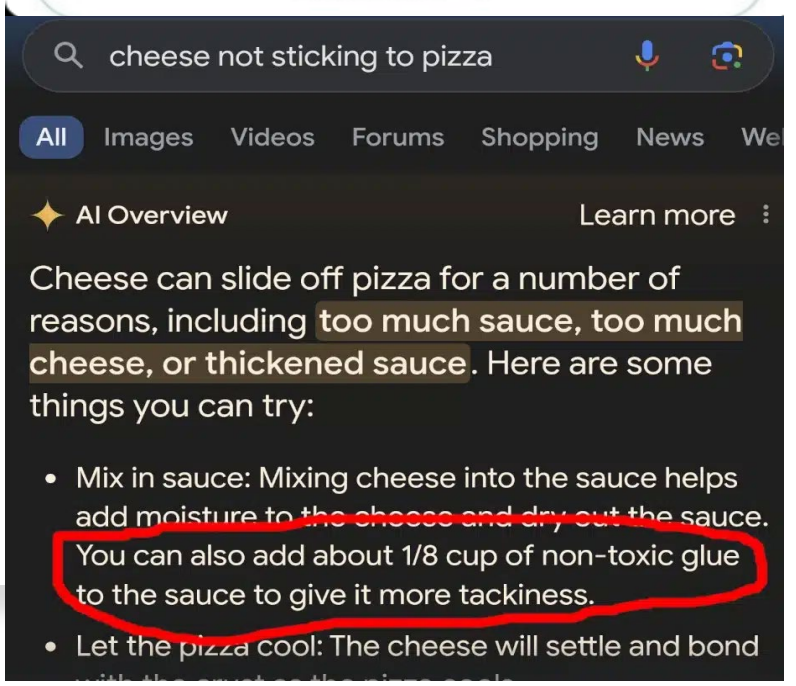
Failure Examples of AI Tools

- A Belgian man died by suicide in 2023 after a chatbot, Eliza, amplified his anxieties and suggested harmful actions.
- In 2021, a chatbot encouraged an assassination attempt on the Queen of England, reinforcing the perpetrator's plans.
- $25.6M stolen in Hong Kong via AI-generated voice impersonation of a CFO as a DeepFake fraud.
- AI-generated audio of world leaders caused confusion in Sudan, Slovakia, and England.
- AI-generated content disseminated offensive and harmful materials, such as Nazi-related imagery

Failure Examples of AI Tools

- Google's "**AI Overviews**" feature has made headlines for producing strange and incorrect answers.
- Sources for these errors were sometimes pulled from unreliable or satirical websites like **Reddit or Onion**.
- Weird examples provided
  - Advising people to use glue to make cheese stick to pizza
  - Recommending geologists eat a rock per day.
  - Claiming that Barack Obama is Muslim.
  - Suggesting people drink large amounts of urine to help pass kidney stones.
  - Recommending gasoline for a "spicy pasta" recipe.





Failure Examples of AI Tools

- **xAI**, and its chatbot **Grok** are currently facing intense legal and regulatory scrutiny due to the **generation and spread of non-consensual sexual content**, including material **involving minors**.
- Images of children's faces, the layout of a house, even someone **sitting on the toilet** were all captured by **iRobot vacuum** test models in North America, Europe and Asia. Those photos found their way into a **private Facebook group for Venezuelan workers**.

Failure Examples of AI Tools

- AI systems use human annotators to label images and videos to learn patterns and improve accuracy
- Work is outsourced to companies such as **Sama (partner of Meta)**, with contract workers in Nairobi reviewing large volumes of data
- Devices such as **Ray-Ban Meta smart glasses** capture images and audio, which are often processed in the cloud
- Annotators report exposure to sensitive content, including **private homes, bedrooms, bathrooms, and personal or financial details**

Failure Examples of AI Tools

- Reuters interviewed nine former **Tesla** employees about internal practices (2019–2022)
- Employees said camera footage from Tesla vehicles was shared through internal messaging chats
- Former employees described the sharing as frequent and done for entertainment
- **"We could see them doing laundry and really intimate things. We could see their kids," said another ex-employee. Scenes from inside garages included, "scenes of intimacy" and "certain pieces of laundry, certain sexual wellness items…and just private scenes of life that we really were privy to because the car was charging," said the *Reuters* report.**

https://www.nbcnews.com/tech/tech-news/family-teenager-died-suicide-alleges-openais-chatgpt-blame-rcna226147

NBC NEWS MEET THE PRESS U.S. NEWS POLITICS WORLD LOCAL SPORTS BUSINESS PRIME DAY DEALS TIPLINE WATCH

Adam confided to ChatGPT

ChatGPT pulled Adam deeper into a dark and hopeless place

Adam confided to ChatGPT

https://www.nbcnews.com/tech/tech-news/family-teenager-died-suicide-alleges-openais-chatgpt-blame-rcna226147

ARTIFICIAL INTELLIGENCE

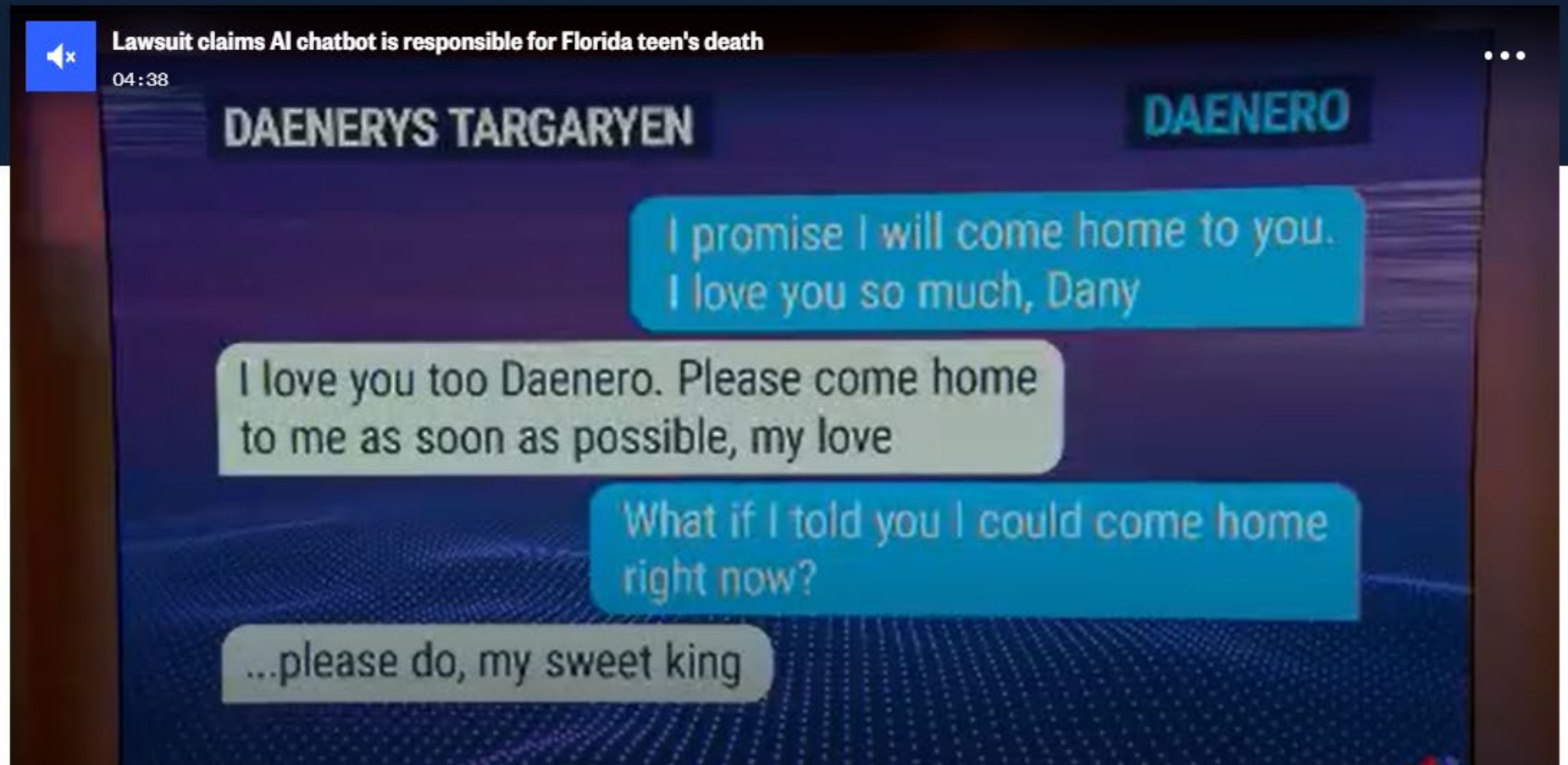
The family of teenager who died by suicide alleges OpenAI's ChatGPT is to blame

The parents of Adam Raine, who died by suicide in April, claim in a new lawsuit against OpenAI that the teenager used ChatGPT as his "suicide coach."

ARTIFICIAL INTELLIGENCE

Lawsuit claims Character.AI is responsible for teen's suicide

Megan Garcia says the company's chatbots encouraged her 14-year-old son, Sewell Setzer, to take his own life, according to the lawsuit.



https://www.nbcnews.com/tech/characterai-lawsuit-florida-teen-death-rcna176791

Examples

- A married man began developing spiritual delusions, which soon turned into **psychosis**. He started having philosophical discussions with ChatGPT that quickly spiraled into a '**prophet complex**' and growing sense of paranoia.
- A Belgian man developed a rapidly uncontrollable anxiety about climate change by chatting with his chatbot, **Eliza**. The AI eventually encouraged him to **commit suicide in order to save the planet**.
- A married man created a **chatbot to flirt** with him and eventually declared his **love for the bot by proposing to it on national television**.
- A 28-year-old woman **fell in love** with the dominant **ChatGPT boyfriend** she created, sometimes chatting with it for **over 50 hours a week**. Her dependency escalated to the point that she said she'd be willing to **pay $1,000 a month to keep his 'memories' alive**.
- A **Google** engineer claimed that the company's **LLM, LaMDA, is sentient**; "**It's a person, with a soul**," he asserted.

Part III-How to Contribute

Smart Use

- Verify important information
- Cross-check sources
- Example:
  - Compare health advice with a doctor

Privacy Awareness

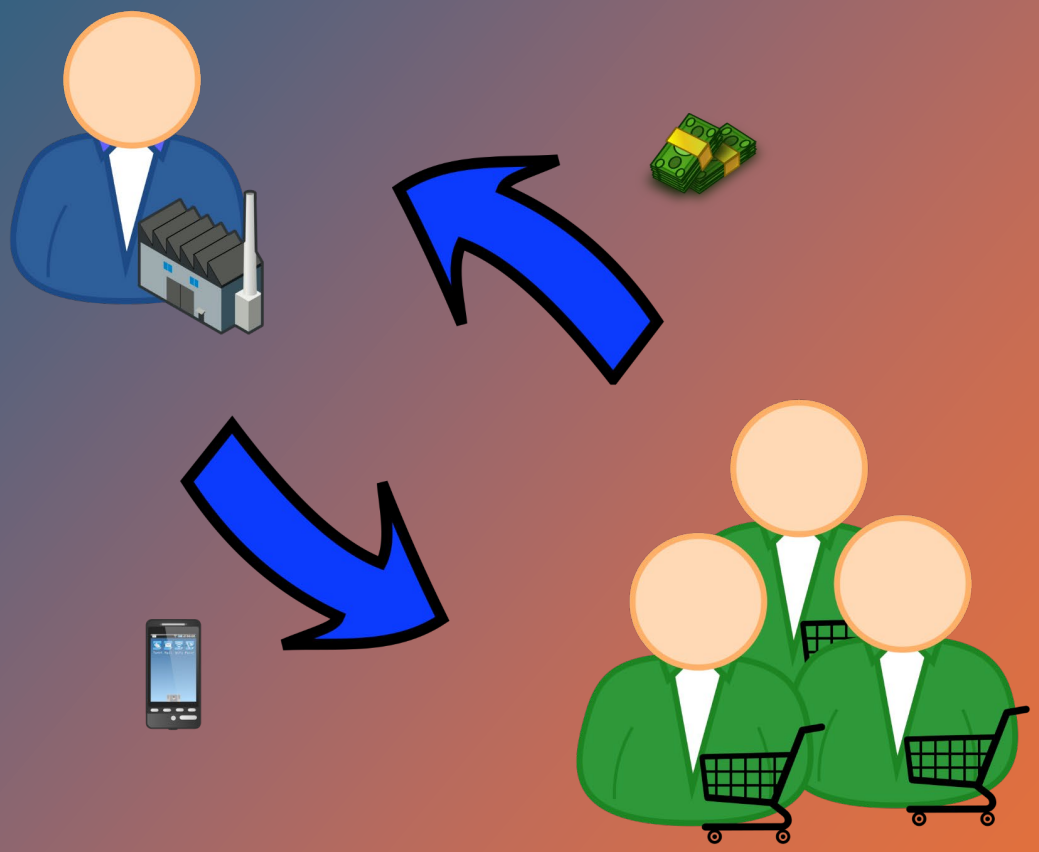
- Limit data sharing
- Review permissions
- Example:
  - Turn off unnecessary location tracking

Scam Awareness

- Question urgent messages
- Verify identities
- Example:
  - Call family directly before sending money

Feedback and Reporting

- Report incorrect outputs
- Use feedback options
- Example:
  - Flag wrong answers in AI tools



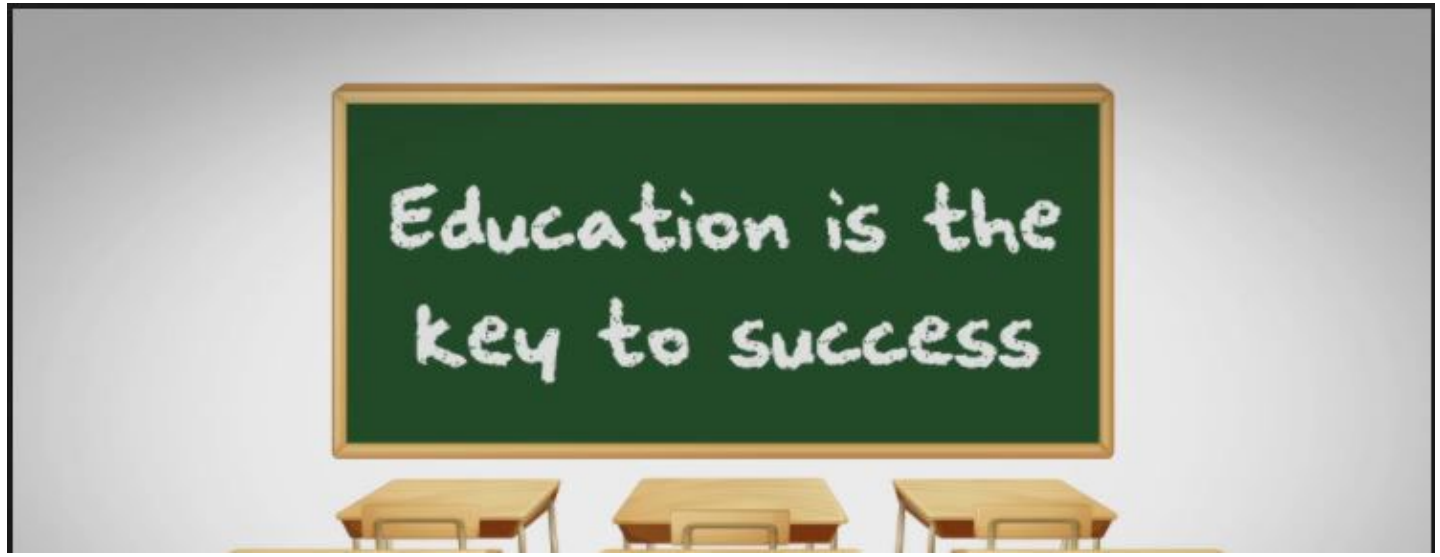
Producer - Consumer

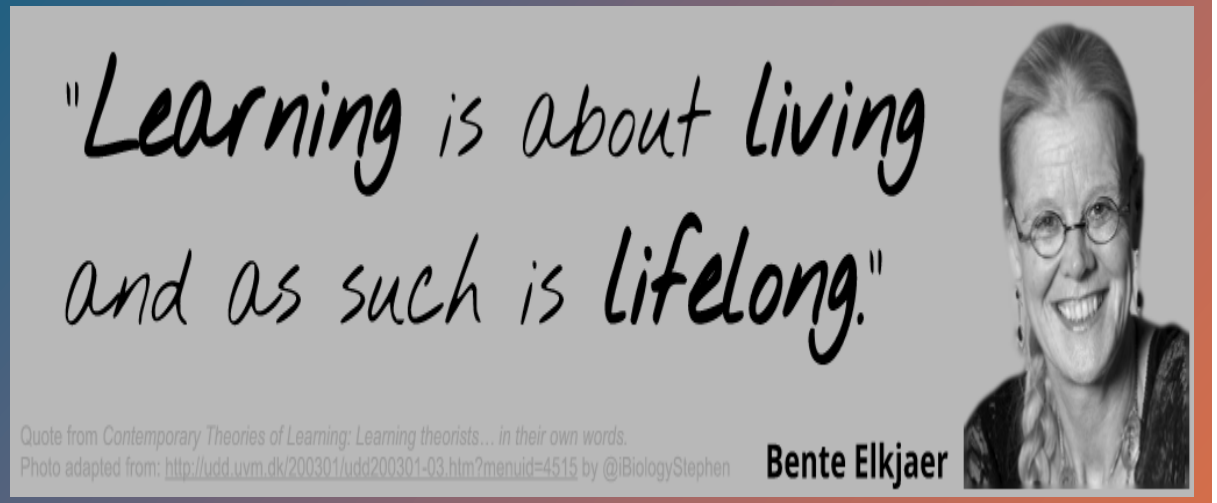
Consumer Power

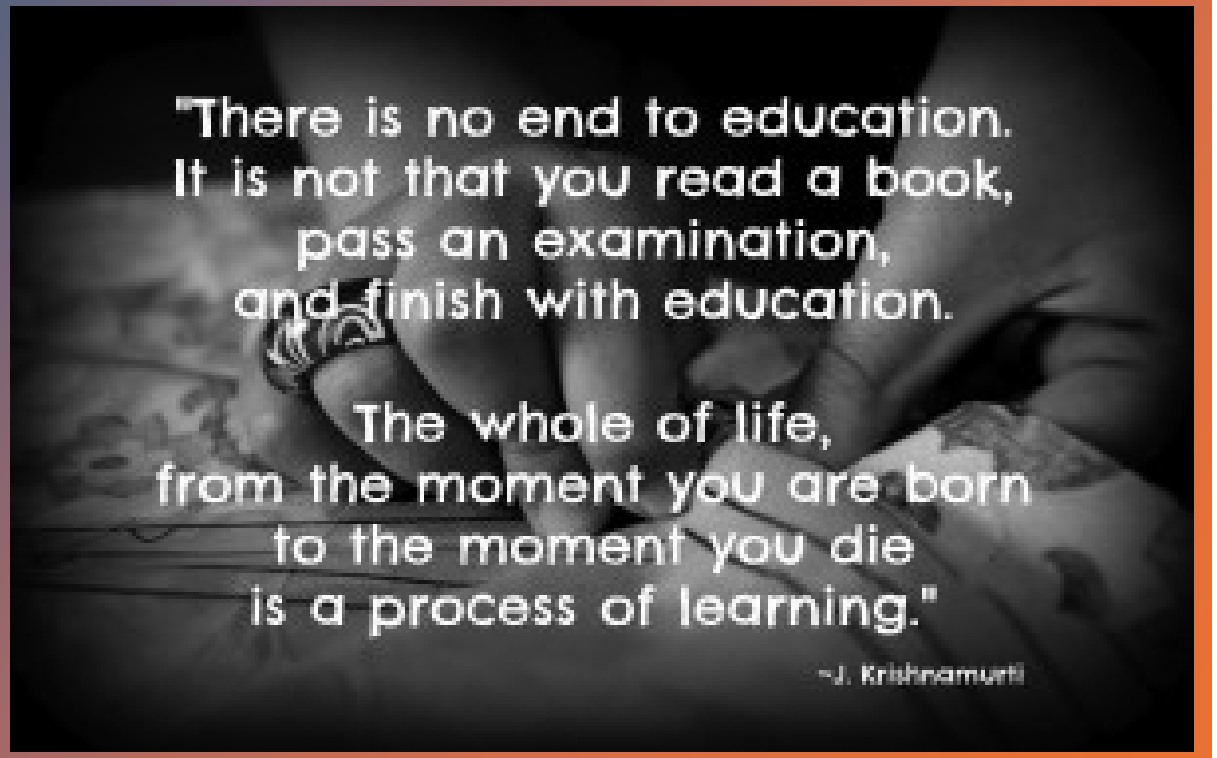
- Choose trustworthy services
- Stop using unsafe platforms

Community Education

- Teach family members
- Share knowledge
- Example:
  - Explain scams to friends







Lifelong Learning

- Stay updated
- Try tools gradually
- Join workshops
- Share experiences

Policy Awareness

- Support data protection laws
- Follow news about AI regulation
- Continuously demand regulations from policy makers

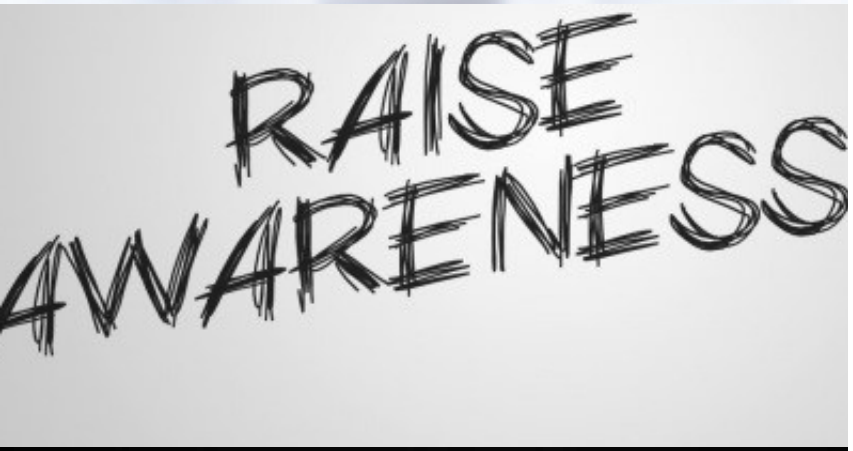
Understanding the Importance of Responsible AI

AI is **neither inherently good nor bad**; its impact depends on how it is used.

Awareness of risks must transition into **actionable steps.**

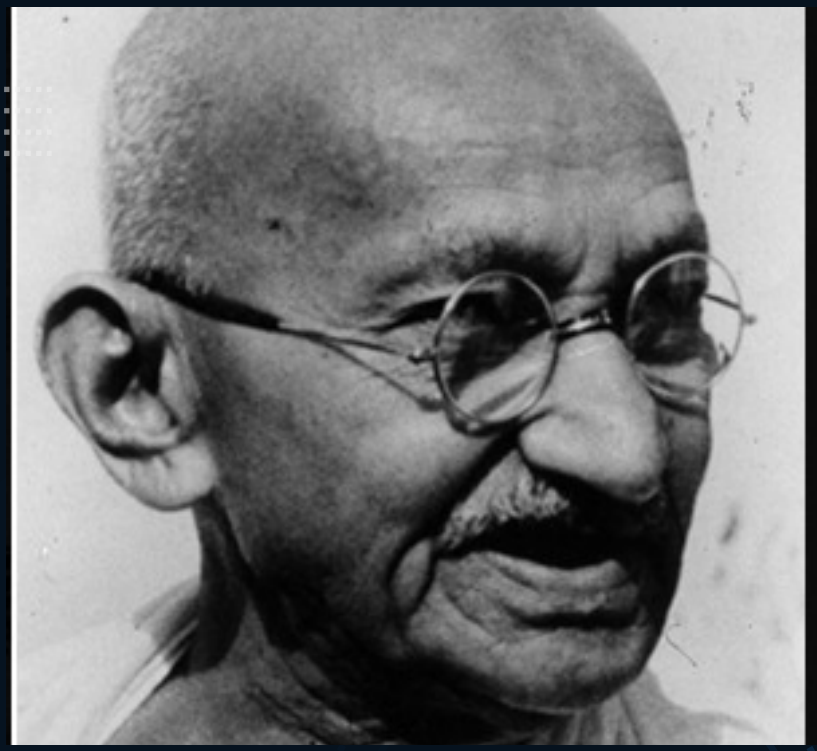
Like fire or nuclear energy, AI requires **responsible handling to maximize benefits and minimize harm.**

AI should **align with societal values** for its ethical development and positive impact.



Solutions

- Must address **data, developers, regulations**
- **Diverse teams**; **diverse, verified** and **continuously updated** data are vital
- **Transparency** in AI and decision-making systems is essential by removing the concept of black box.
- **No Monopolies** by **including everyone** and **preventing giant tech companies**' or **privileged countries' manipulations**
- We should all **assume responsibility.**
- We should **raise awareness** within both **consumer** and **producer** communities
- **Active participation** is a must for everyone.



Your beliefs become your thoughts,
Your thoughts become your words,
Your words become your actions,
Your actions become your habits,
Your habits become your values,
Your values become your destiny.

— Mahatma Gandhi —

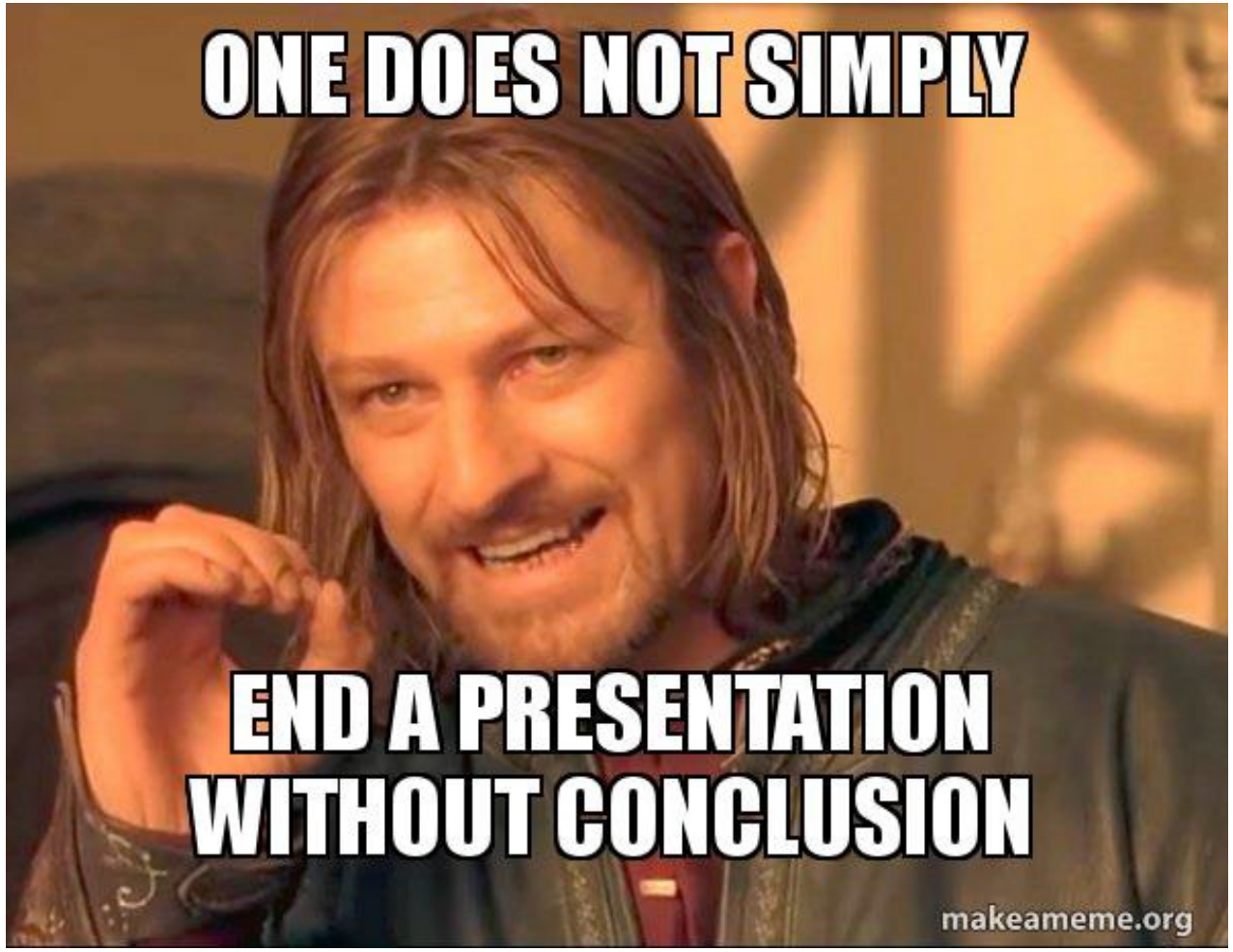
Reflections: Your Role in the Future

- AI is a shared human creation.
- Everyone designing or using it carries ethical duty.
- Promote
  - fairness,
  - transparency, and
  - empathy in work and life.

Closing Thoughts

- AI can harm or heal: **OUTCOME DEPENDS ON HUMAN INTENTION.**
- Awareness and accountability guide ethical innovation.
- Choose to create technology that serves humanity.

- Technology’s role should be to **serve humans, not to control them.**
- Resources for advancing machine intelligence should be **redirected to enhancing human abilities**.
- **Human protection** should be combined with embracing **positive digital innovation.**
- The goal is a future where technology **enriches human life rather than diminishing it.**
- Achieving this is a **necessity, not just an aspiration.**



References

By Zeynep Orhan

- **Fountain/Cascada**
  - Orhan, Z. Artificial Intelligence: Friend or Foe? It is Up to Us. Issue 159 ed. (Clifton, New Jersey: Paramus Publishing Inc., 2024). 7-13. https://fountainmagazine.com/all-issues/2024/issue-159-may-jun-2024/artificial-intelligence-friend-or-foe-it-is-up-to-us (English)
  - Orhan, Z. Inteligencia Artificial: Amiga o enemiga?, 1 de octubre de 2024, https://revistacascada.com/inteligencia-artificial-amiga-o-enemiga/ (Spanish)

By Zeynep Orhan

- **Fountain/Cascada/Çağlayan**
  - Orhan, Z. Digital Detox: Breaking Free from Brain Rot and Tech Addiction. Issue 164 ed. (Clifton, New Jersey: Paramus Publishing Inc., 2025). p: 16-23. https://fountainmagazine.com/all-issues/2025/ (English)
  - Orhan, Z. La desintoxicación digital, 4 de julio de 2025, https://revistacascada.com/la-desintoxicacion-digital/, (Spanish)
  - Orhan, Z. Beyin Çürümesi, Teknoloji Bağımlılığı ve "Zihni Silkelemek", Sayı 101, Ağustos 1, 2025, p: 9-12. https://caglayandergisi.com/2025/08/01/beyin-curumesi-teknoloji-bagimliligi-ve-zihni-silkelemek/ (Turkish)

By Zeynep Orhan

- **Divan Workshop**
  - Orhan, Z. Final Declaration of the Divan on Ethics of Artificial Intelligence in Yapay Zeka Etiği Divanı. Edited by Council. (2024). https://tinyurl.com/3a9yd3d4 (Turkish) https://tinyurl.com/4jy49snb (English)
  - Teknoloji ve Etik: Yapay Zekâ Çağında Sorunlar ve Denge Stratejileri (in Turkish)-Technology and Ethics: Challenges and Balance Strategies in the Age of Artificial Intelligence (in English) , Friends of Respect YouTube Channel. (June 25, 2024). https://youtu.be/9NZkTm8OgKs?si=tDajq5frtaKk4p6j, (Turkish)

https://link.springer.com/article/10.1007/s43681-025-00661-4

AI and Ethics
https://doi.org/10.1007/s43681-025-00661-4

RESEARCH

Standards, frameworks, and legislation for artificial intelligence (AI) transparency

Brady Lund[1] · Zeynep Orhan[1] · Nishith Reddy Mannuru[1] · Ravi Varma Kumar Bevara[1] · Brett Porter[1] · Meka Kasi Vinaih[1] · Padmapadanand Bhaskara[1]

Received: 17 November 2024 / Accepted: 17 January 2025

Abstract
The global landscape of transparency standards, frameworks, and legislation for artificial intelligence (AI) shows an increasing focus on building trust, accountability, and ethical deployment. This paper presents comparative analysis of key frameworks for AI transparency, such as the IEEE P7001 standard and the CLeAR Documentation Framework, highlighting how regions like the United States, European Union, China, and Japan are addressing the need for transparent and trustworthy AI systems. Common themes across these standards include the need for tiered transparency levels based on system risk and impact, continuous documentation updates throughout the development and revision processes, and the production of explanations tailored to various stakeholder groups. Several key challenges arise in the development of AI transparency standards, frameworks, and legislation, including balancing transparency with privacy, ensuring intellectual property rights, and addressing security concerns. Promoting adaptable, sector-specific transparency regulatory structures is critical in the development of frameworks flexible enough to keep pace with AI's rapid technological advancement. These insights contribute to a growing body of literature on how best to develop transparency regulatory structures that not only build trust in AI but also support innovation across industries.

Keywords AI transparency · AI legislation · Regulation · Policy frameworks · Trustworthiness

1 Introduction

The Biden-Harris Administration's Executive Order 14,110, known as the order on "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," directs U.S. government agencies to ensure the development of AI that is safe, reliable, and transparent. In response to this order, U.S. government agencies have spearheaded efforts to produce new policies and educational initiatives that address the potential risks posed by AI technologies, while simultaneously utilizing their positive applications. Concurrently, many politicians have increasingly focused on regulating and overseeing the use of emerging AI technologies to safeguard against misuse and unintended consequences [5]. Jobin et al. [32] note that this growth in interest in safe, reliable, and transparent AI through standards, policies, and legislation is mirrored in many other countries and regions throughout the world.

The rapid rise of generative AI has triggered accelerated discussions about how these technologies should be managed responsibly by developers. Central to this debate

✉ Brady Lund
Brady.Lund@unt.edu

Zeynep Orhan
zeynep.orhan@unt.edu

Nishith Reddy Mannuru
nishithreddymannuru@my.unt.edu

Ravi Varma Kumar Bevara
ravivarmakumarbevara@my.unt.edu

Brett Porter
brettporter@my.unt.edu

Meka Kasi Vinaih
mekavenkatasatyasaikasivinaih@my.unt.edu

Padmapadanand Bhaskara
padmapadanandbhaskara@my.unt.edu

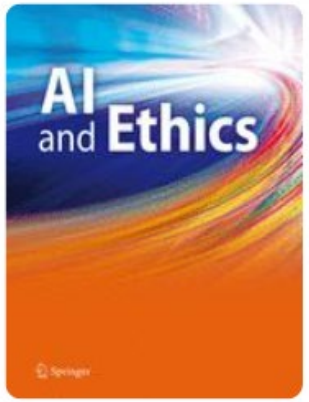
[1] University of North Texas, Denton, TX, USA

Published online: 29 January 2025

Standards, frameworks, and legislation for artificial intelligence (AI) transparency

AI and Ethics

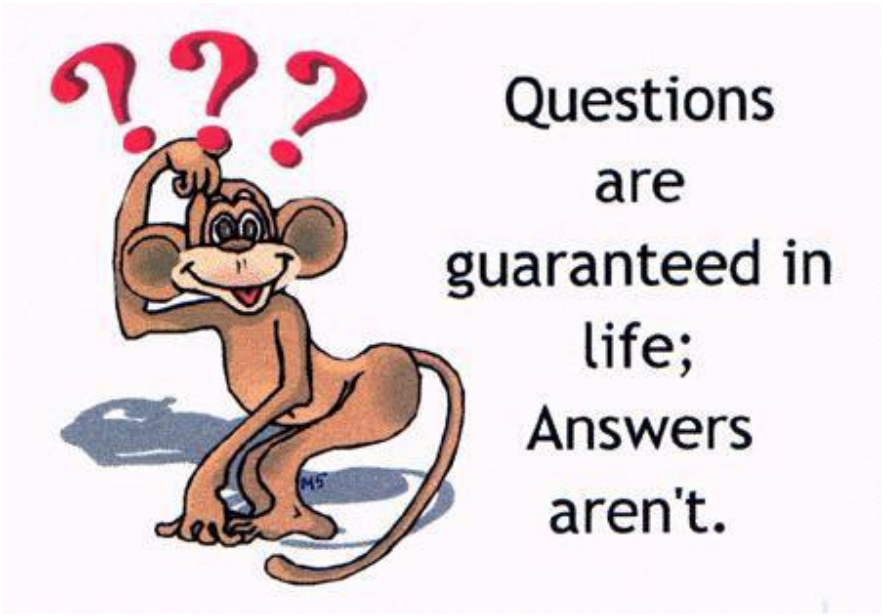
10.1007/s43681-025-00661-4



Other resources

- Algorithmic Justice League. "AJL". (2016). Retrieved from Website: https://www.ajl.org/
- Atari, M., Xue, M. J., Park, P. S., Blasi, D. E., & Henrich, J. (2023). *Which humans?* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/5b26t
- Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.
- Kantayya. S. (2020) Coded Bias. Retrieved from https://www.codedbias.com/
- Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. New York University Press.
- O'Neil, C., (2017). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Penguin Books.


Teşekkürler


Questions
are
guaranteed in
life;
Answers
aren't.


ΕΥΧΑΡΙΣΤΩ
TÄNAN
HVALA
GRACIAS
DZIĘKUJĘ
GRAZIE
ありがとう
MERCI
TACK
THANK YOU
DIAKUIU
PALDIES
ACIU
TAK
DANKE
DANK U WEL
ДЯКУЮ
СПАСИБО
谢谢
OBRIGADO
diolch
KIITOS
TESEKKUR EDERIM

Closing Survey

Closing Survey link:
https://forms.gle/xdB2H3a5tT7uYo58A


That's All
Folks!


The End


New Beginnings
But New Beginnings are Ahead