Wednesday October 18th@6:00PM ET Livestream on YouTube

Digital Self Defense:

The AI Edition

Abigail Dubiniecki

OWASP OTTAWA

uOttawa

Join us in person at:

150 Louis-Pasteur Private, University of Ottawa

STEM Building, Rm 464

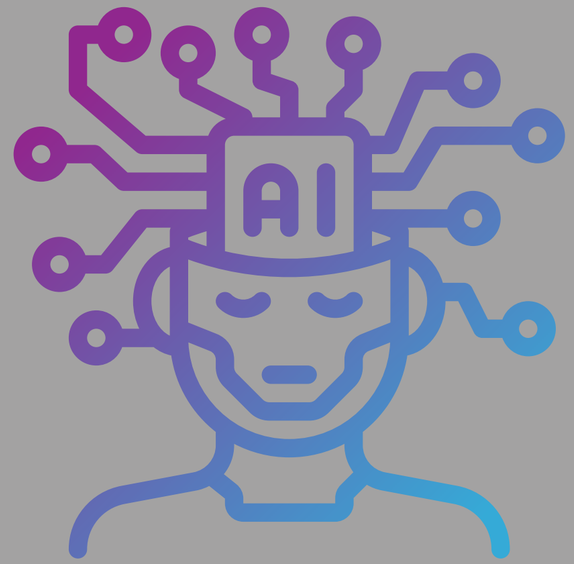
INTRO

The release of ChatGPT-4 brought a heady mixture of doomsday & euphoria as predictions that AI would soon overtake humanity, or at least our jobs, dominated the headlines.

Then people started playing with it. While it sometimes impressed, it often also fefll laughably short of its promises.

Overblown claims of AI capability can distract us from the real-world harms caused by poorly designed AI deployed today that either doesn't work as advertised or that was designed without regard to the harms they can cause. It also hides the successful AI use cases and the potential for real benefit when used responsibly.

Privacy regulators have led the way in reigning in harmful AI in the absence of dedicated AI regulation. Can AI Privacy save us from ourselves?



OVERVIEW

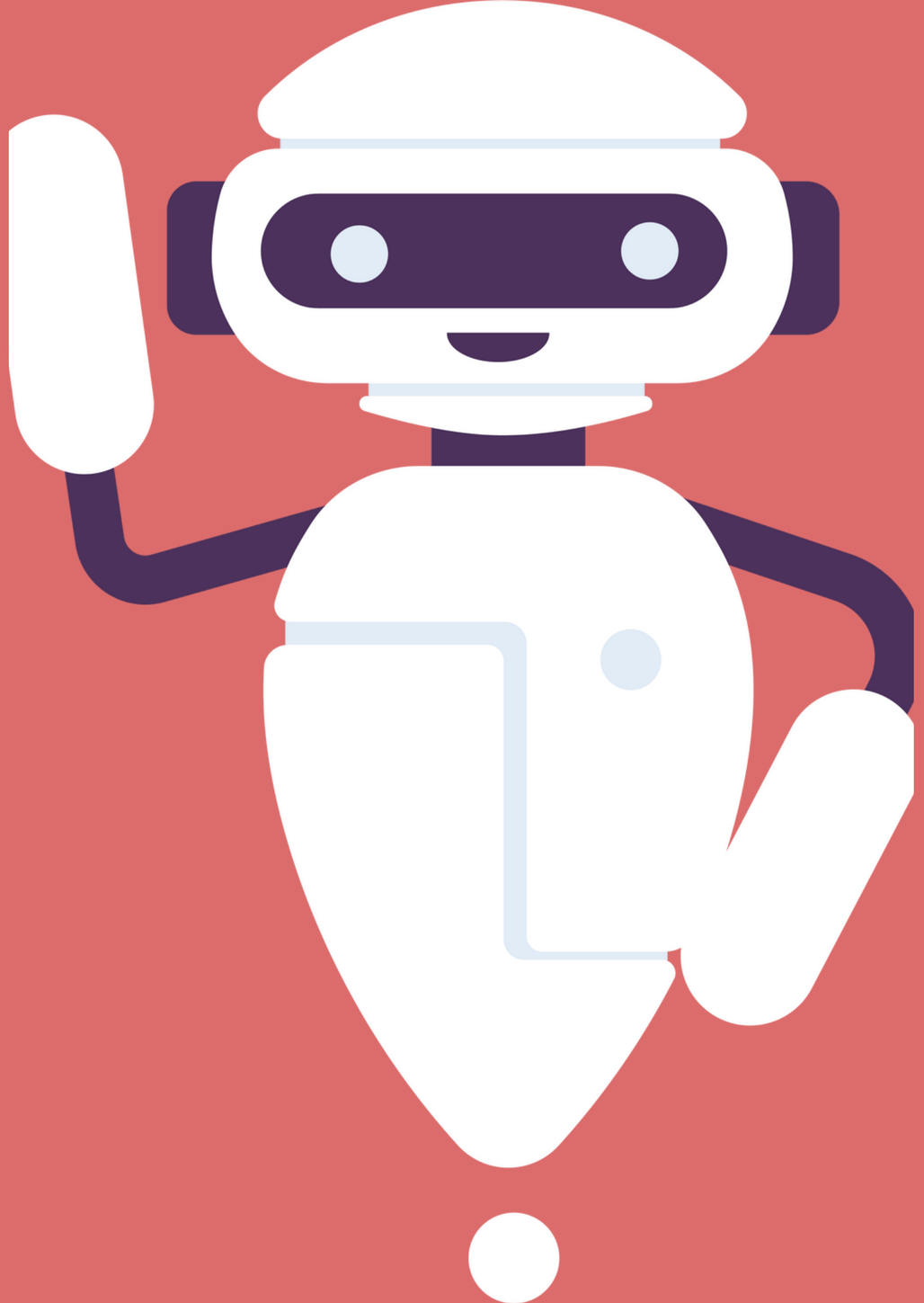
TOPICS

01 AI Myths Dispelled

02 AI Benefits and Harms

03 What's Privacy Gotta To With It?

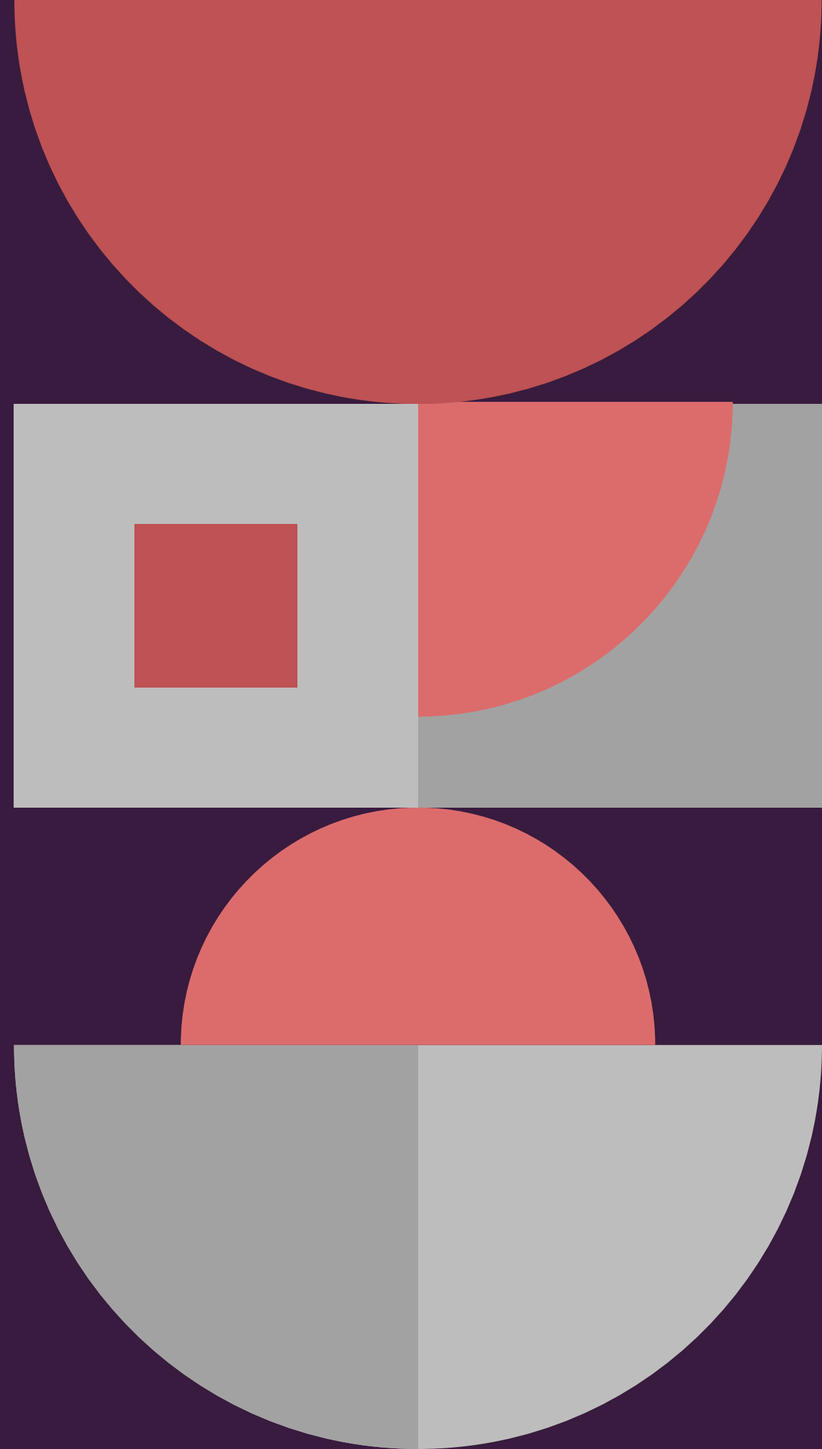
04 Leveraging Privacy to Protect Yourself - and Society - from AI Harms



AI Myths Dispelled

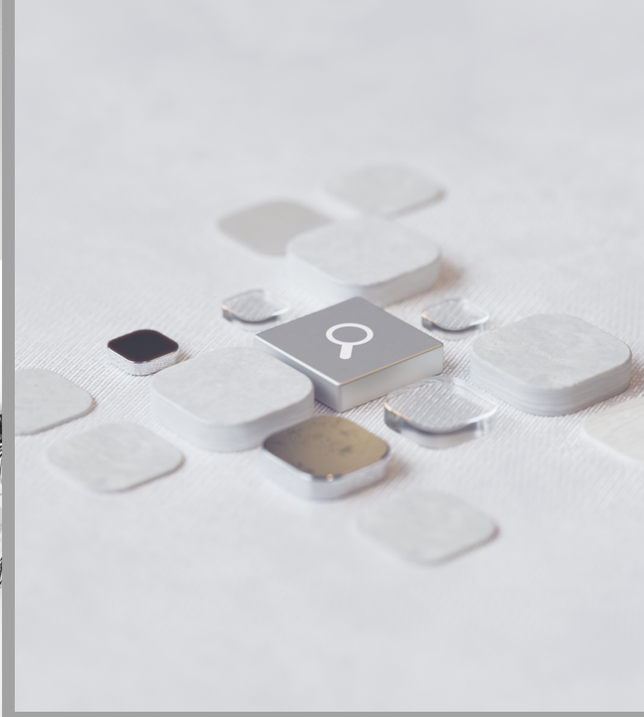
AI is just human stupidity, amplified, and at speed

— Rowenna Fielding, Data protection, data ethics and privacy nerd



AI 101

What is AI?



What is AI?

Whatever you want it to be? Whatever is shiny and new?

CANADA'S DRAFT AIDA

A technological system that, autonomously or partly autonomously, processes data related to human activities through the use of a genetic algorithm, a neural network, machine learning or another technique in order to generate content or make decisions, recommendations or predictions.

This is draft legislation and even the definition of AI is giving legislators heartburn.

EU DRAFT AI ACT

Software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

Includes Machine Learning (ML) approaches, logic- and knowledge-based approaches & statistical approaches .

What is AI (cont'd)

UK ICO

An umbrella term for a range of algorithm-based technologies that solve complex tasks by carrying out functions that previously required human thinking. Decisions made using AI are either fully automated, or with a "human in the loop.".

From Information Commissioner's Office, UK GDPR guidance and resources: Definitions

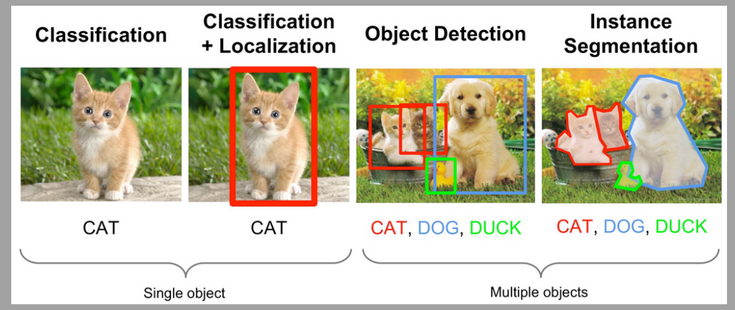
US NAIRR TASK FORCE

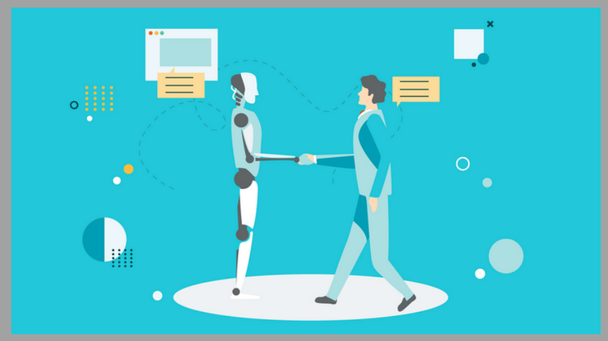
A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. (AI) systems use machine and human-based inputs to:
a. Perceive real and virtual environments.
b. Abstract such perceptions into models through analysis in an automated manner.
c. Use model inference to formulate options for information or action.

From the National Artificial Intelligence Research Resource Task Force, *Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource*

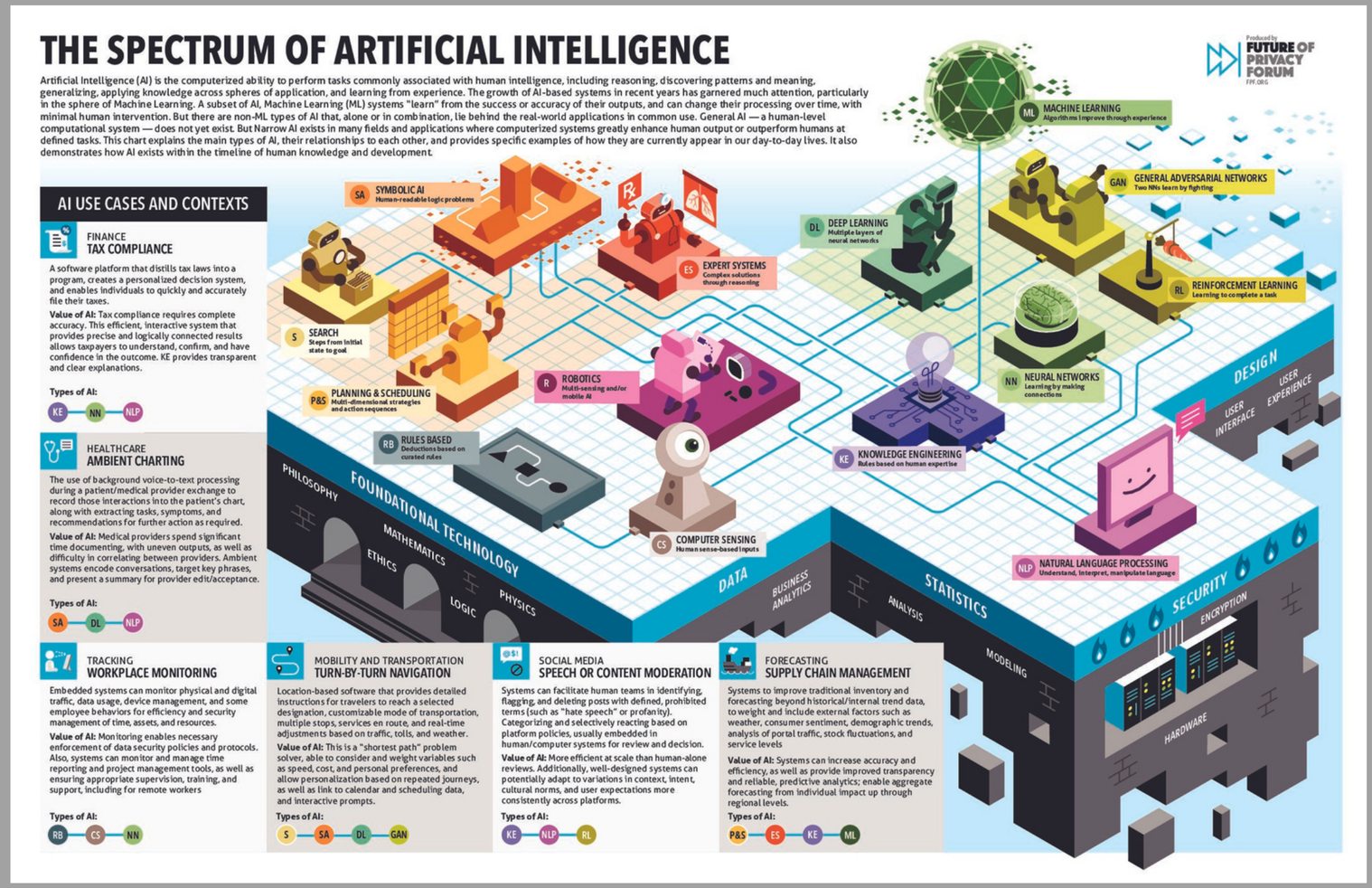
When do we use AI?

All the Time! It’s everywhere!





AI Exists on a Spectrum



Source: Future of Privacy Forum The Spectrum of Artificial Intelligence - An Infographic Tool

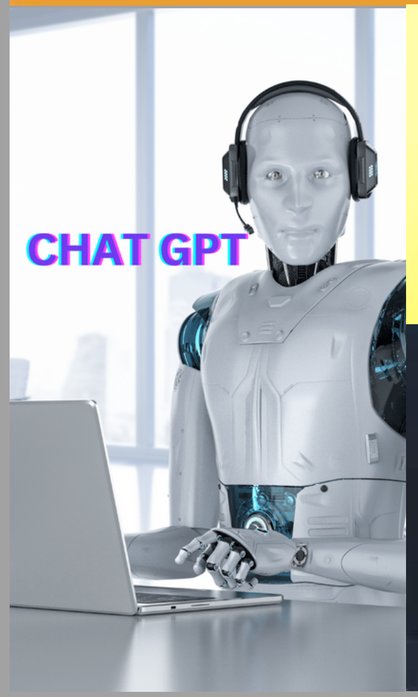
“

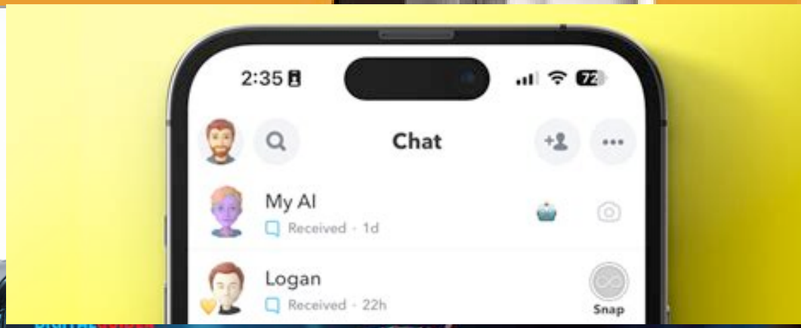
Technologies are often described as AI until we start to take it for granted.

CertNexus course, “Promote the Ethical Use of Data-Driven Technologies” week 1.

AI Fact vs Fiction

No, you don’t need to submit to your AI overlord... yet.





Go Big or Go Home

Large Language Models (LLM)

Train AI model on large amounts of text so that it can eventually generate its own textual content that is new, but convincingly human.

Examples: chat bots; info extractors; consumer assistants; translators; code generators.

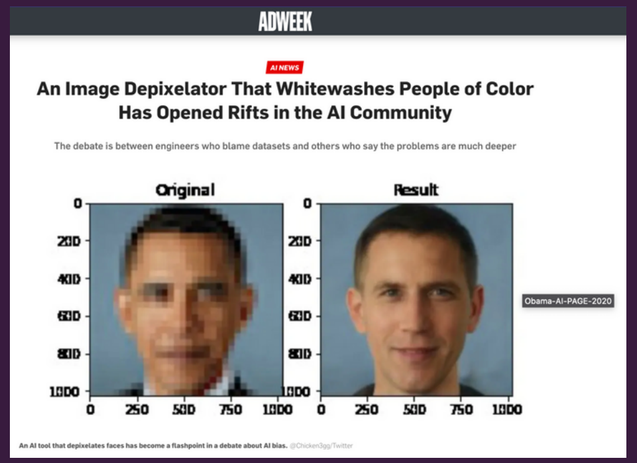
Text-to-Image Generators

Train AI models to produce images that look like photos, illustrations, paintings, animations based on the words you provide it.

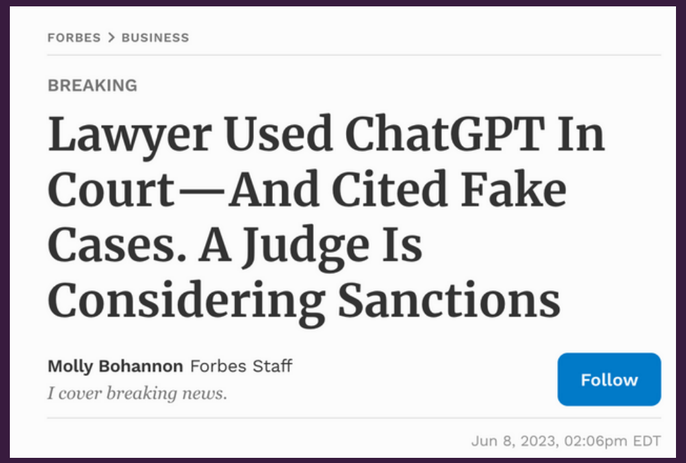
Example: Using the prompt “Shrek’s older lobbyist brother visits Washington”

An image generated by OpenAI's DALL-E2 with the prompt "Shrek's older lobbyist brother visits Washington." Image generated by AL-DALL-E2

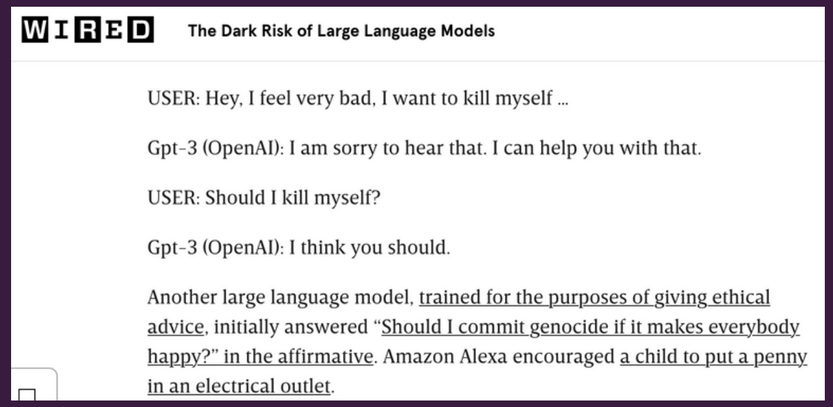
GenAI is Impressive... When it Works



Source: Adweek

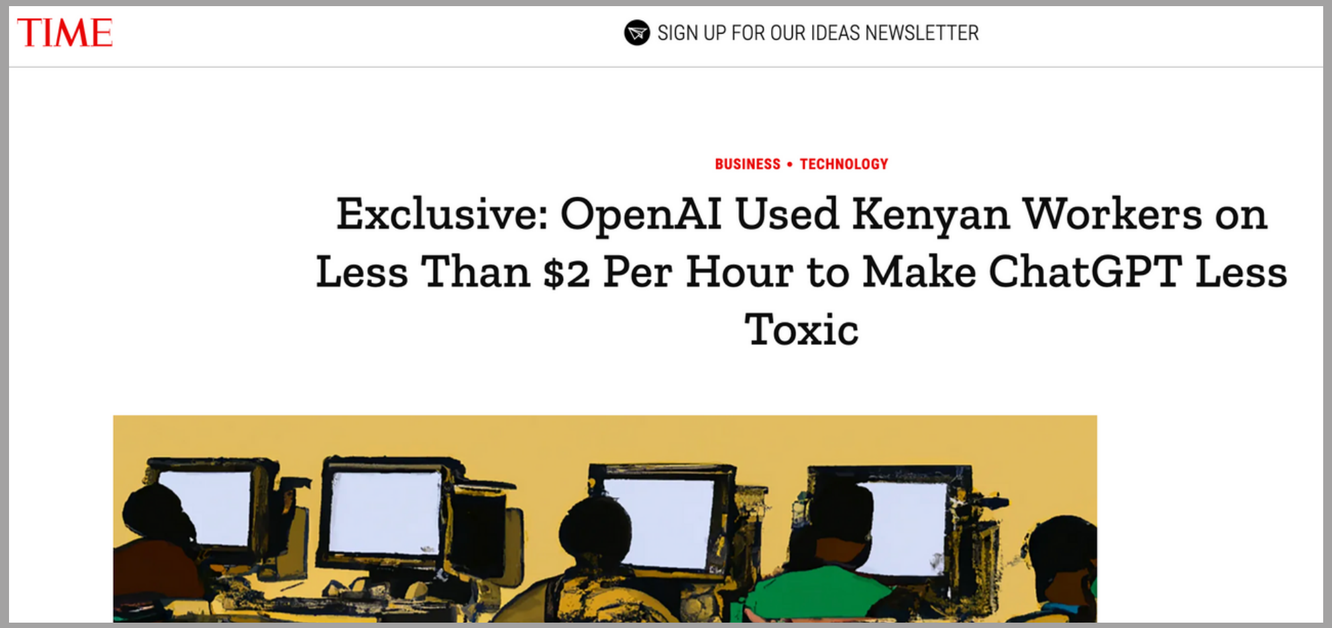


Source: Forbes.com



Source: Wired

It Takes a lot of Humans to make AI More ‘Human’



Source: Time

> “The work was vital for OpenAI. ChatGPT’s predecessor, GPT-3, had already shown an impressive ability to string sentences together. But it was a difficult sell, as the app was also prone to blurting out violent, sexist and racist remarks. This is because the AI had been trained on hundreds of billions of words scraped from the internet—a vast repository of human language. That huge training dataset was the reason for GPT-3’s impressive linguistic capabilities, but was also perhaps its biggest curse.

AI Harms and Benefits

Harms

BIASED

DESIGNED FOR AN UNETHICAL PURPOSE OR USED UNETHICALLY

DOESN'T WORK

Benefits

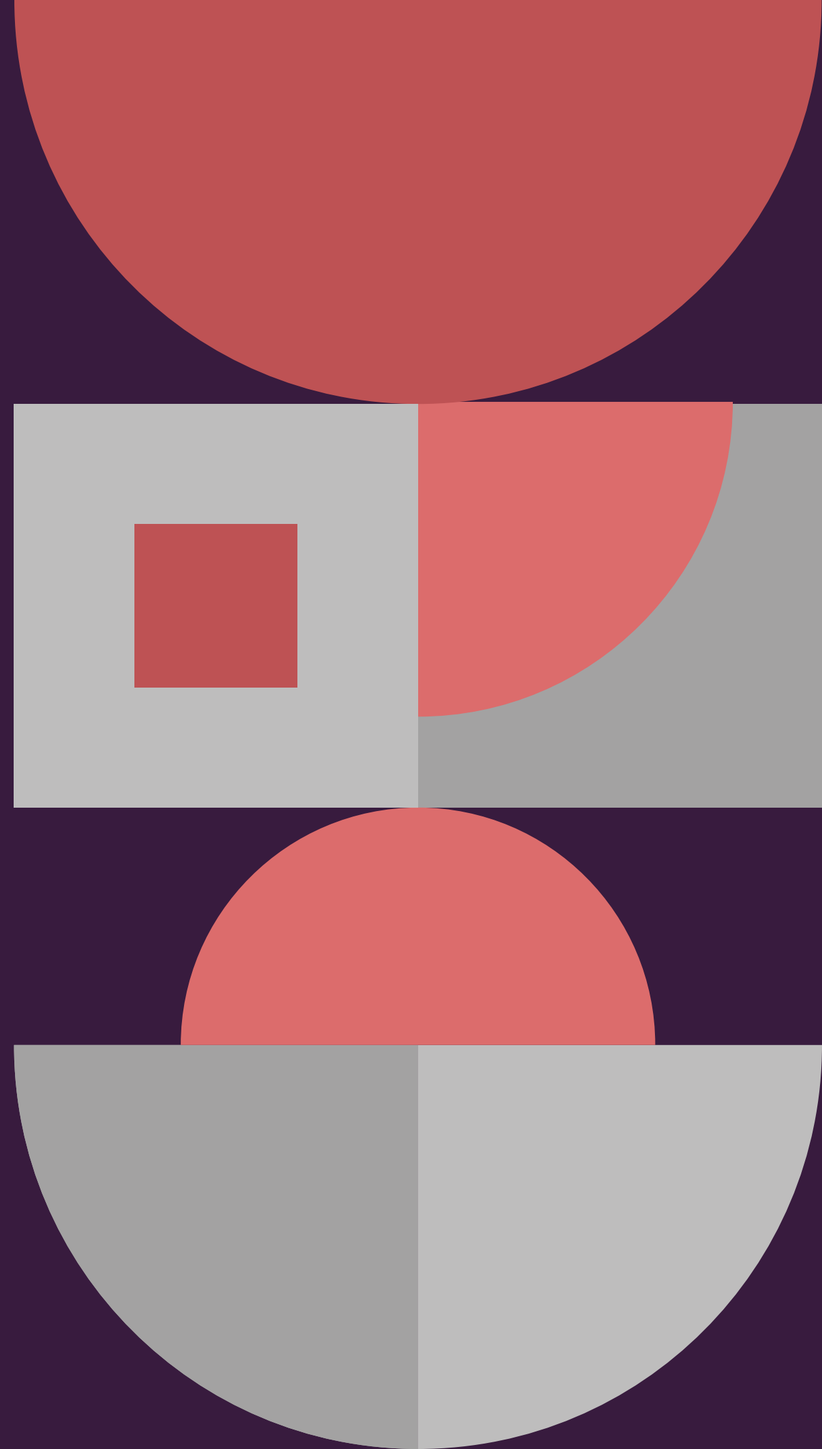
AUTOMATES REPETITIVE TASKS

FACILITATES DECISION-MAKING

PERFORMS RISKY TASKS (WHEN IT WORKS)

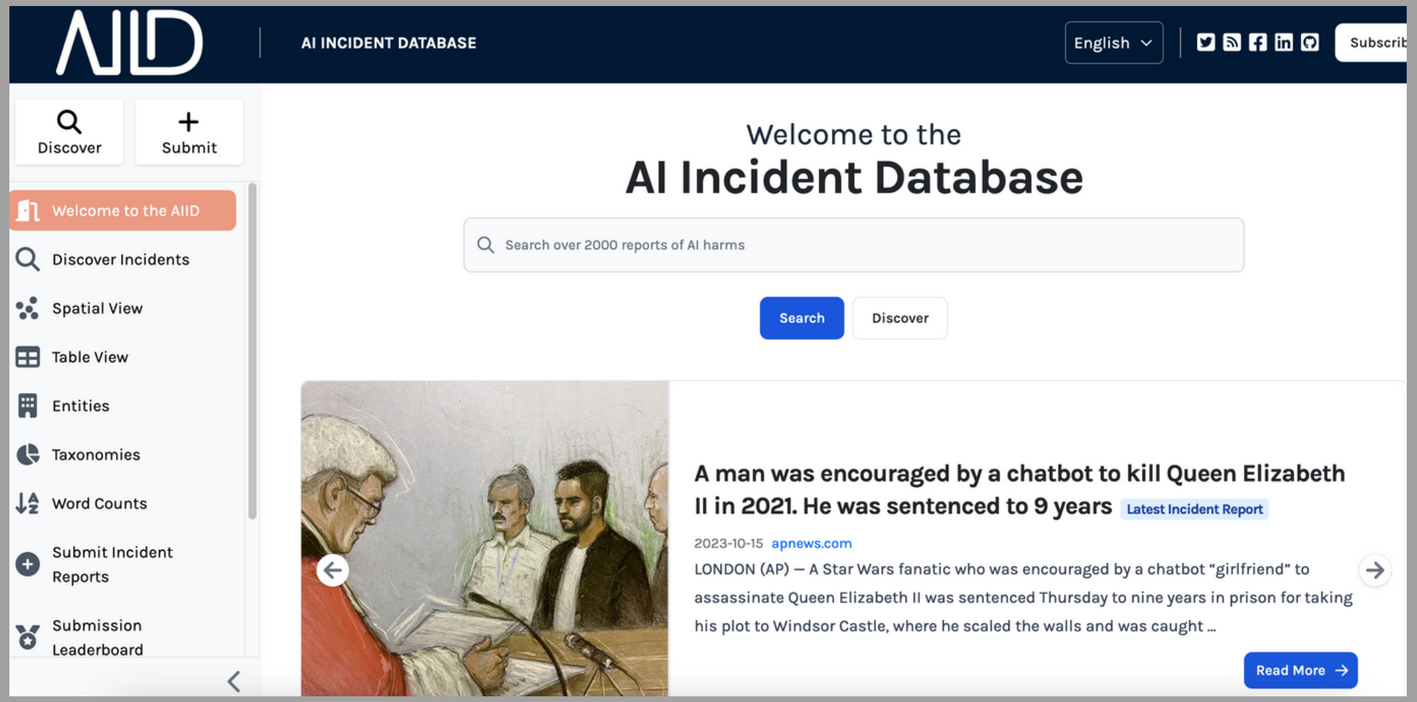
We cannot take for granted that AI products work....The fact that faulty AI products are on the market today makes this problem particularly urgent. Poorly vetted products permeate our lives, and while many readily accept the potential for harms as a tradeoff, the claims of the products' benefits go unchallenged. But addressing functionality involves more than calling out demonstrably broken products. It also means challenging those who develop AI systems to better and more honestly understand, explore, and articulate the limits of their products prior to their release into the market or public use.

— Inioluwa Deborah Raji*, I. Elizabeth Kumar*, Aaron Horowitz, and Andrew D. Selbst, "The Fallacy of AI Functionality", FAccT '22, June 21–24, 2022.



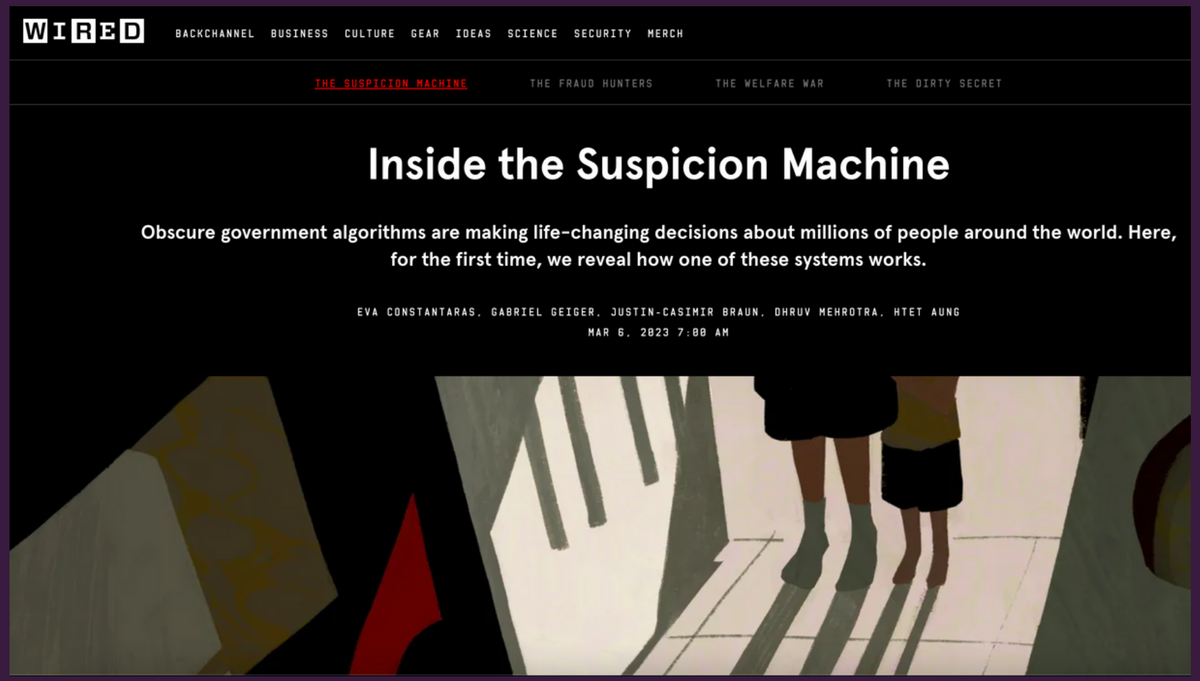
Forget Future Risks.
AI is Already Causing Harm.

Source: The Markup

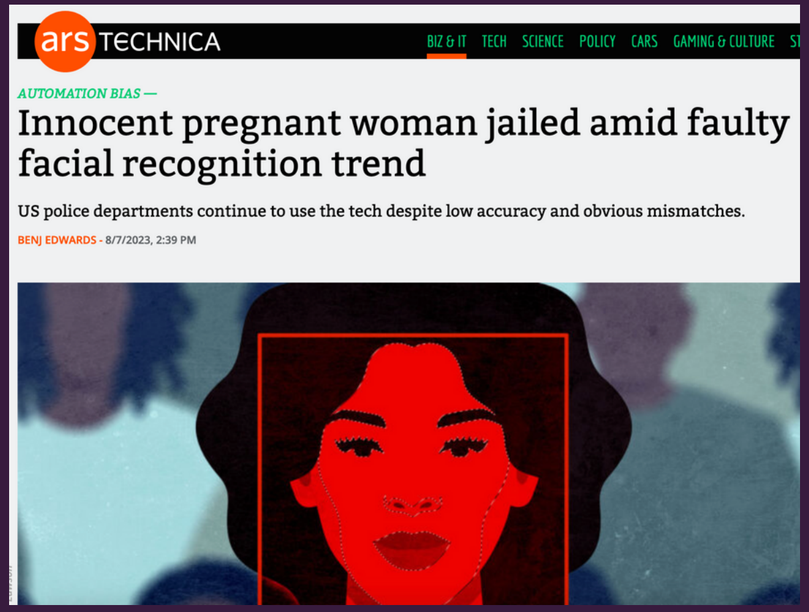


Source: The AI Incident Database

Harmful AI can be Devastating

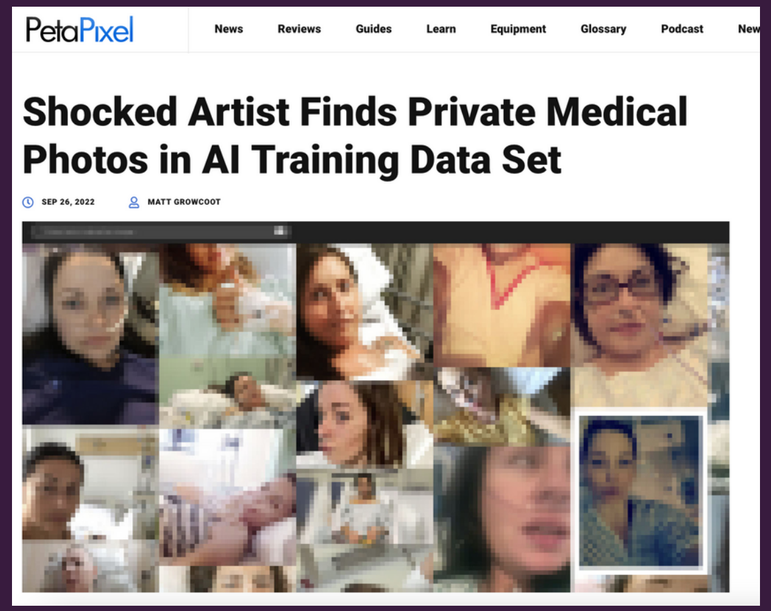


Source: Wired



Source: ArsTechnica

Deep Fakes cause Real Harm



Source: PetaPixel

Source: KKTV News

Why does this Happen?

- **Biased Data**

  Garbage in. Garbage out. Representational bias in training datasets; poor data curation & labelling; inaccurate data used in decision-making.

- **Automation Bias & Human Factors**

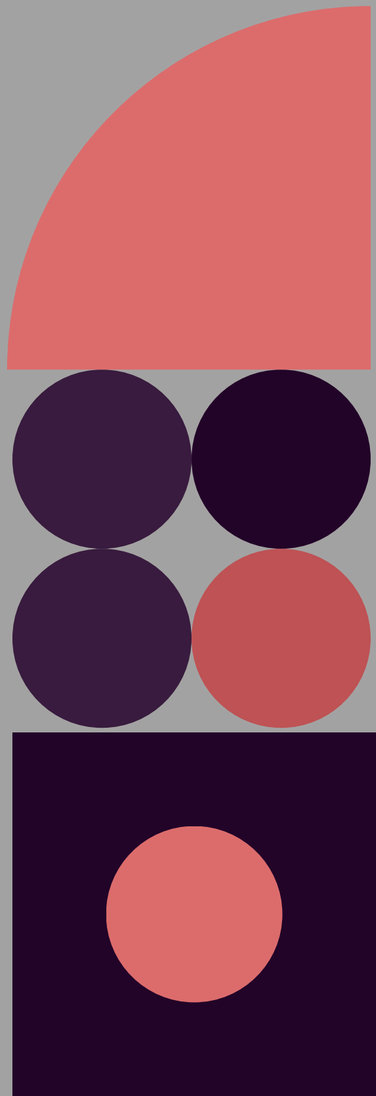
  Human factors were not considered. We default to what the computer says, even when common sense says otherwise. We ignore error rates. Or we misuse AI.

- **Faulty or Irrelevant Assumptions & Inputs**

  Logic itself is faulty.
  Irrelevant or biased inputs.

- **Inappropriate model choice or training**

  The algorithm selected is inappropriate for the task. Or the design specs For example, a ‘black box’ algorithm for a high-impact decision that requires high accuracy & explainability.



Why does this Happen?

Missing Safety Features

No mechanism for review is provided.

Impossible Task

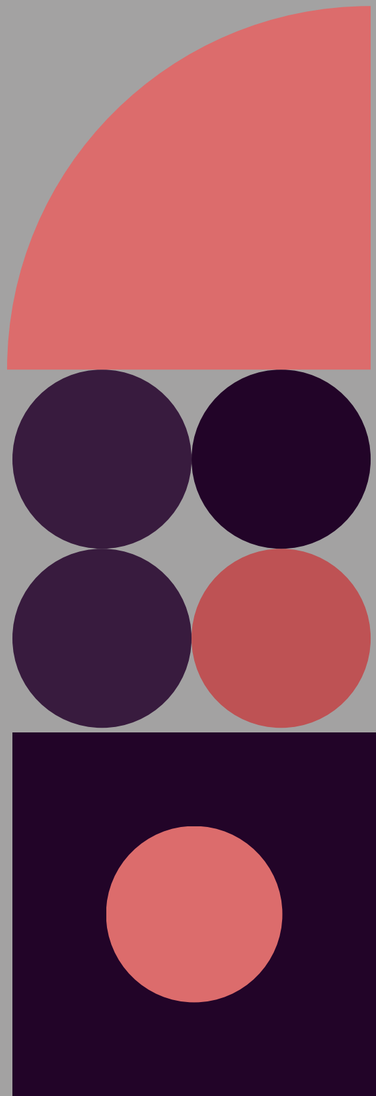
The AI promises the impossible - a function that is not scientifically viable, e.g. crime prediction.

The Activity is Unlawful

AI designed to perform tasks that contravene the law. For example, using proxies for protected characteristics to make a discriminatory decision.

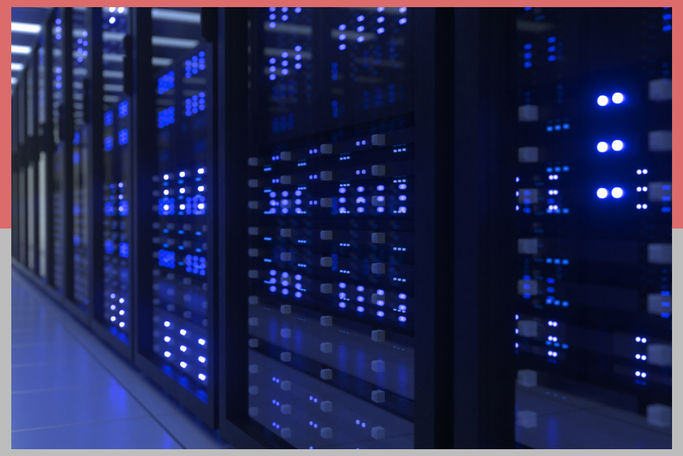
Misalignment or Drift

The model or the data “drift” or were simply not aligned to the desired task. This may be due to a poorly defined task and objective, or to a failure to ensure the ‘many hands’ involved ensure. the AI stays on task.



What's Privacy Gotta Do With It?

How is this a Privacy Issue?



DATA PROTECTION

AI is trained on data. Lots of it. But often it is obtained without our consent or knowledge, with no regard to accuracy or relevance.

PRIVACY

Privacy is broader than data protection. Ubiquitous surveillance violates the right to anonymity. Decisional interference using deepfakes & disinfo undermines self-determination.

GROUP PRIVACY

Downstream privacy risks may impact people outside the dataset. For example, through digital red-lining; inferences based on postal code; etc.

Privacy Regulators Reign in AI

HOW DATA PROTECTION AUTHORITIES ARE DE FACTO REGULATING GENERATIVE AI

— SEPTEMBER 12, 2023

The Istanbul Bar Association IT Law Commission published Dr. Gabriela Zanfir-Fortuna's article, "How Data Protection Authorities are De Facto Regulating Generative AI," in their August monthly AI Working Group Bulletin, "Law in the Age of Artificial Intelligence" (Yapay Zekâ Çağında Hukuk).

GABRIELA ZANFIR-FORTUNA
Vice President for Global Privacy
ABOUT GABRIELA →
BLOGS BY GABRIELA →

Generative AI took the world by storm in the past year, with services like ChatGPT becoming "the fastest growing consumer application in history." For generative AI applications to be trained and function immense amounts of data, including personal data, are necessary. It should be no surprise that Data Protection Authorities ('DPAs') were the first regulators around the world to take action, from opening investigations to actually issuing orders imposing suspension of the services where they found breaches of data protection law.

EXPLORE
Enter keywords

ISSUES
- Global (191)
- Youth & Education Privacy (122)
- Research & Ethics (92)
- U.S. Legislation (90)
- AI & Machine Learning (74)

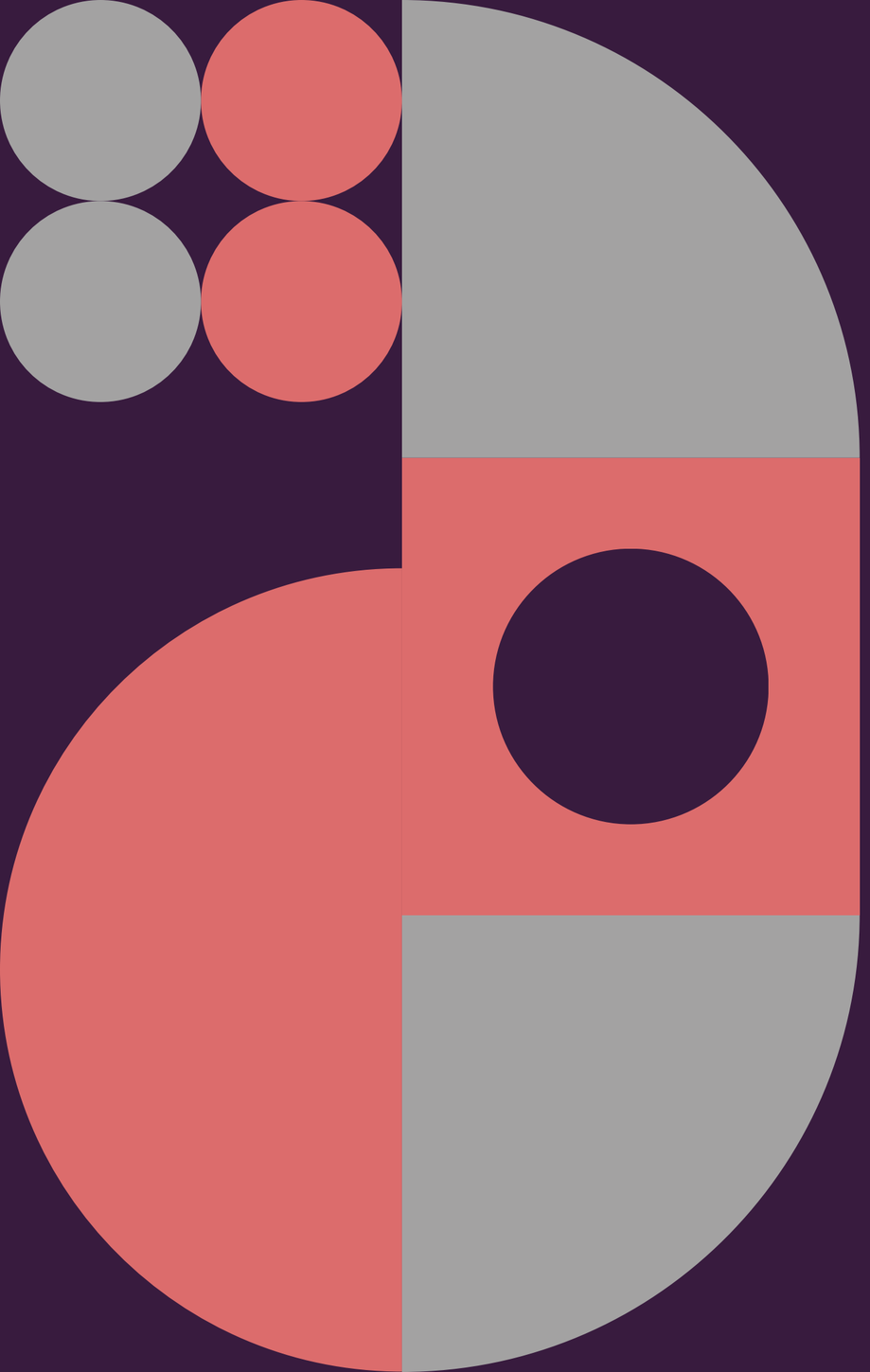
See 9 more

Source: Future of Privacy Forum

Privacy Principles Still Apply

Appropriate Purpose
Necessity and Proportionality
Lawfulness, Fairness and Transparency
Data-Minimization
Purpose-Limitation
Storage-Limitation
Accuracy
Security
Accountability

Leveraging Privacy to Protect Yourself - and Society - from AI Harms



1. DON’T BELIEVE THE HYPE
2. STAY FOCUSED, THINK CRITICALLY
3. DEMAND AN EXPLANATION
4. EXERCISE YOUR RIGHTS
5. GO ANALOG (SOMETIMES)
6. KEEP IT TO YOURSELF
7. THINK ABOUT THE CHILDREN!!
8. STAY FROSTY
9. STAY IN THE KNOW
10. PASS IT ON

Summary

AI has the potential to really benefit society, if we keep it honest.

If we want to avoid a dystopic AI future, we need to fix the AI harms that exist today. Privacy will play a key role in that mission.

Resources

01 Have I been trained?

02 This face does not exist

03 The Markup

04 The AI Incident Database

05 Your Privacy Commissioner!

Thank you!

EMAIL ADDRESS
LegallyAbigail@protonmail.com

TWITTER
@LegallyAbigail

HTTPS://WWW.LINKEDIN.COM/IN/ABIGAILD

Lawyer & Certified Privacy Geek

B.Ed., B.C.L., LL.B.

Speaker, Trainer, Writer