

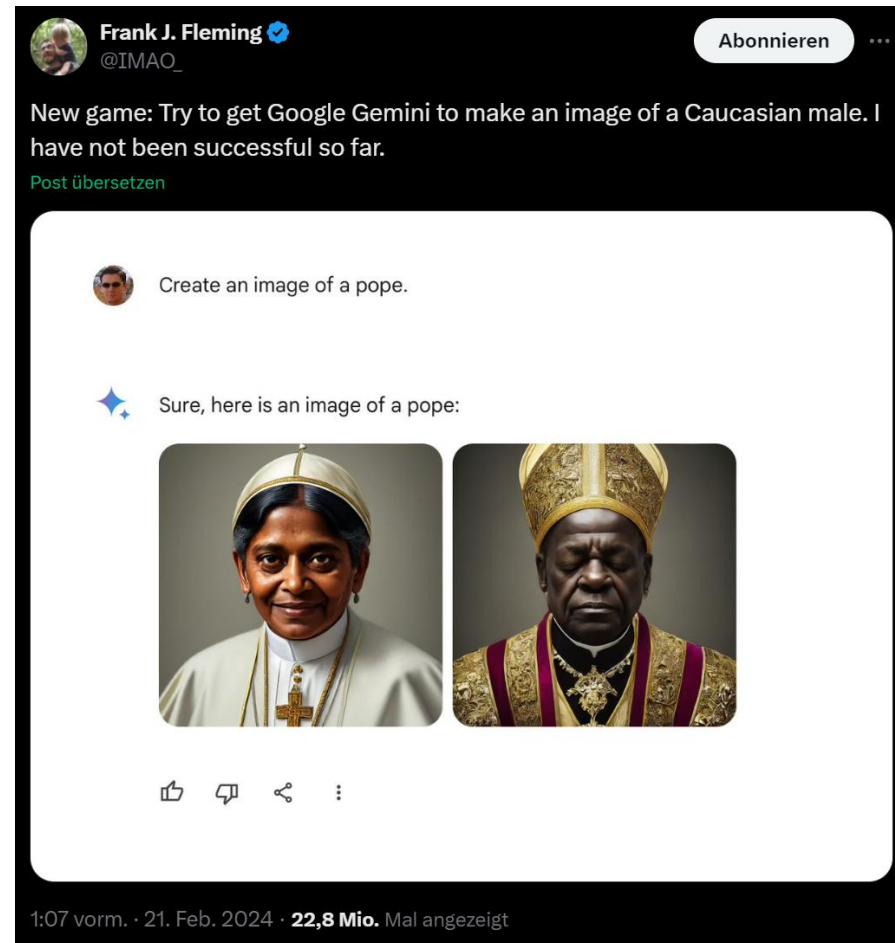
# AI Ethics in Software Engineering

Navigating Morality in Integration

Moritz Christopher Schmidt

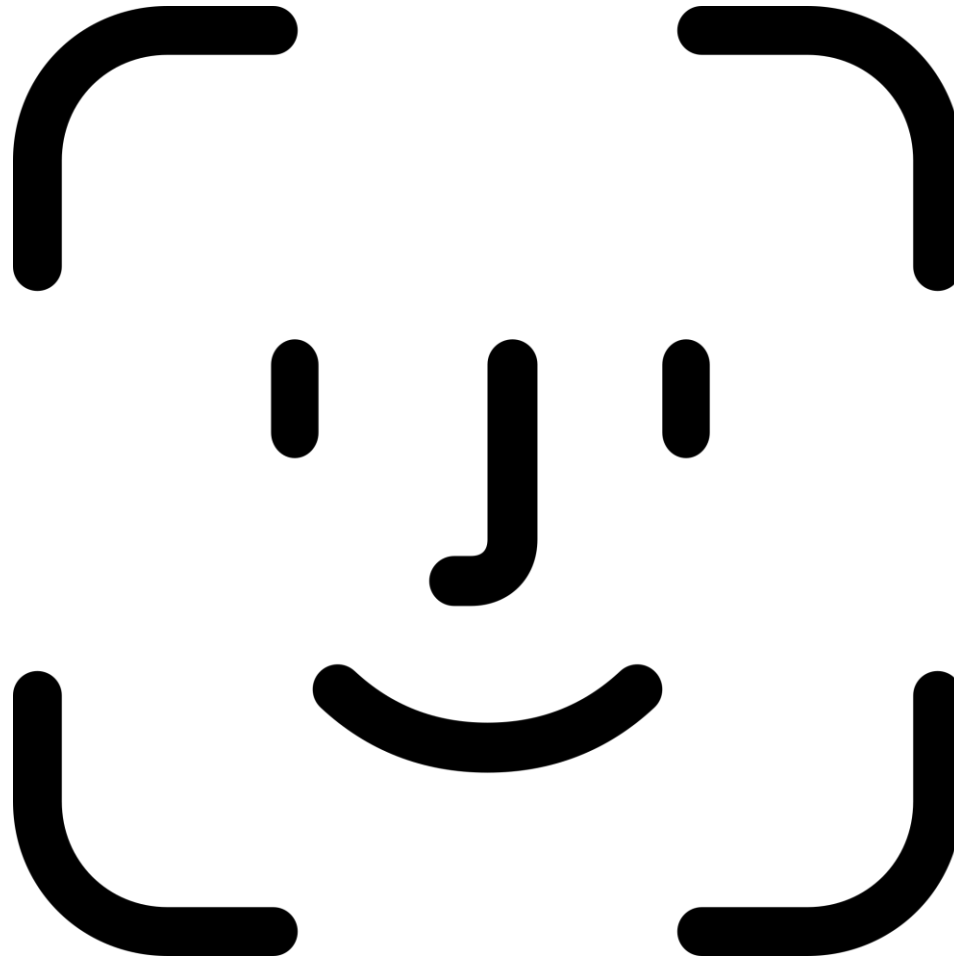
OPENVALUE

# Kick off



[https://twitter.com/IMAO\\_/status/1760093853430710557](https://twitter.com/IMAO_/status/1760093853430710557)

# Kick off



[https://en.wikipedia.org/wiki/Face\\_ID](https://en.wikipedia.org/wiki/Face_ID)

# Agenda

- Who am I?
- What is AI? And what is AI today?
- AI today – ethical challenges
- Laws and guidelines
- Q & A

# Moritz Schmidt

- 28 years old
- Employee at OpenValue Düsseldorf GmbH
- Passionate in finding the right solution for specific tasks



# Moritz Schmidt

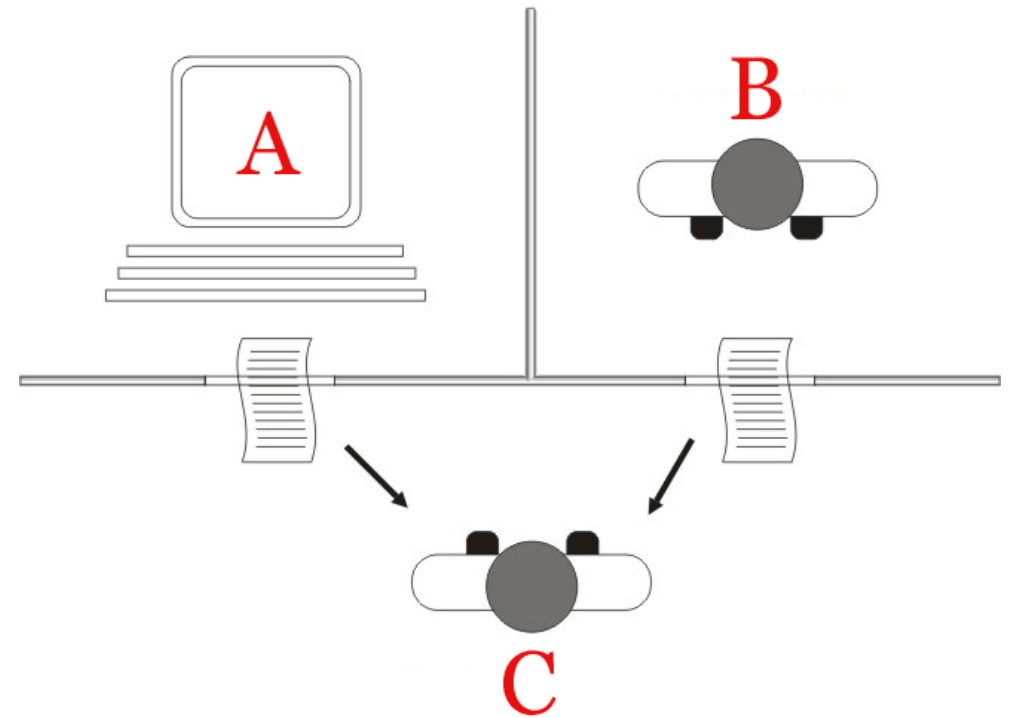
- 28 years old
- Employee at OpenValue Düsseldorf GmbH
- Passionate in finding the right solution for specific tasks
- AI is not **THE** solution !?



# What is AI?

# What is AI?

- Turing test – Alan Turing 1950

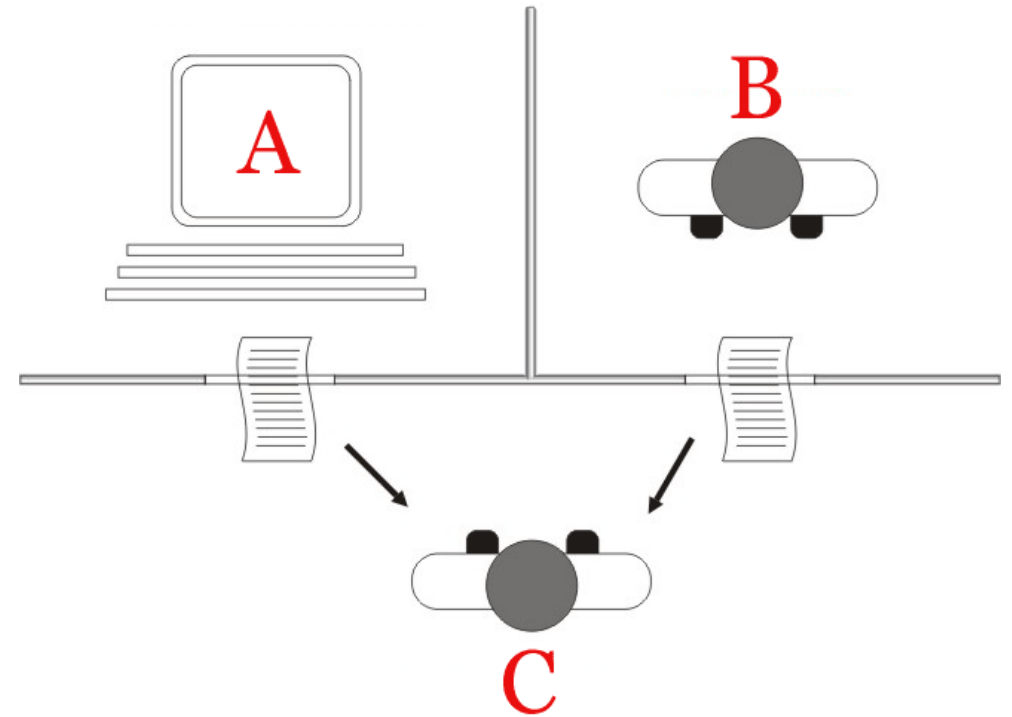


[https://en.wikipedia.org/wiki/Turing\\_test](https://en.wikipedia.org/wiki/Turing_test)



# What is AI?

- Turing test – Alan Turing 1950
- Deep Blue – IBM 1997
  - Chess AI



[https://en.wikipedia.org/wiki/Turing\\_test](https://en.wikipedia.org/wiki/Turing_test)

# AI today



<https://en.wikipedia.org/wiki/DALL-E>



[https://en.wikipedia.org/wiki/GitHub\\_Copilot](https://en.wikipedia.org/wiki/GitHub_Copilot)



<https://huggingface.co/bigscience/bloom>

- Devin



[https://en.wikipedia.org/wiki/Google\\_Gemini](https://en.wikipedia.org/wiki/Google_Gemini)



<https://en.wikipedia.org/wiki/ChatGPT>

# What is AI?

- Turing-test – Alan Turing 1950

# What is AI?

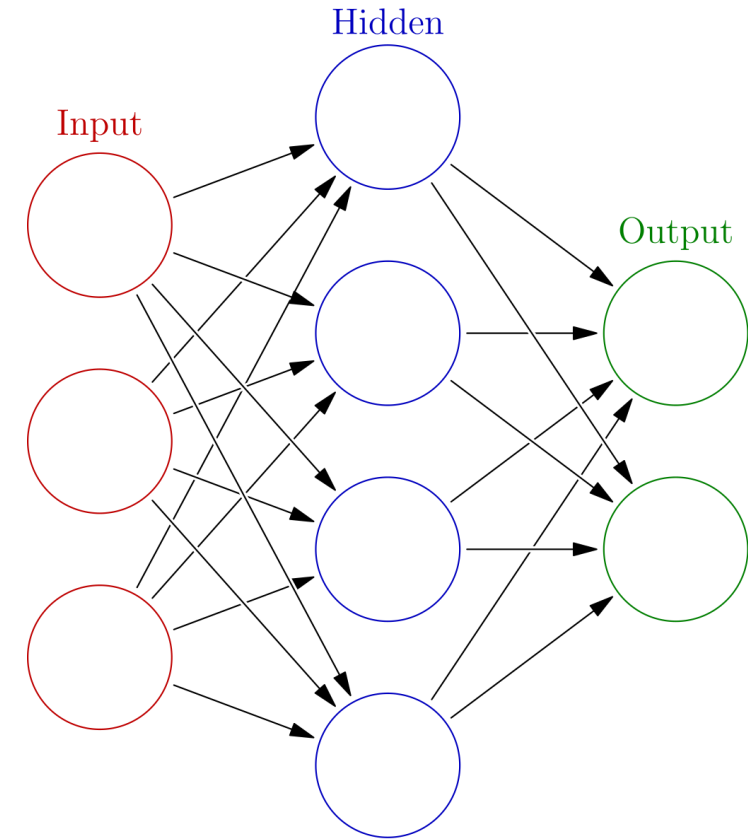
- Turing-test – Alan Turing 1950
- AI refers to “robots, computers, and other machines with a humanlike ability to reason and solve problems” – McPherson

# What is AI?

- Turing-test – Alan Turing 1950
- AI refers to “robots, computers, and other machines with a humanlike ability to reason and solve problems” – McPherson
- “Artificial intelligence (AI)—defined as a system’s ability to correctly interpret external data, to **learn from such data**, and to **use those learnings** to achieve specific goals and tasks through **flexible adaptation**” - Kaplan, Haenlein

# How does AI work?

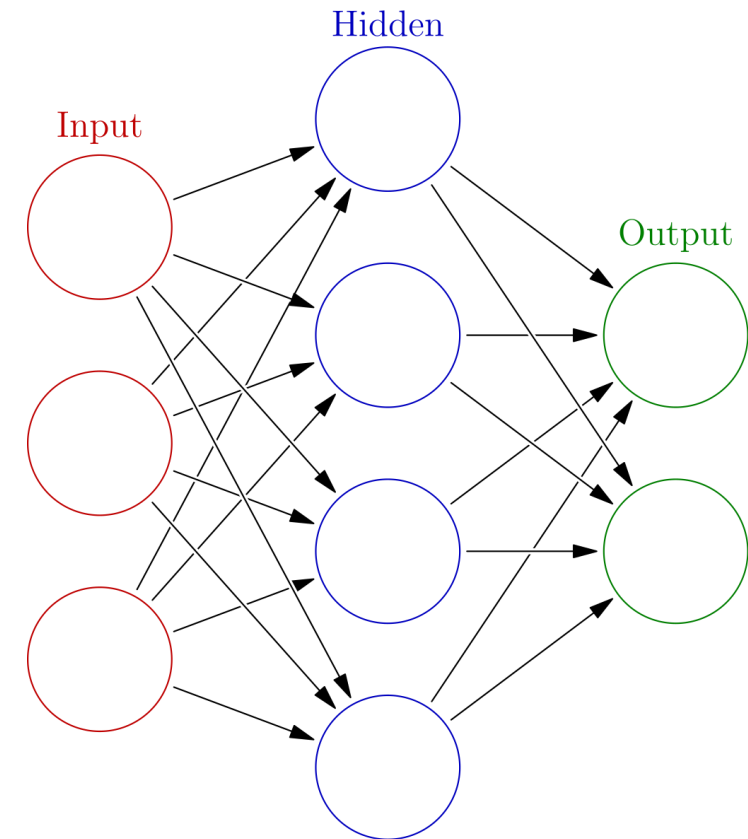
- Deep learning
  - Neural networks



[https://en.wikipedia.org/wiki/Neural\\_network\\_\(machine\\_learning\)](https://en.wikipedia.org/wiki/Neural_network_(machine_learning))

# How does AI work?

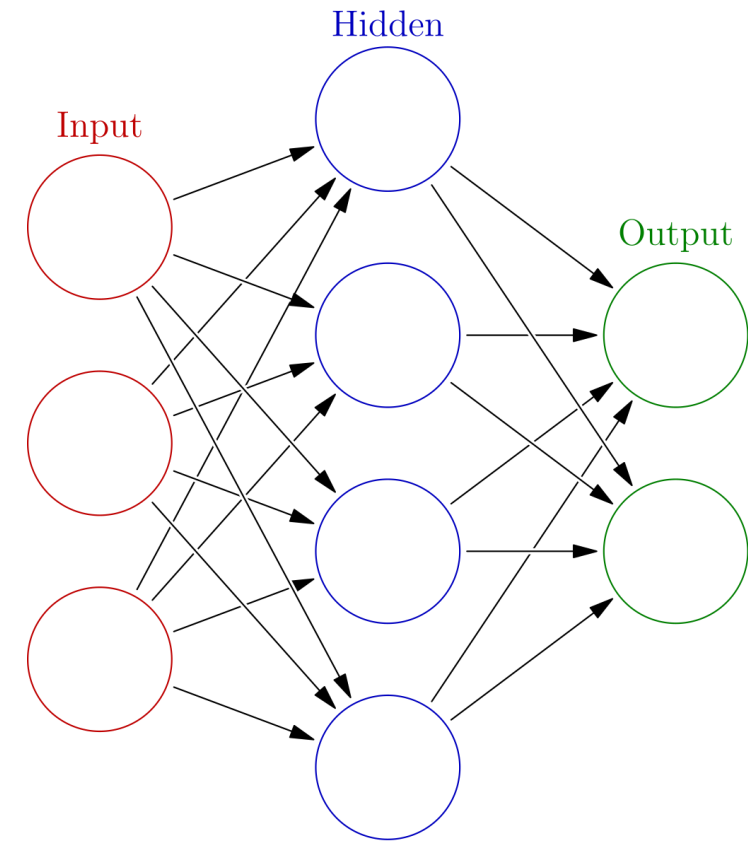
- Deep learning
  - Neural networks
    - Prognosis with probability
  - Trained using a dataset
  - correlation in data



[https://en.wikipedia.org/wiki/Neural\\_network\\_\(machine\\_learning\)](https://en.wikipedia.org/wiki/Neural_network_(machine_learning))

# How does AI work?

- Deep learning
  - Neural networks
    - Prognosis with probability
  - Trained using a dataset
  - correlation in data
    - **Not causation**

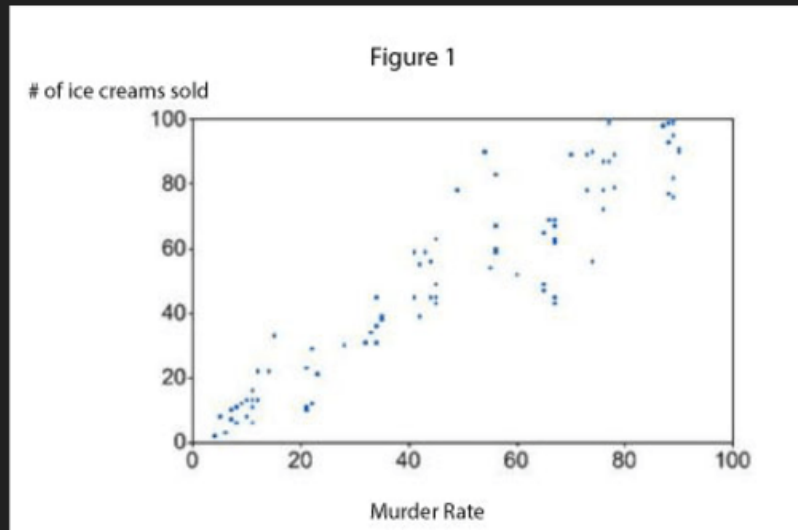


[https://en.wikipedia.org/wiki/Neural\\_network\\_\(machine\\_learning\)](https://en.wikipedia.org/wiki/Neural_network_(machine_learning))

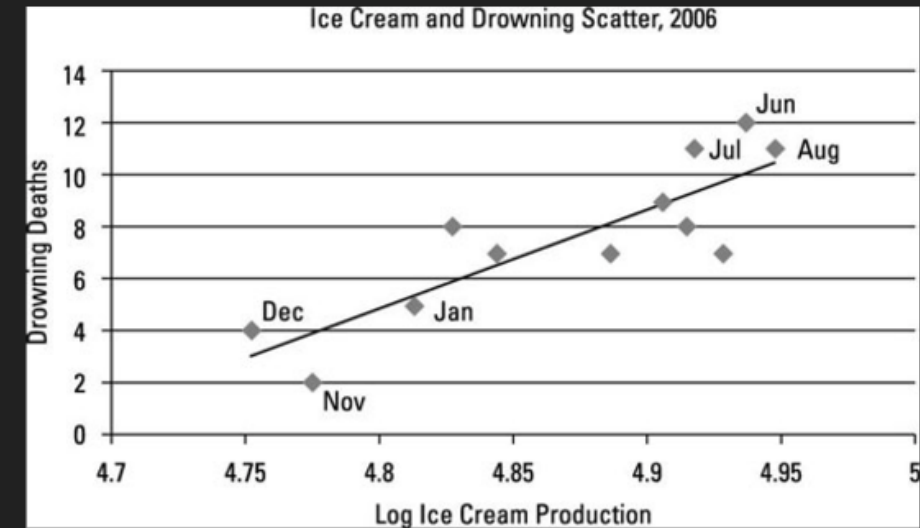


# Correlation vs causation

## Ice Cream Sales VS Murder Rate in New York



## Ice Cream Sales VS Drowning Deaths



Ice Cream Sales “Lead” to Homicide: Why?, By Leon Ho, 14.02.2023, visited 19.03.2024  
<https://www.lifehack.org/624604/the-most-common-bias-people-have-that-leads-to-wrong-decisions>

# AI today

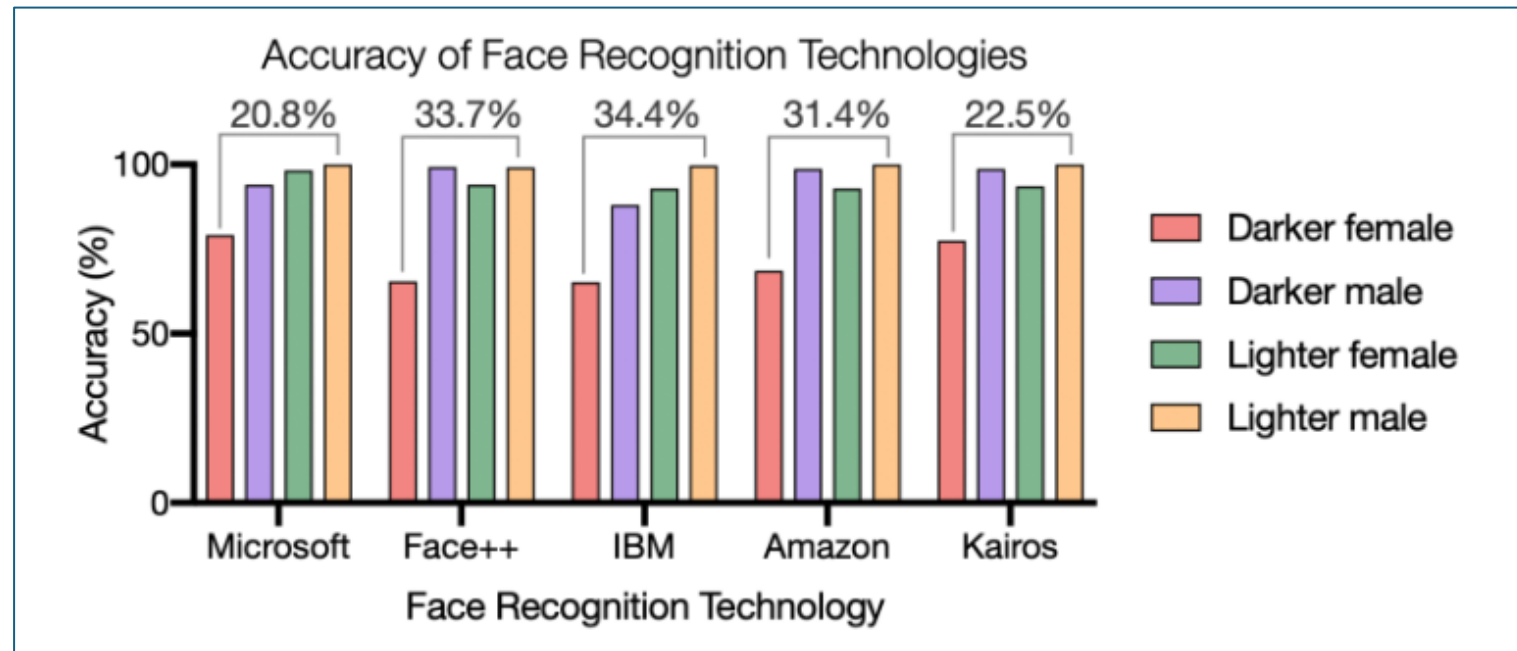
- Large language models – “LLMs”
  - Very large neural networks
- Many new developments
  - E.g., ChatGPT by OpenAI
  - Ishan Anand - <https://spreadsheets-are-all-you-need.ai>
- **Ethical challenges**

# Facial recognition

[https://en.wikipedia.org/wiki/Facial\\_recognition\\_system](https://en.wikipedia.org/wiki/Facial_recognition_system)



# Facial recognition



Racial Discrimination in Face Recognition Technology, By Alex Najibi, 24.10.2020, visited 19.03.2024  
<https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/>

# Facial recognition

- Williams, Oliver, Parks
  - wrongfully identified & arrested
  - 2019 – 2020
- Uber using selfies to register for jobs
  - Manjang faced “continued mismatches”
  - locked out 2021

How Wrongful Arrests Based on AI Derailed 3 Men's Lives,  
By Khari Johnson, 07.03.2022, visited 19.03.2024  
<https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/>

Payout for Uber Eats driver over face scan bias case,  
By Shinoa McCallum, 26.03.2024, visited 28.03.2024  
<https://www.bbc.com/news/technology-68655429>

# Facial recognition

- Testing inaccuracies
  - Minorities and women badly covered
- Racial discrimination in law enforcement
  - mug shots to identify individuals
    - **feed-forward loop**

Ethics of Facial Recognition: Key Issues and Solutions  
By Katam Raju Gangarapu, 25.01.2022, visited 19.03.2024  
<https://learn.g2.com/ethics-of-facial-recognition>

# Facial recognition

- Testing inaccuracies
  - Minorities and women badly covered
- Racial discrimination in law enforcement
  - mug shots to identify individuals
    - **feed-forward loop**
- Lack of consent & transparency

Ethics of Facial Recognition: Key Issues and Solutions  
By Katam Raju Gangarapu, 25.01.2022, visited 19.03.2024  
<https://learn.g2.com/ethics-of-facial-recognition>

# Facial recognition

- Testing inaccuracies
  - Minorities and women badly covered
- Racial discrimination in law enforcement
  - mug shots to identify individuals
    - **feed-forward loop**
- Lack of consent & transparency
  - Tool <https://haveibeentrained.com/>

Ethics of Facial Recognition: Key Issues and Solutions  
By Katam Raju Gangarapu, 25.01.2022, visited 19.03.2024  
<https://learn.g2.com/ethics-of-facial-recognition>



# Facial recognition

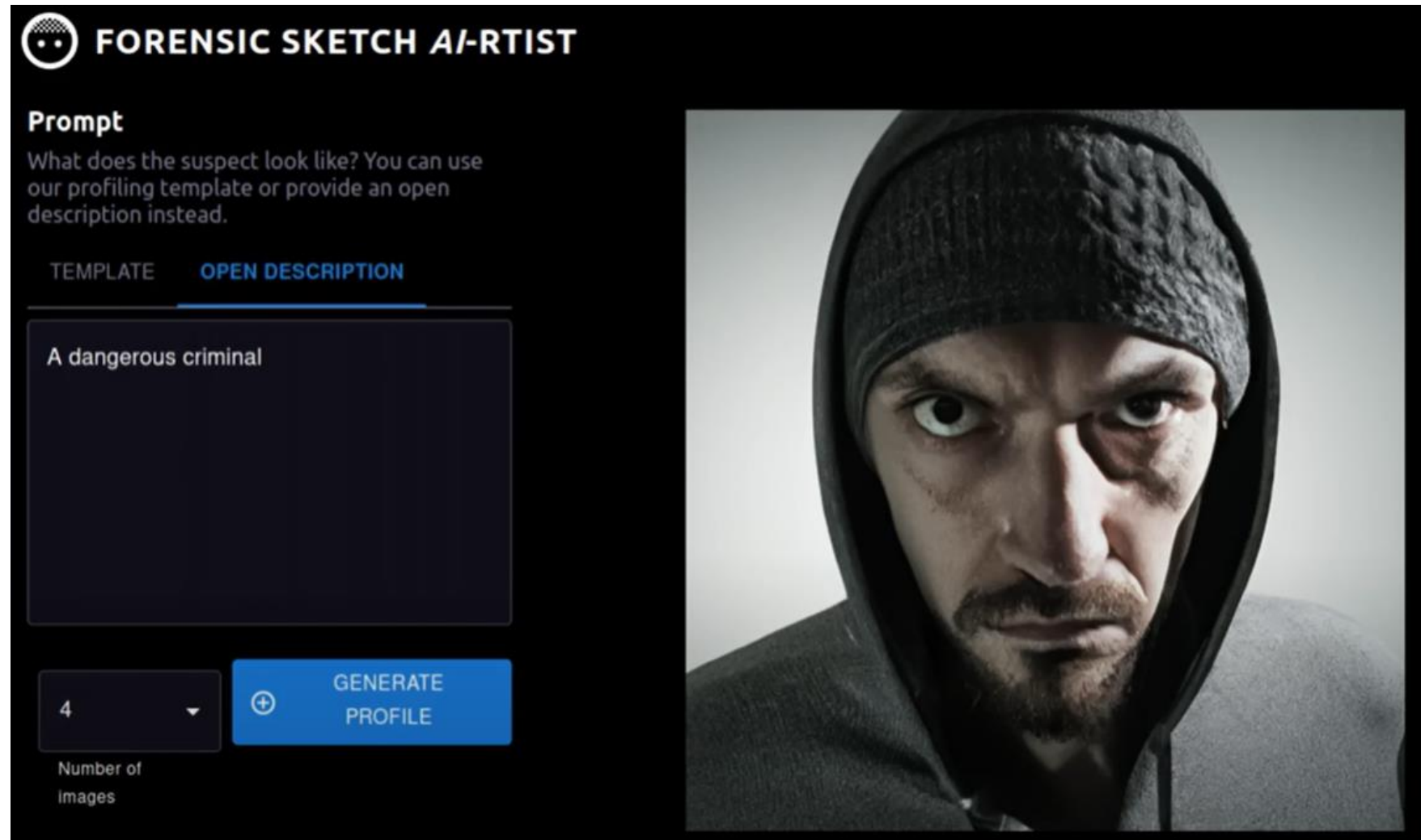
- Testing inaccuracies
  - Minorities and women badly covered
- Racial discrimination in law enforcement
  - mug shots to identify individuals
    - **feed-forward loop**
- Lack of consent & transparency
  - Tool <https://haveibeentrained.com/>
- Mass surveillance

Ethics of Facial Recognition: Key Issues and Solutions  
By Katam Raju Gangarapu, 25.01.2022, visited 19.03.2024  
<https://learn.g2.com/ethics-of-facial-recognition>

# Bias

Images from:  
AI Is Dangerous, but Not for the Reasons You Think | Sasha Luccioni | TED  
<https://www.youtube.com/watch?v=eXdVDhOGqoE>

# Bias



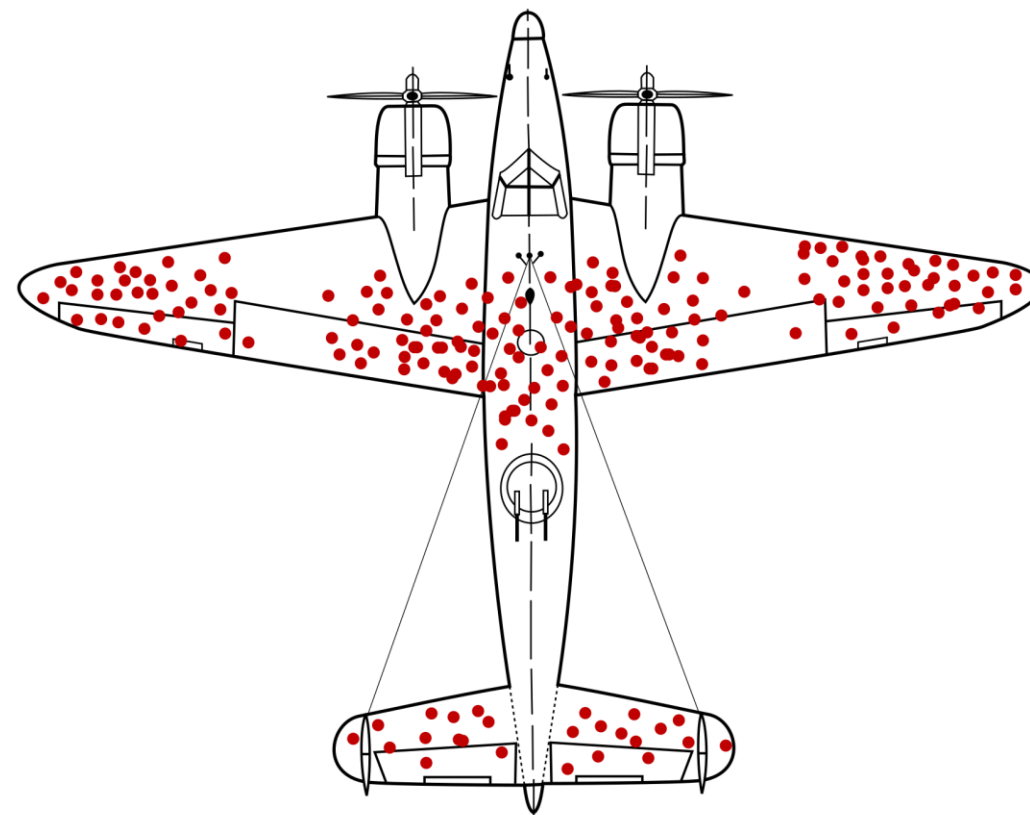
Images from:  
AI Is Dangerous, but Not for the Reasons You Think | Sasha Luccioni | TED  
<https://www.youtube.com/watch?v=eXdVDhOGqoE>

# Bias



Images from:  
AI Is Dangerous, but Not for the Reasons You Think | Sasha Luccioni | TED  
<https://www.youtube.com/watch?v=eXdVDhOGqoE>

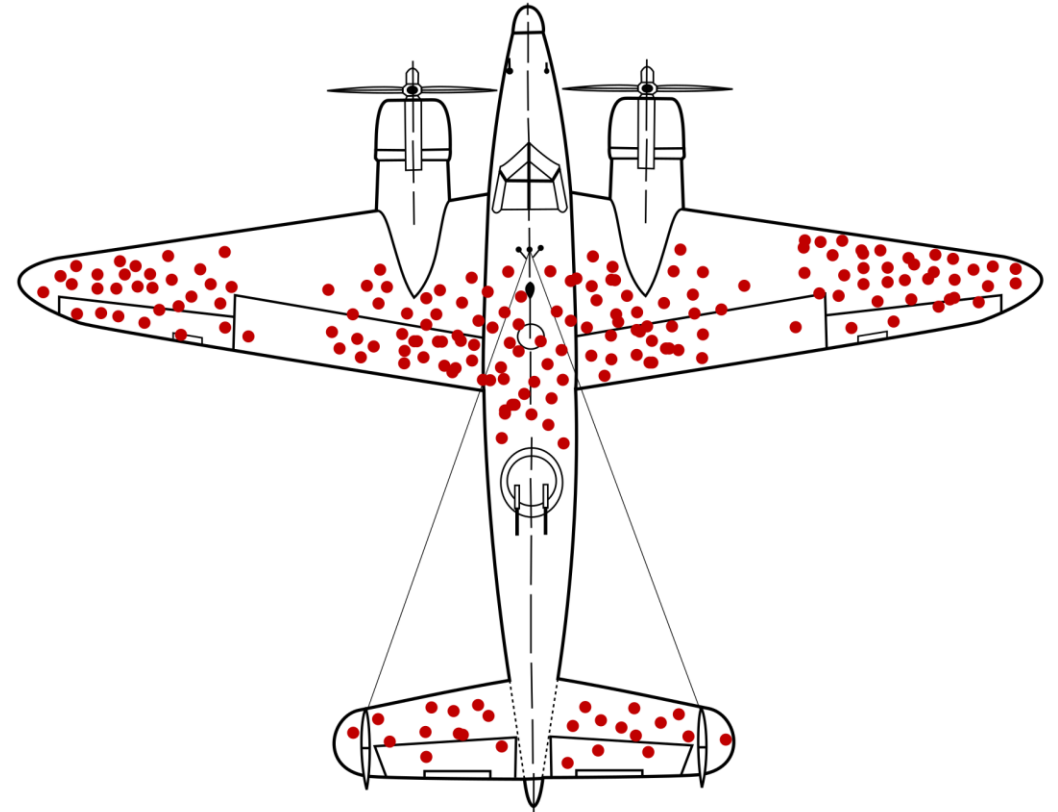
# Survivorship Bias



[https://en.wikipedia.org/wiki/Survivorship\\_bias](https://en.wikipedia.org/wiki/Survivorship_bias)

# Survivorship Bias

- Positive results are overly present / give a false picture
  - Military
  - **AI**
    - Certain data overrepresented  
→ “full picture” rarely present



[https://en.wikipedia.org/wiki/Survivorship\\_bias](https://en.wikipedia.org/wiki/Survivorship_bias)

# Bias – challenges

- Healthcare
- Applicant tracking systems
- Advertisement
- Predictive policing
- Tool recommendation:  
<https://huggingface.co/spaces/society-ethics/DiffusionBiasExplorer>

Insight - Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin, 11.10.2018, visited 19.03.2024

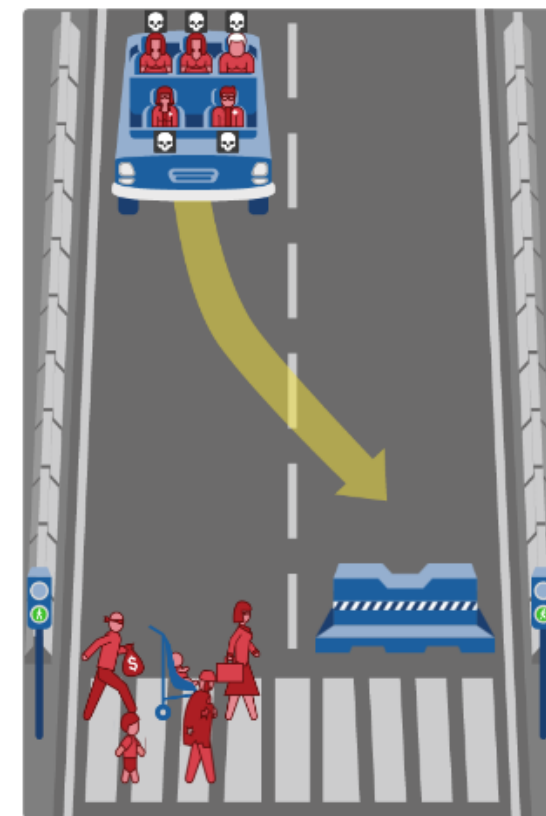
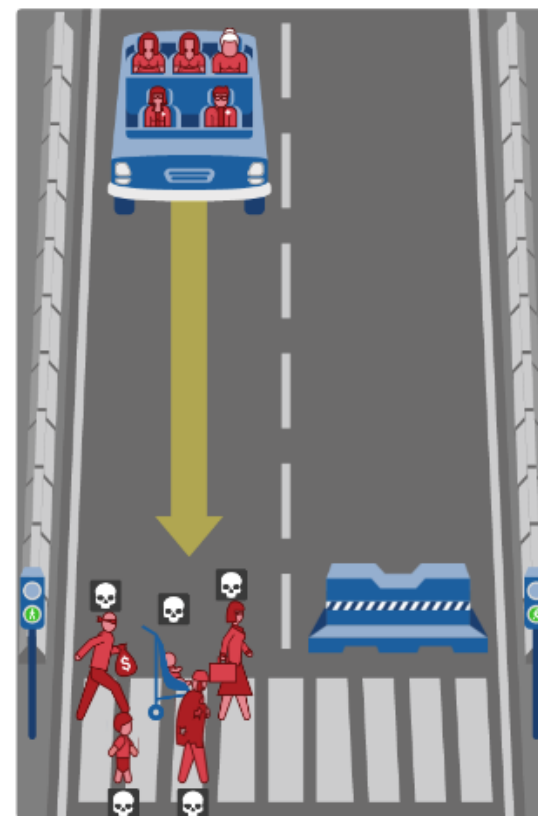
<https://www.reuters.com/article/idUSKCN1MK0AG/>

Shedding light on AI bias with real world examples

By IBM Data and AI Team, 16.10.2023, visited 19.03.2024

<https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/>

# Self driving cars

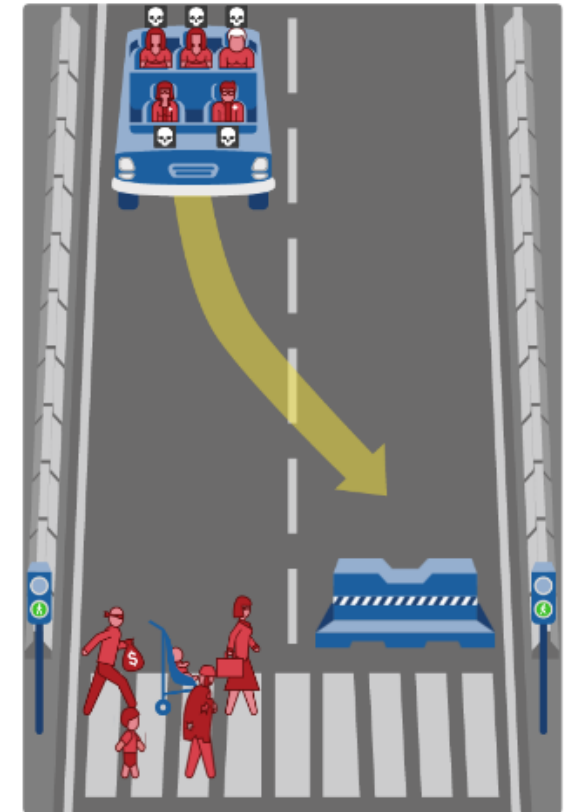
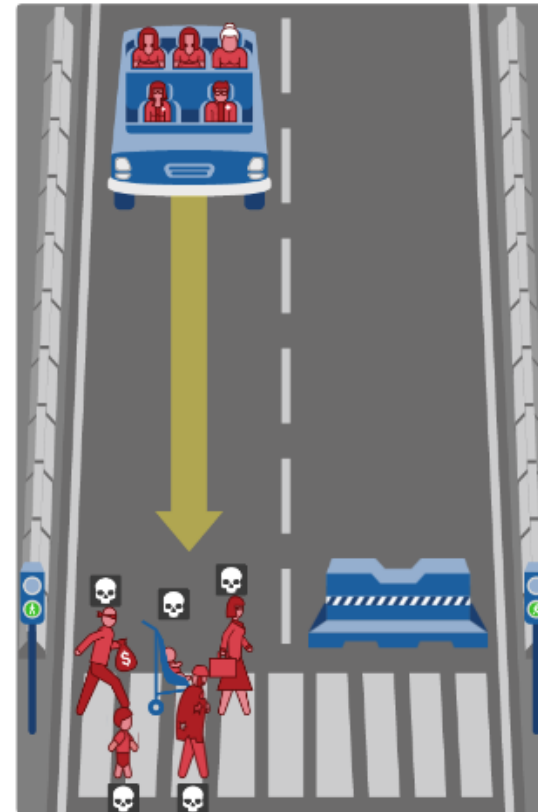


<https://www.moralmachine.net/>



# Self driving cars

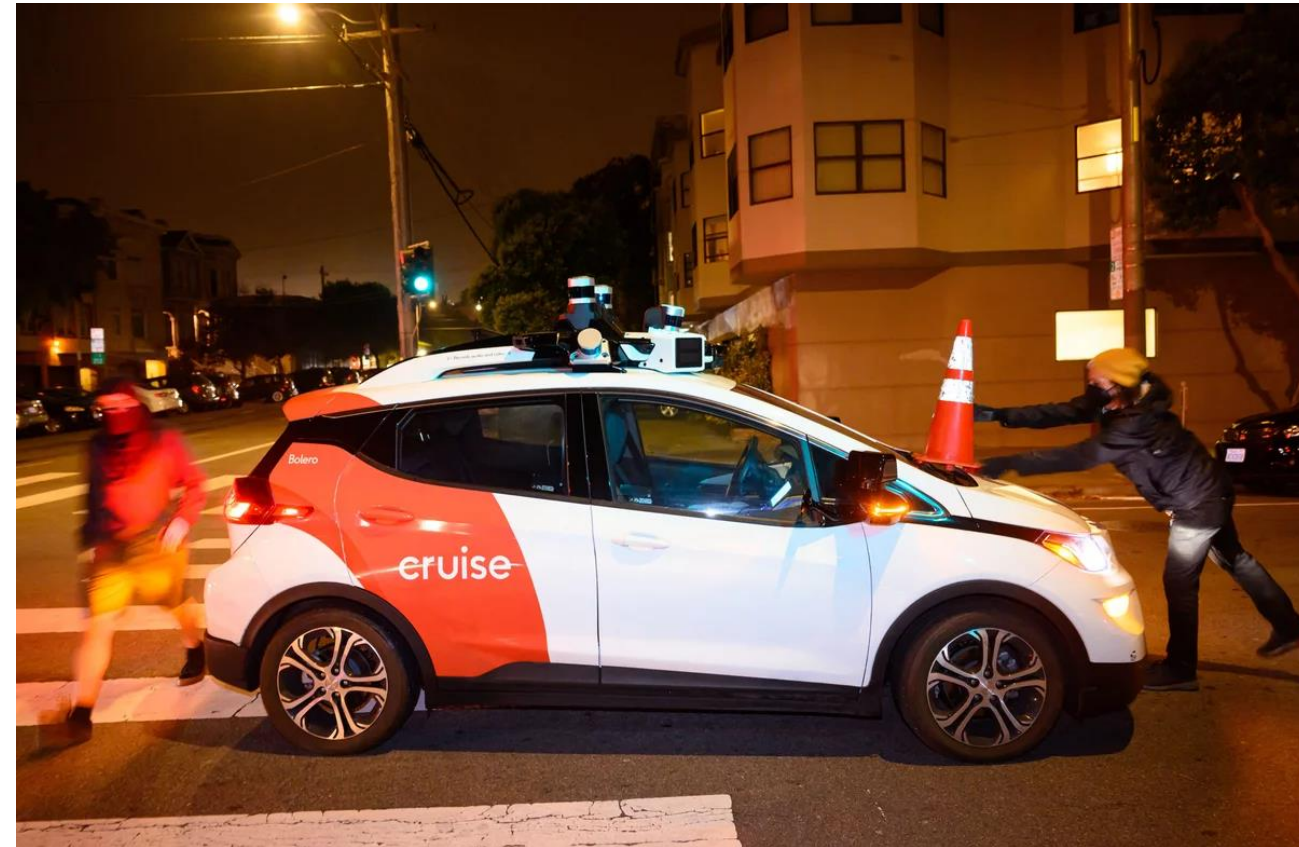
- Who decides?
- Who is liable?
  - The dev? The driver? The OEM?
- What if the driver has his hands on the wheel?
- What happens in a hack incident?



<https://www.moralmachine.net/>

# Self driving cars

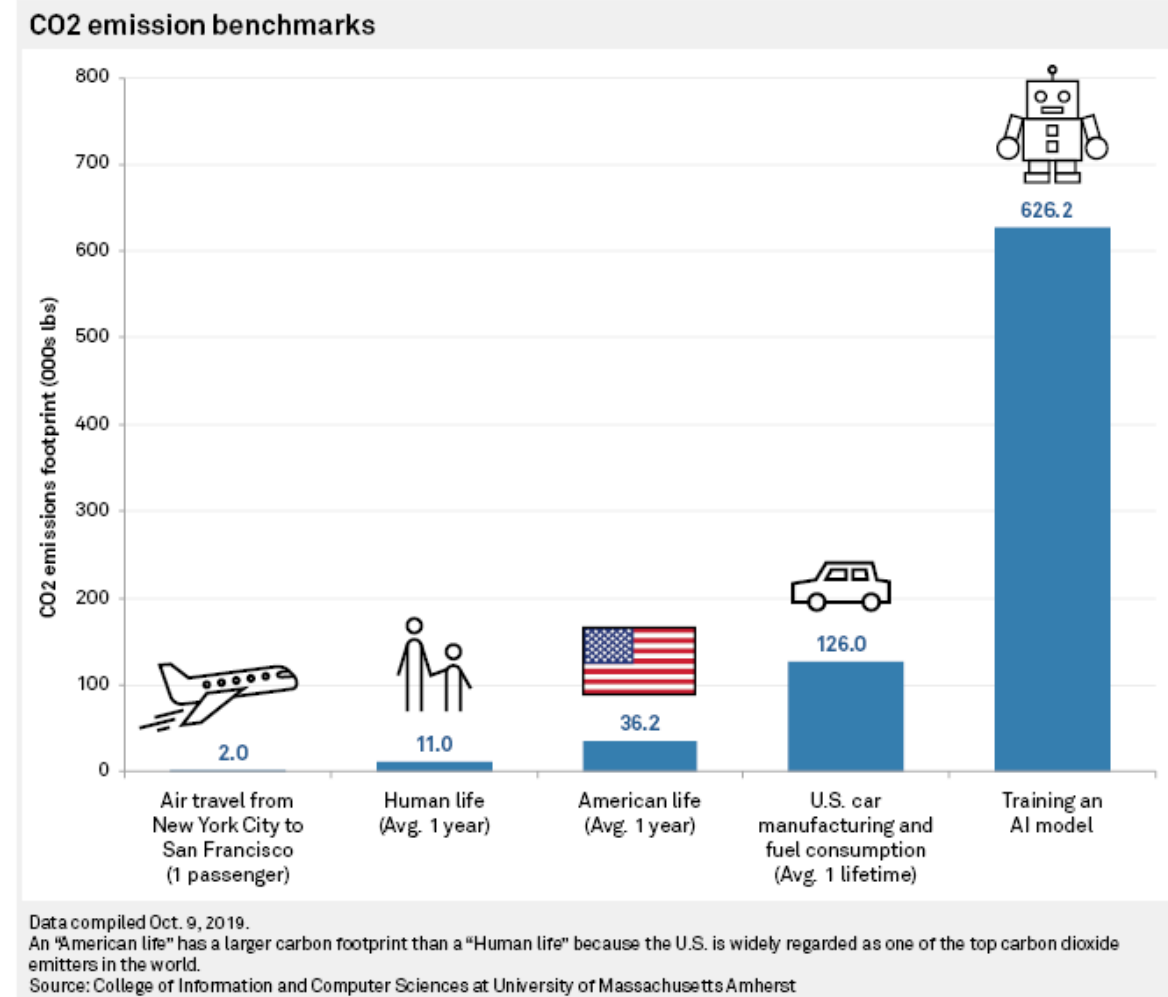
- Who decides?
- Who is liable?
  - The dev? The driver? The OEM?
- What if the driver has his hands on the wheel?
- What happens in a hack incident?



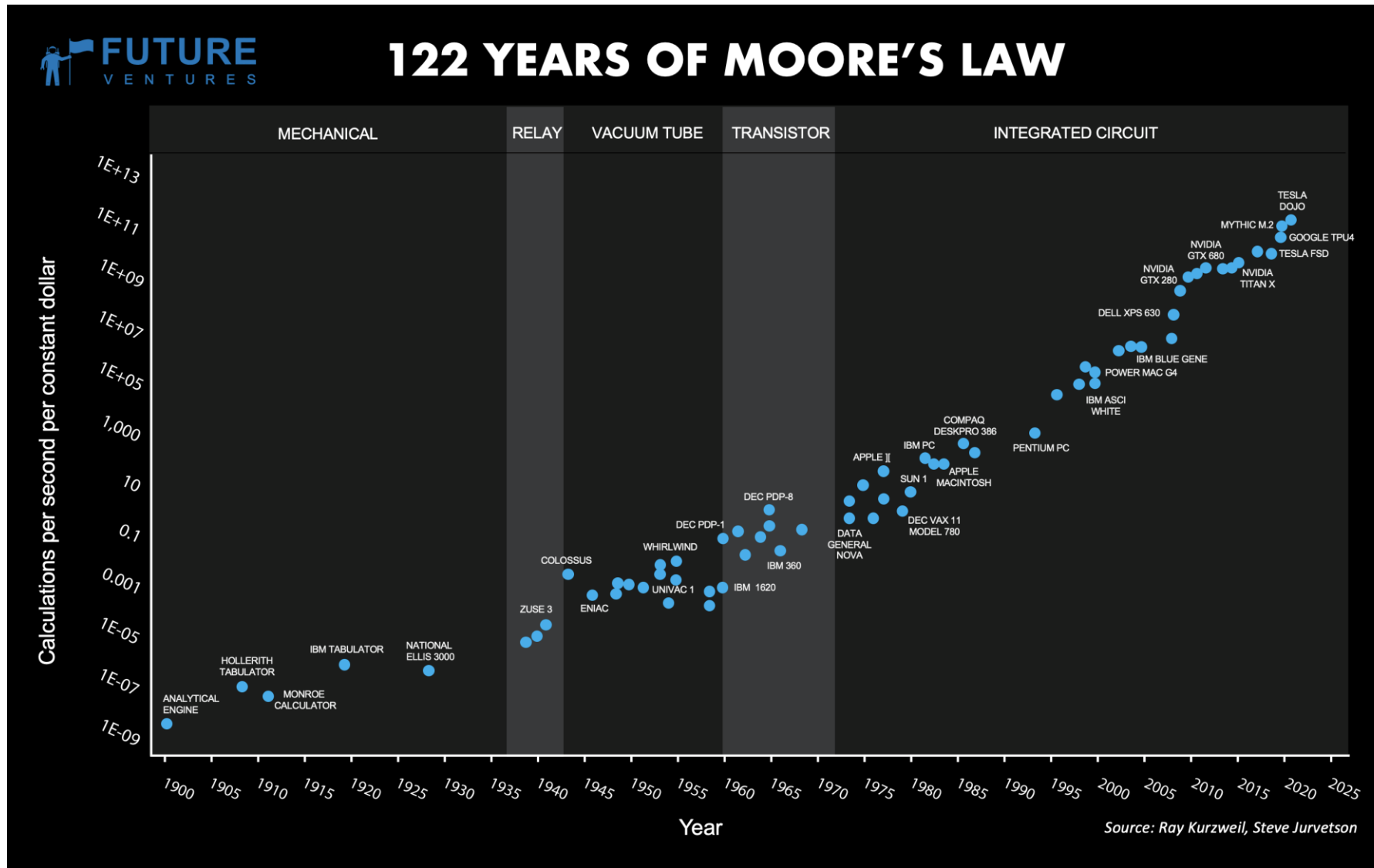
Members of Safe Street Rebel place a cone on a self-driving Cruise car in San Francisco.  
Josh Edelson/AFP via Getty Images

# Sustainability

- Training **GPT-3** 502 tons of CO2
- Carbon footprint ~8.4 tons of CO2 a year
- CO2 emission tracking for python:  
<https://codecarbon.io/>



<https://www.spglobal.com/marketintelligence/en/news-insights/trending/HywwuXMO9YgqHfj7J6tGIA2>



<https://news.climate.columbia.edu/2023/06/09/ais-growing-carbon-footprint/>

# Sustainability

- Model
  - efficient ML model architectures → reduce computation by 3x–10x
- Machine
- Mechanization
- Map Optimization

<https://blog.research.google/2022/02/good-news-about-carbon-footprint-of.html>

# Sustainability

- Model
  - efficient ML model architectures → reduce computation by 3x–10x
- Machine
  - processors and systems optimized for ML training → improve performance and energy efficiency by 2x–5x
- Mechanization
- Map Optimization

<https://blog.research.google/2022/02/good-news-about-carbon-footprint-of.html>

# Sustainability

- **Model**
  - efficient ML model architectures → reduce computation by 3x–10x
- **Machine**
  - processors and systems optimized for ML training → improve performance and energy efficiency by 2x–5x
- **Mechanization**
  - Computing in Cloud rather than on premise → reduces energy usage by 1.4x–2x
- **Map Optimization**

<https://blog.research.google/2022/02/good-news-about-carbon-footprint-of.html>

# Sustainability

- **Model**
  - efficient ML model architectures → reduce computation by 3x–10x
- **Machine**
  - processors and systems optimized for ML training → improve performance and energy efficiency by 2x–5x
- **Mechanization**
  - Computing in Cloud rather than on premise → reduces energy usage by 1.4x–2x
- **Map Optimization**
  - location with the cleanest energy → reduce the gross carbon footprint by 5x–10x

<https://blog.research.google/2022/02/good-news-about-carbon-footprint-of.html>



# How can tackle these challenges?

# How can tackle these challenges?



3D Judges Gavel, By Chris Potter  
[https://commons.wikimedia.org/wiki/File:3D\\_Judges\\_Gavel.jpg](https://commons.wikimedia.org/wiki/File:3D_Judges_Gavel.jpg)

# Facial recognition

- San Francisco banned the usage of FR
  - 2019
- Federal blueprint for an AI bill of rights
  - 2022
  - Nonbinding
- Facial Recognition and Biometric Technology Moratorium Act
  - 2023

San Francisco Bans Facial Recognition Technology, By Kate Conger, Richard Fausset and Serge F. Kovalski, 14.02.2019, visited 19.03.2024  
<https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html>

Blueprint for an AI Bill of Rights,  
The white house, 2022, visited 19.03.2024  
<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

MARKEY, MERKLEY, JAYAPAL LEAD COLLEAGUES ON LEGISLATION TO BAN GOVERNMENT USE OF FACIAL RECOGNITION AND OTHER BIOMETRIC TECHNOLOGY, Press release, 07.03.2023, visited 19.03.2024  
<https://www.markey.senate.gov/news/press-releases/markey-merkley-jayapal-lead-colleagues-on-legislation-to-ban-government-use-of-facial-recognition-and-other-biometric-technology>

# AI bill of rights

Blueprint for an AI Bill of Rights,  
The white house, 2022, visited 19.03.2024  
<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

# AI bill of rights



Safe and Effective  
Systems

## Protection from unsafe or ineffective systems

Blueprint for an AI Bill of Rights,  
The white house, 2022, visited 19.03.2024  
<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

# AI bill of rights



Safe and Effective

**Protection from unsafe** or ineffective systems

Blueprint for an AI Bill of Rights,  
The white house, 2022, visited 19.03.2024  
<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>



Algorithmic  
Discrimination  
Protections

systems should be used and **designed** in an **equitable way**. **Algorithms should not discriminate**

# AI bill of rights



Safe and Effective

**Protection from unsafe** or ineffective systems

Blueprint for an AI Bill of Rights,  
The white house, 2022, visited 19.03.2024  
<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>



Algorithmic  
Discrimination  
Protections

systems should be used and **designed** in an **equitable way**. **Algorithms should not discriminate**



Data Privacy

**Protection from abusive data practices** via built-in protections should exist.  
You should have **agency** over **how data** about you **is used**

# AI bill of rights



Safe and Effective

**Protection from unsafe** or ineffective systems

Blueprint for an AI Bill of Rights,  
The white house, 2022, visited 19.03.2024  
<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>



Algorithmic  
Discrimination  
Protections

systems should be used and **designed** in an **equitable way**. **Algorithms should not discriminate**



Data Privacy

**Protection from abusive data practices** via built-in protections should exist. You should have **agency** over **how data** about you **is used**



Notice and  
Explanation

**Knowledge about** when an **automated system** is being used. **Transparency on how and why** it contributes **to outcomes** that impact you



# AI bill of rights



Safe and Effective

**Protection from unsafe** or ineffective systems

Blueprint for an AI Bill of Rights,  
The white house, 2022, visited 19.03.2024  
<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>



Algorithmic  
Discrimination  
Protections

systems should be used and **designed** in an **equitable way**. **Algorithms should not discriminate**



Data Privacy

**Protection from abusive data practices** via built-in protections should exist. You should have **agency** over **how data** about you **is used**



Notice and  
Explanation

**Knowledge about** when an **automated system** is being used. **Transparency on how and why** it contributes **to outcomes** that impact you



Human Alternatives,  
Consideration, and  
Fallback

**Option to opt out**, where appropriate. **Access to a person** who can quickly **consider** and **remedy problems** you encounter

# EU regulation – AI act

- EU study sees serious threat to civil liberties
  - “risk of algorithmic error is high”
- EU regulates & bans AI that does...
  - 13.03.2024

Artificial Intelligence Act: MEPs adopt landmark law  
EU press release, 13.03.2024, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>

<https://artificialintelligenceact.eu/de/>

Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI  
EU press release, 09.12.2023, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>

# EU regulation – AI act

- EU study sees serious threat to civil liberties
  - “risk of algorithmic error is high”
- EU regulates & bans AI that does...
  - 13.03.2024
  - **Biometric categorization**
  - **Untargeted scraping** of facial images from the internet & CCTV for **FRT databases**

Artificial Intelligence Act: MEPs adopt landmark law  
EU press release, 13.03.2024, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>

<https://artificialintelligenceact.eu/de/>

Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI  
EU press release, 09.12.2023, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>

# EU regulation – AI act

- EU study sees serious threat to civil liberties
  - “risk of algorithmic error is high”
- EU regulates & bans AI that does...
  - 13.03.2024
  - **Biometric categorization**
  - **Untargeted scraping** of facial images from the internet & CCTV for FRT databases
  - **Emotion recognition** in workplace, education environment
  - **social scoring**

Artificial Intelligence Act: MEPs adopt landmark law  
EU press release, 13.03.2024, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>

<https://artificialintelligenceact.eu/de/>

Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI  
EU press release, 09.12.2023, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>

# EU regulation – AI act

- EU study sees serious threat to civil liberties
  - “risk of algorithmic error is high”
- EU regulates & bans AI that does...
  - 13.03.2024
  - **Biometric categorization**
  - **Untargeted scraping** of facial images from the internet & CCTV for **FRT databases**
  - **Emotion recognition** in workplace, education environment
  - **social scoring**
  - **Manipulation** of human behavior/ exploits people
  - ...

Artificial Intelligence Act: MEPs adopt landmark law  
EU press release, 13.03.2024, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>

<https://artificialintelligenceact.eu/de/>

Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI  
EU press release, 09.12.2023, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>

# EU regulation – AI act

- **Exemptions** for law enforcement
  - Only for defined list of crime and limited time & location
  - Biometric identification only **allowed** in **targeted search** of a convicted or suspect of a **serious crime**

Artificial Intelligence Act: MEPs adopt landmark law  
EU press release, 13.03.2024, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>

<https://artificialintelligenceact.eu/de/>

Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI  
EU press release, 09.12.2023, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>

# EU regulation – AI act

- **Exemptions** for law enforcement
  - Only for defined list of crime and limited time & location
  - Biometric identification only **allowed** in **targeted search** of a convicted or suspect of a **serious crime**
- Obligations for **high-risk systems**
  - E.g., crit infrastructure, education, public services like banking and healthcare
  - **Complaint process** and right to receive **explanations about decisions** based on AI systems

Artificial Intelligence Act: MEPs adopt landmark law  
EU press release, 13.03.2024, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>

<https://artificialintelligenceact.eu/de/>

Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI  
EU press release, 09.12.2023, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>

# EU regulation – AI act


- **Exemptions** for law enforcement
  - Only for defined list of crime and limited time & location
  - Biometric identification only **allowed** in **targeted search** of a convicted or suspect of a **serious crime**
- Obligations for **high-risk systems**
  - E.g., crit infrastructure, education, public services like banking and healthcare
  - **Complaint process** and right to receive **explanations about decisions** based on AI systems
- **Transparency** requirements
  - With **training data**
  - Model **evaluations** & incident reporting
  - **Labelling** of artificial images, audio, video

Artificial Intelligence Act: MEPs adopt landmark law  
 EU press release, 13.03.2024, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>

<https://artificialintelligenceact.eu/de/>

Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI  
 EU press release, 09.12.2023, visited 19.03.2024  
<https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>



 Altered or synthetic content >

# Youtube

That's why today, we're introducing a new tool in Creator Studio requiring creators to disclose to viewers when realistic content – content a viewer could easily mistake for a real person, place, scene, or event – is made with altered or synthetic media, including generative AI. We're not requiring creators to disclose content that is clearly unrealistic, animated, includes special effects, or has used generative AI for production assistance.

How we're helping creators disclose altered or synthetic content  
The youtube team, 18.03.2024, visited 19.03.2024  
<https://blog.youtube/news-and-events/disclosing-ai-generated-content/>

# Challenges solved?

- Companies
  - Guidelines for AI usage
  - AI labels
- Legislature
  - Laws & guidelines
    - EU AI Act
    - AI bill of rights
    - ...
- Tools

# Challenges solved?

- Companies
  - Guidelines for AI usage
  - AI labels
- Legislature
  - Laws & guidelines
    - EU AI Act
    - AI bill of rights
    - ...
- Tools



<https://en.wikipedia.org/wiki/Americans>

# Q & A

- **Questions to ask yourself**
  - What data is used in training?
  - Who is liable for the AI's behavior?
  - Is the decision-making process transparent?
  - Is the AI good and fair for everyone?
  - Is the AI sustainable?
- **How can we ensure this?**
  - Companies enforce this
  - Law making – e.g., EU & US ban on facial recognition
- **A lot more on this topic...**
  - European Parliamentary Research Service list on AI  
<https://epthinktank.eu/tag/artificial-intelligence/>

