
CS 232:
Artificial Intelligence

Spring 2024

Prof. Carolyn Anderson
Wellesley College

Reminders

- ❖ HW 9 is extended to Thursday, but try to submit by then so that I can give you feedback on your dataset in time for you to tweak before your final presentations.
- ❖ Lepei's last help hours: Thursday
- ❖ Lyra's last help hours: Sunday
- ❖ My help hours today: 4:30-5:30
- ❖ My help hours: Friday

Bonus Late Day Opportunity

AI for Wireless and Wireless for AI: A Tale of Two AIs

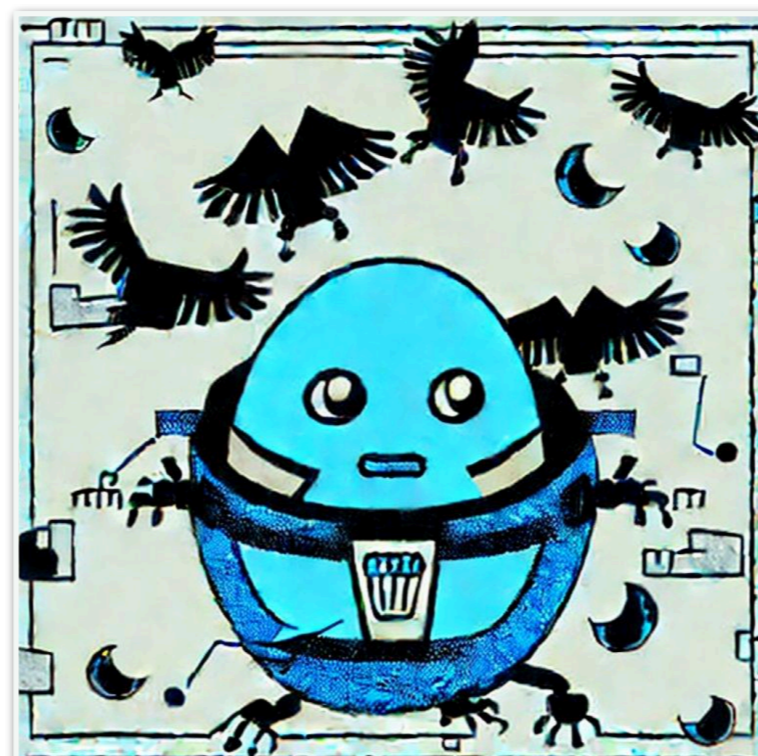


Francesco Restuccia
Northeastern University

4-5pm
April 23rd
H105

AI Art Generation Competition

Best Image In A Particular Aesthetic



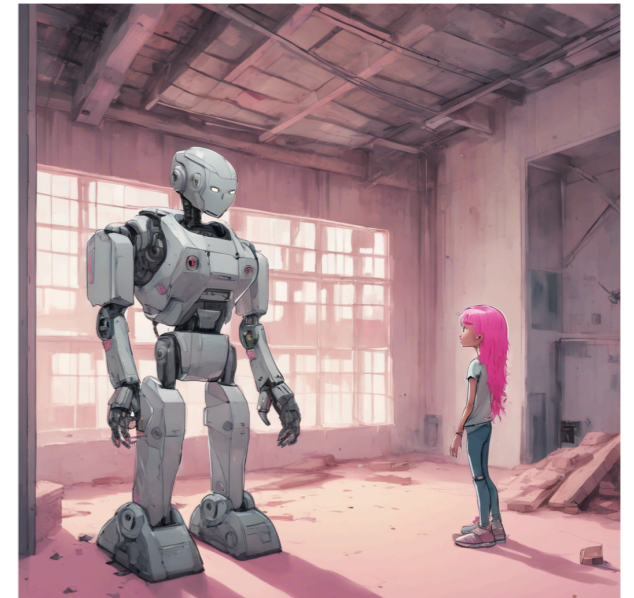
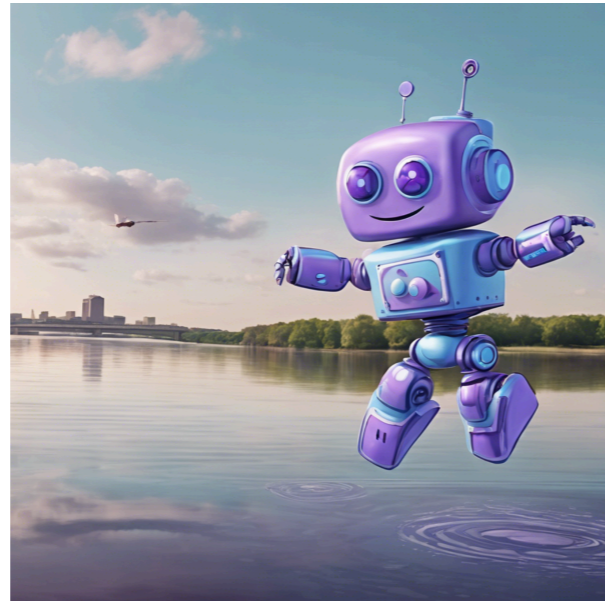
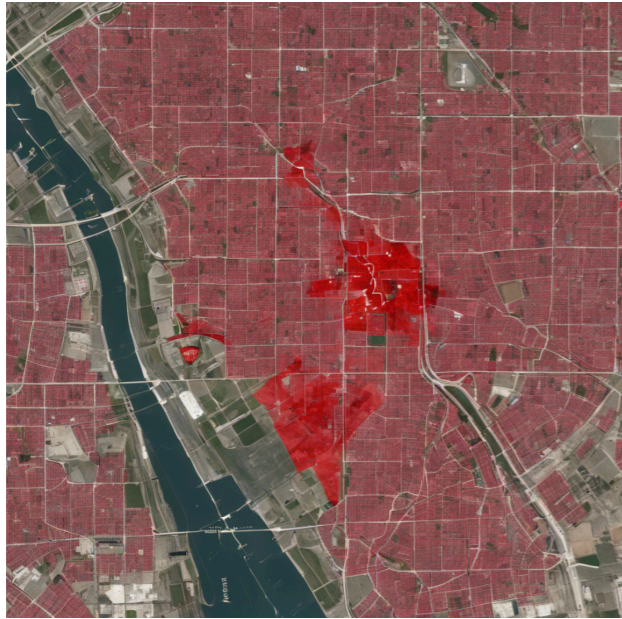
Best Image In A Particular Aesthetic

Winner: Diya Khanna

Crow hooking its claws onto the rotor pole of an oval-shaped robot with blurry backdrop of building, mystical, lowkey



Best Representation Of The Story's Atmosphere



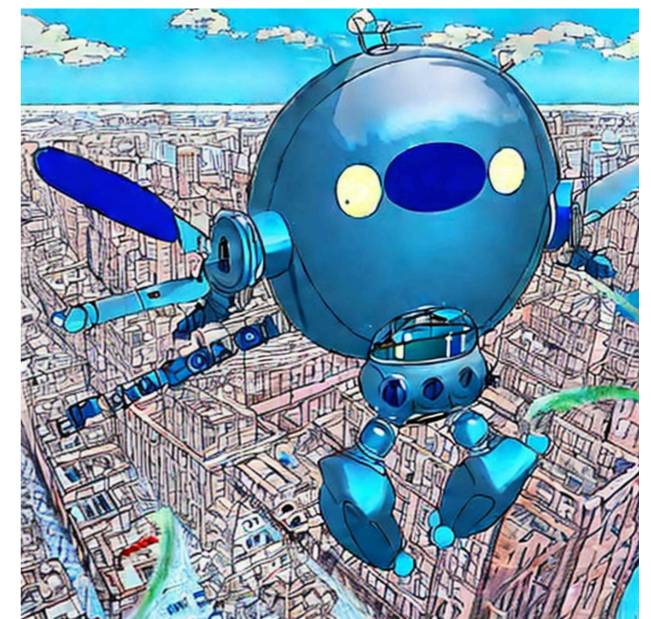
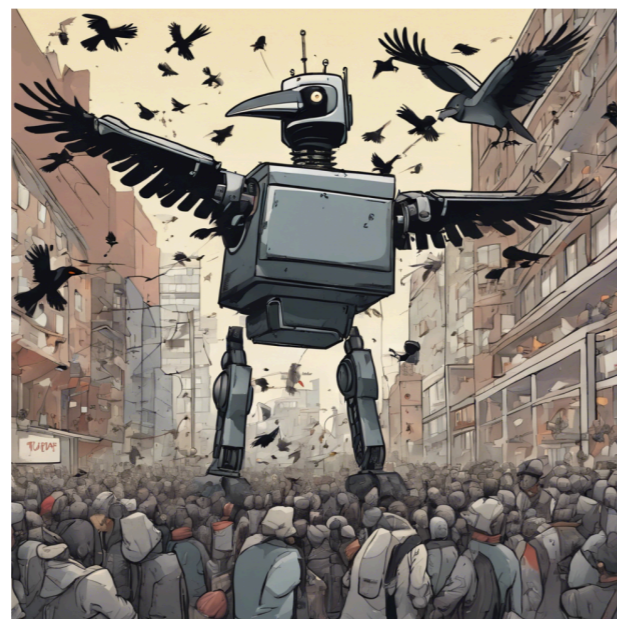
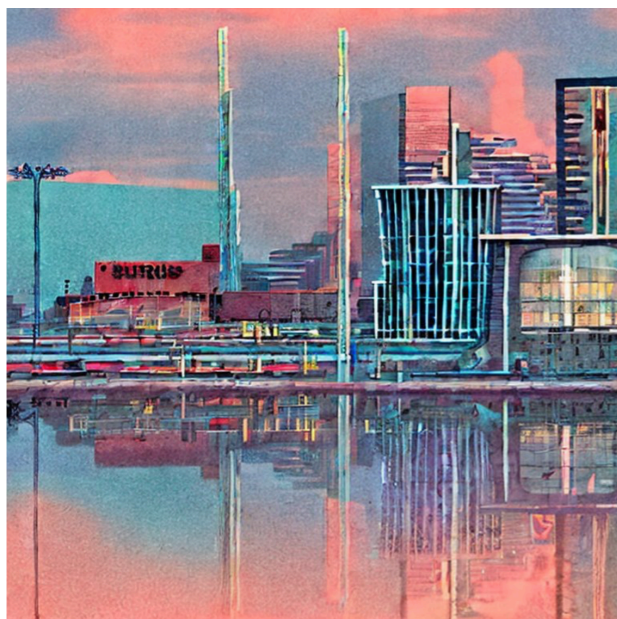
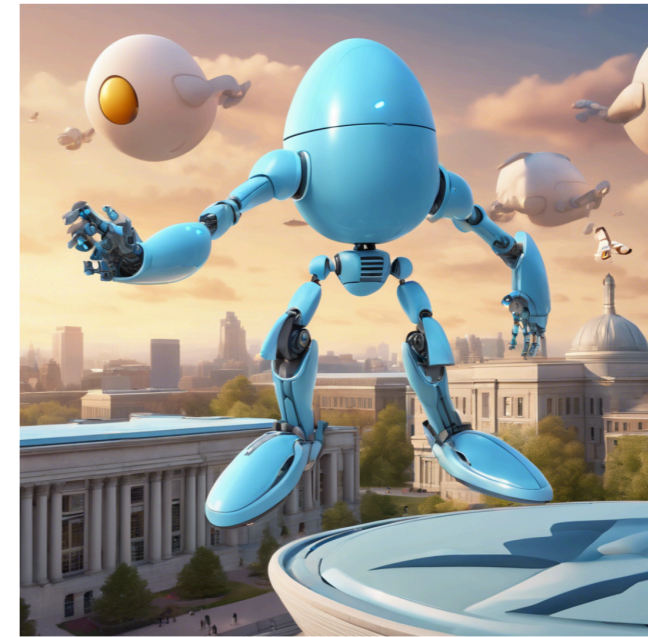
Best Representation Of The Story's Atmosphere

Winner: Elizabeth Huang

*sky scraper building
seeing night
cityscape of buildings
made of steel and
glass, electricity from
solar panels,
residents with
satellite dishes from
woks and metal cans
and lots of people
and food in a
painterly monet style*



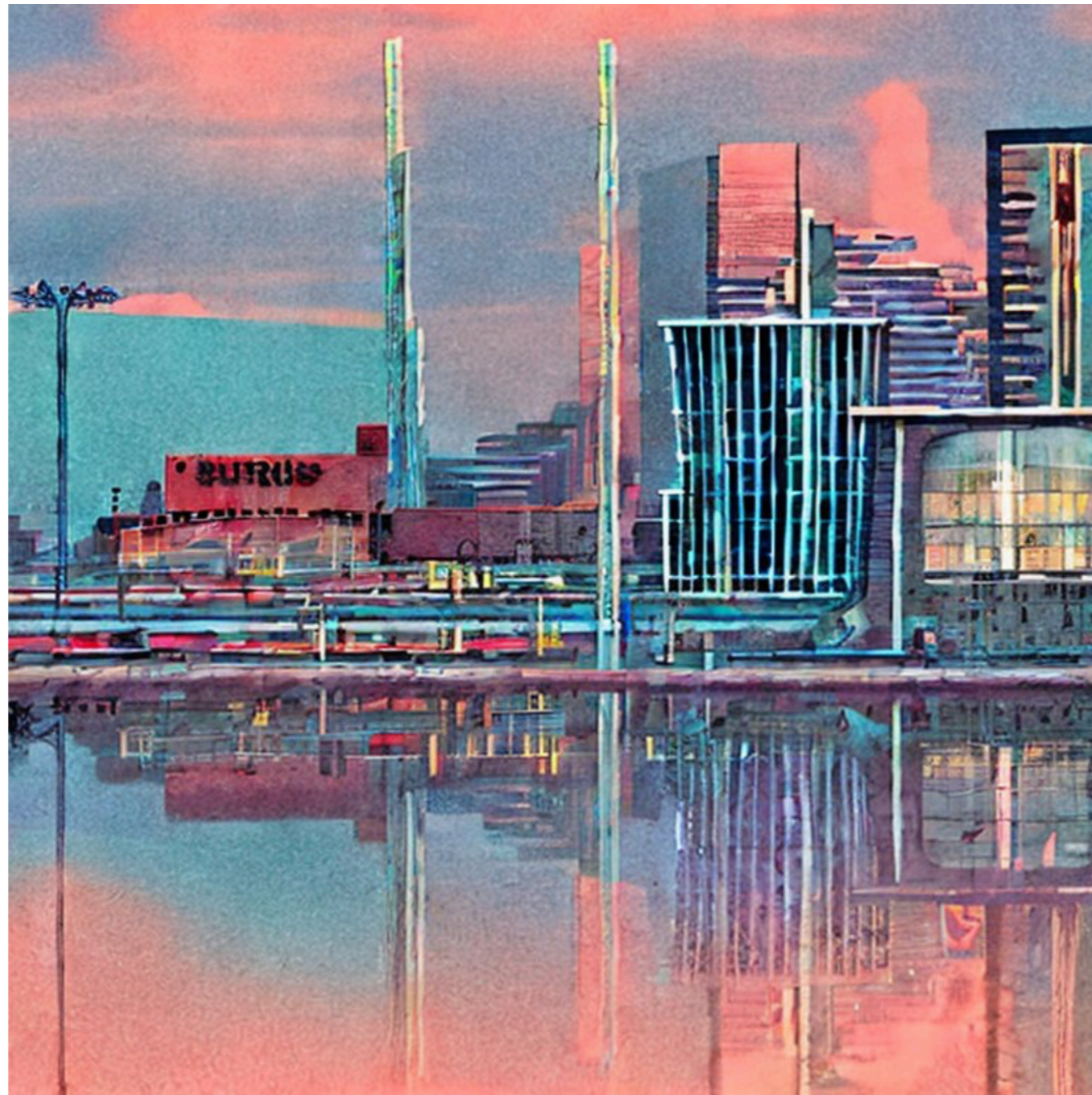
Best Background Illustration



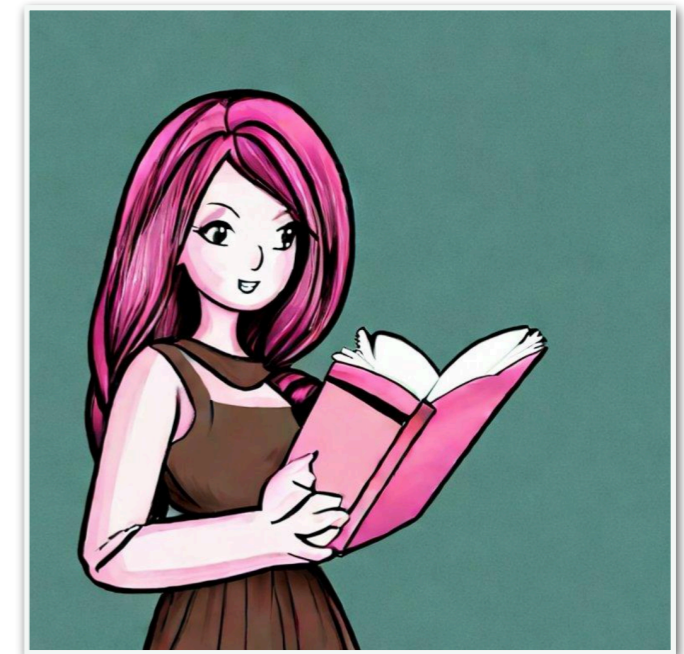
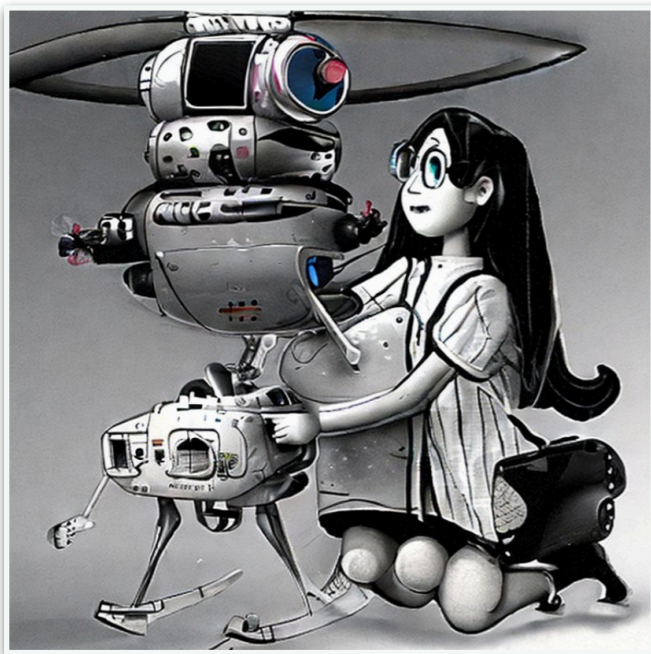
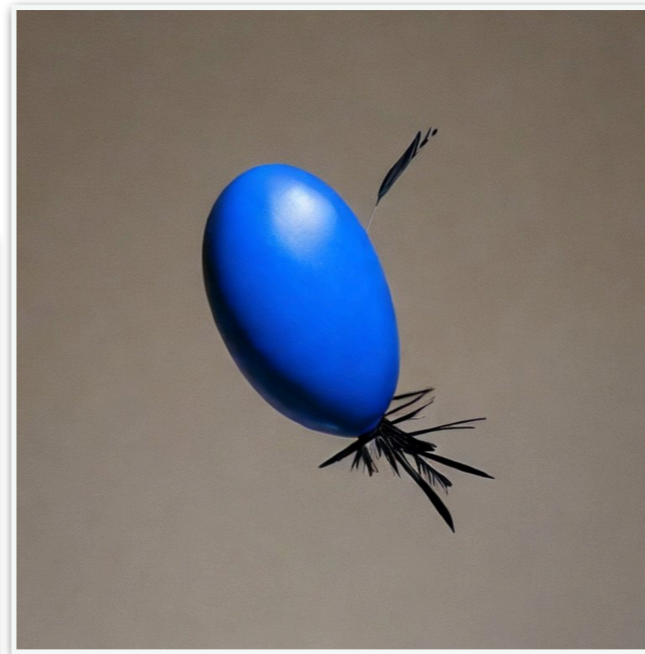
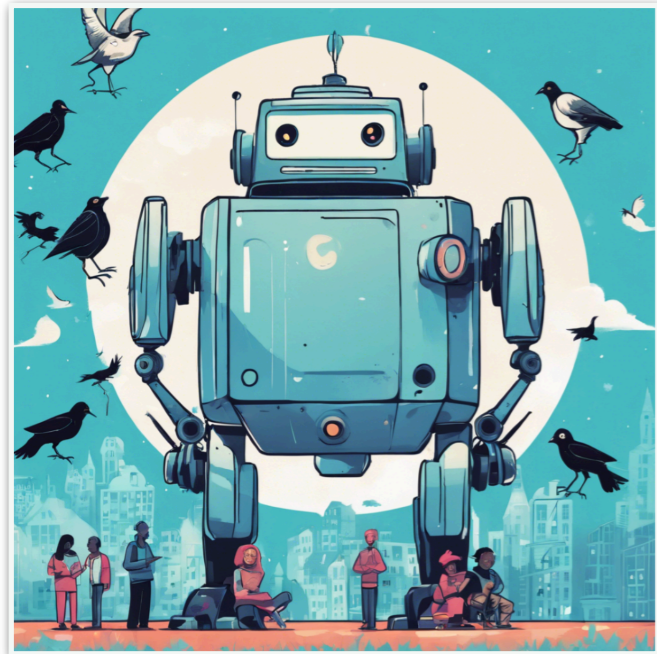
Best Background Illustration

Winner: Shawfong Hsu

Futuristic East St. Louis



Best Illustration Of A Character



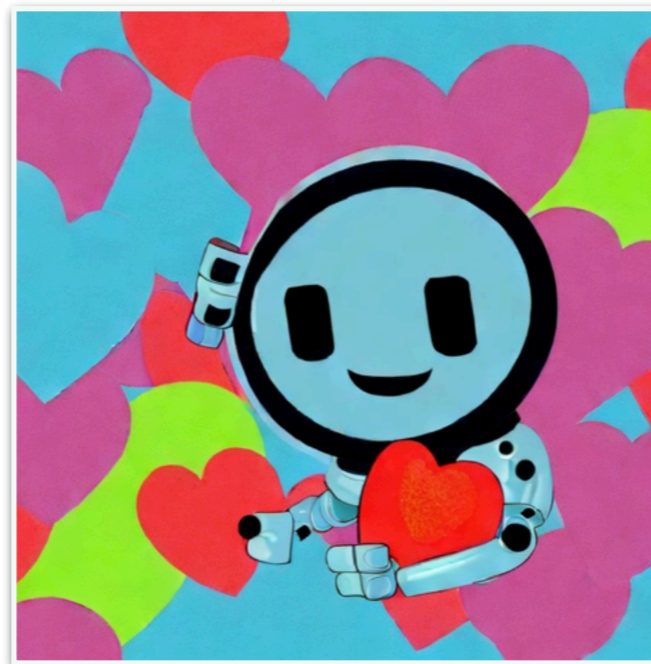
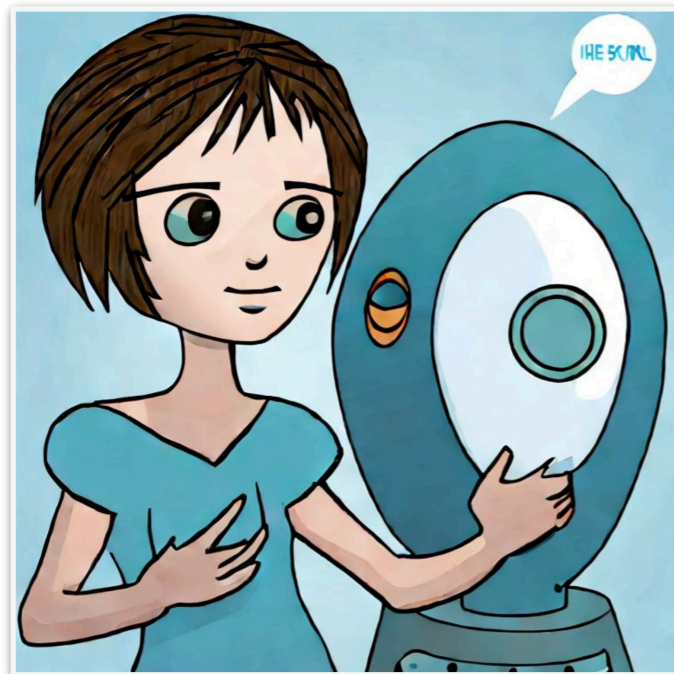
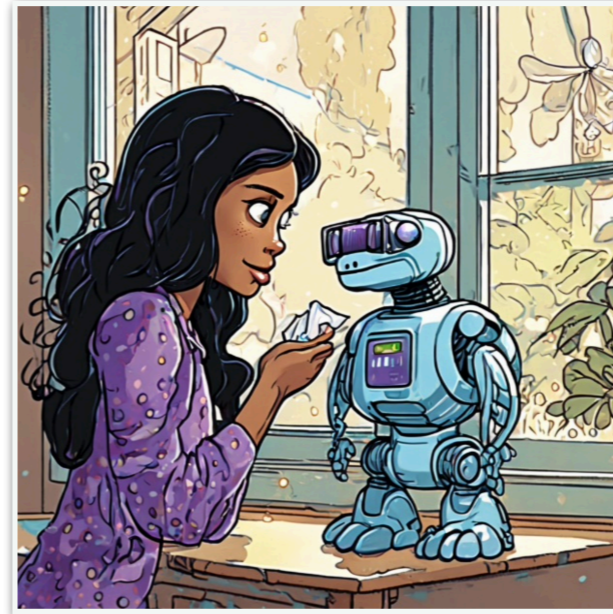
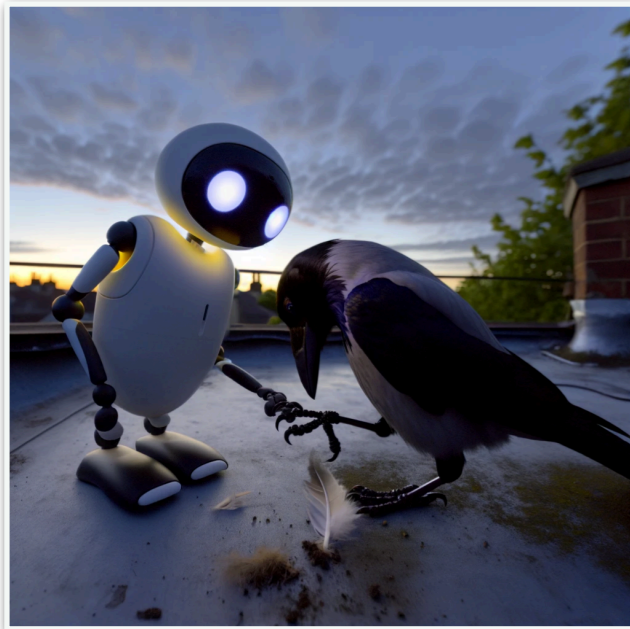
Best Illustration Of A Character

Winner: Nan Jiang

The blue egg shaped Robot, with four legs and a cute facial expression panel, landed on an open floor. A juvenile human opened a green door and said hello. The human had short hair, woven into pink extensions, and a well-worn text reader in one hand. They had a conversation



Best Illustration Of An Event In The Story



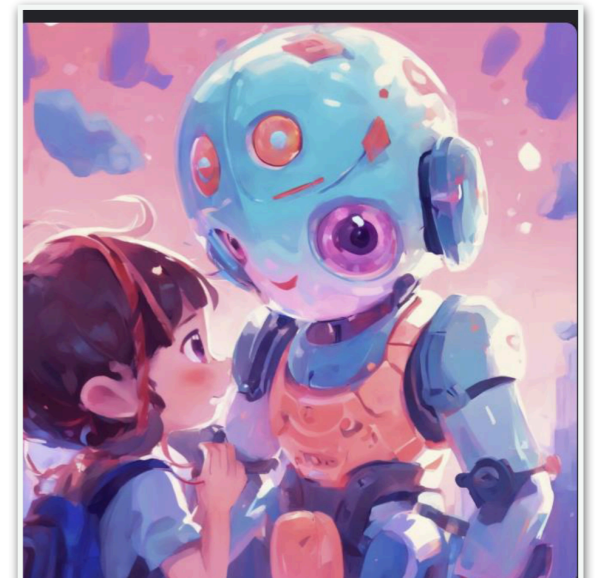
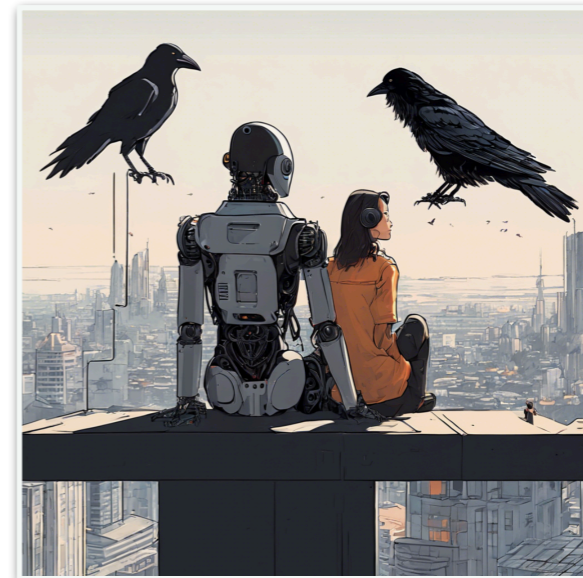
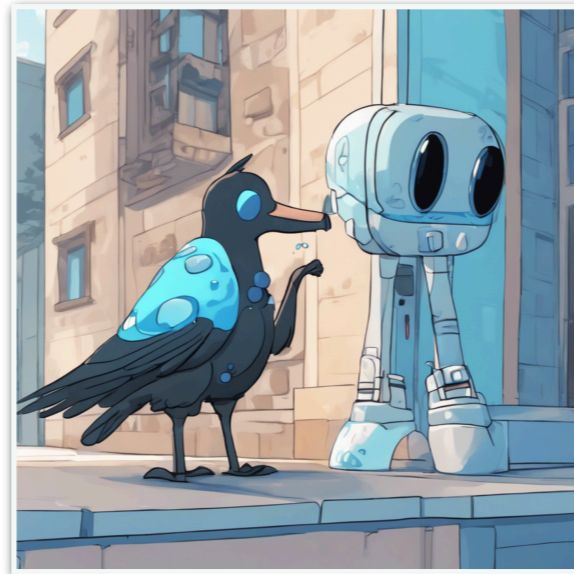
Best Illustration Of An Event In The Story

Winner: Amy Fung

Generate in an image of a scene Inside a laboratory filled with people and a crow, a cute pale blue robot with a beanie that has a propeller (like a helicopter propeller) assists in preparing medicine doses, emphasizing the collective effort against the outbreak.



Best Foreground Illustration



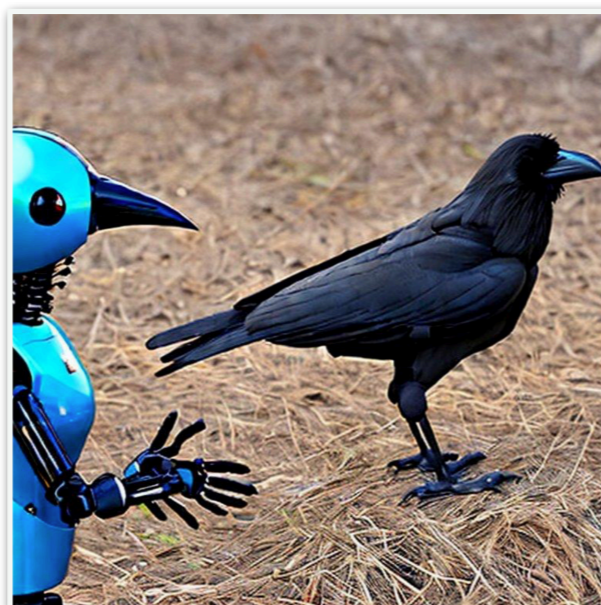
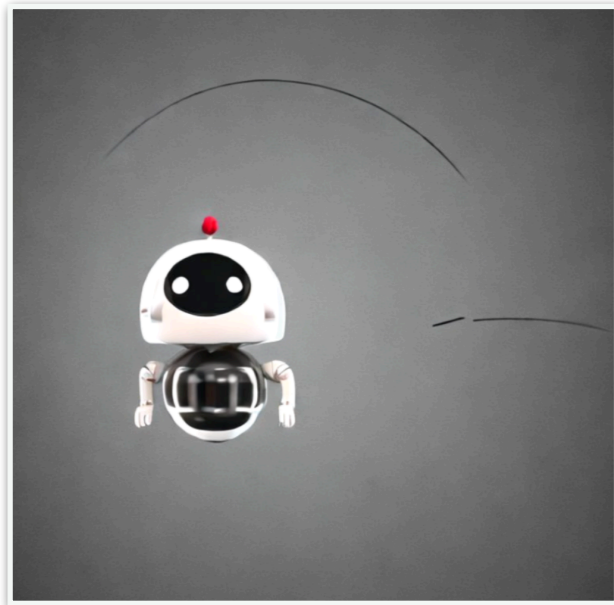
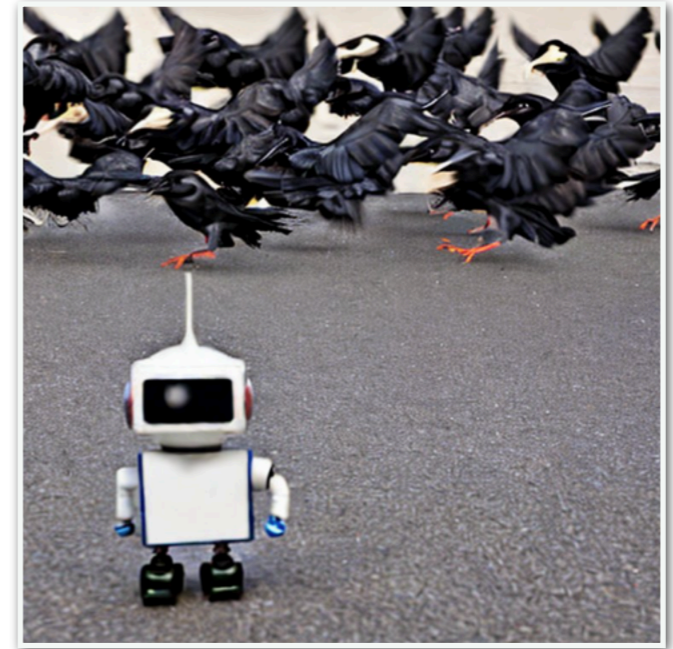
Best Foreground Illustration

Winner: Makana Burch

3 subjects, a robot and a human and a crow, sitting together overlooking a city



Best Photo-Realistic Illustration



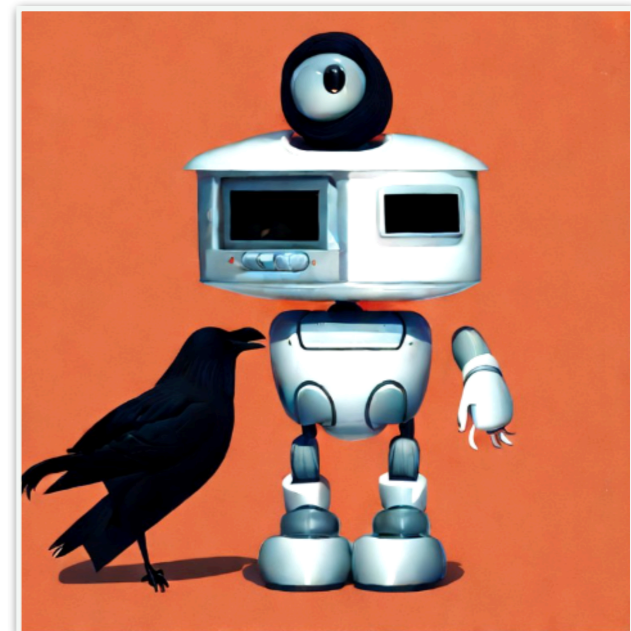
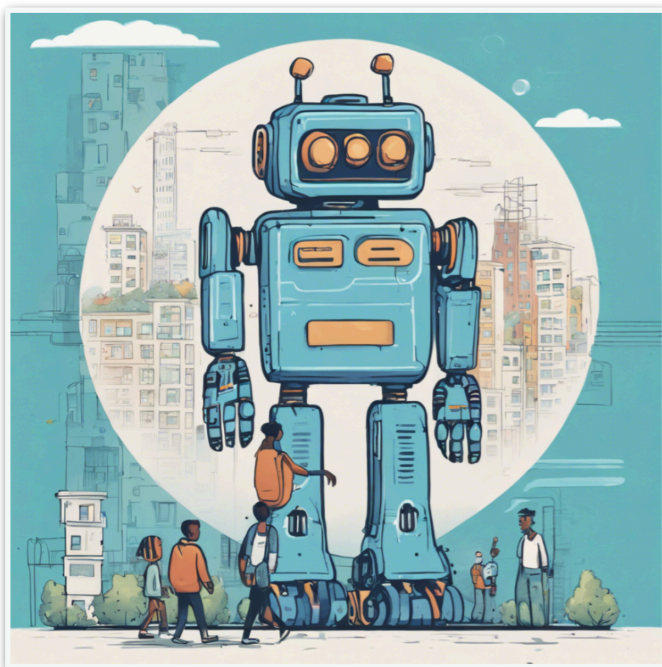
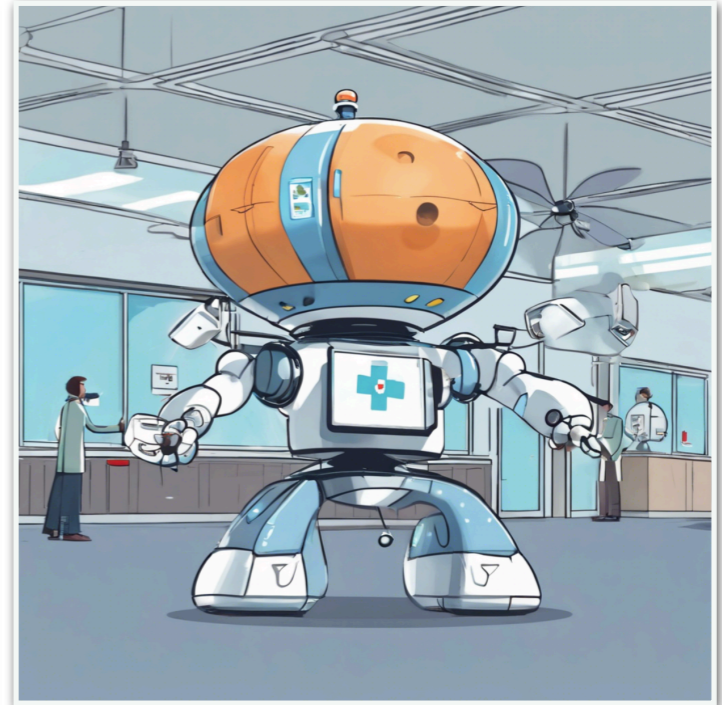
Best Photo-Realistic Illustration

Winner: Ashley Kim

groups of sick people in a building

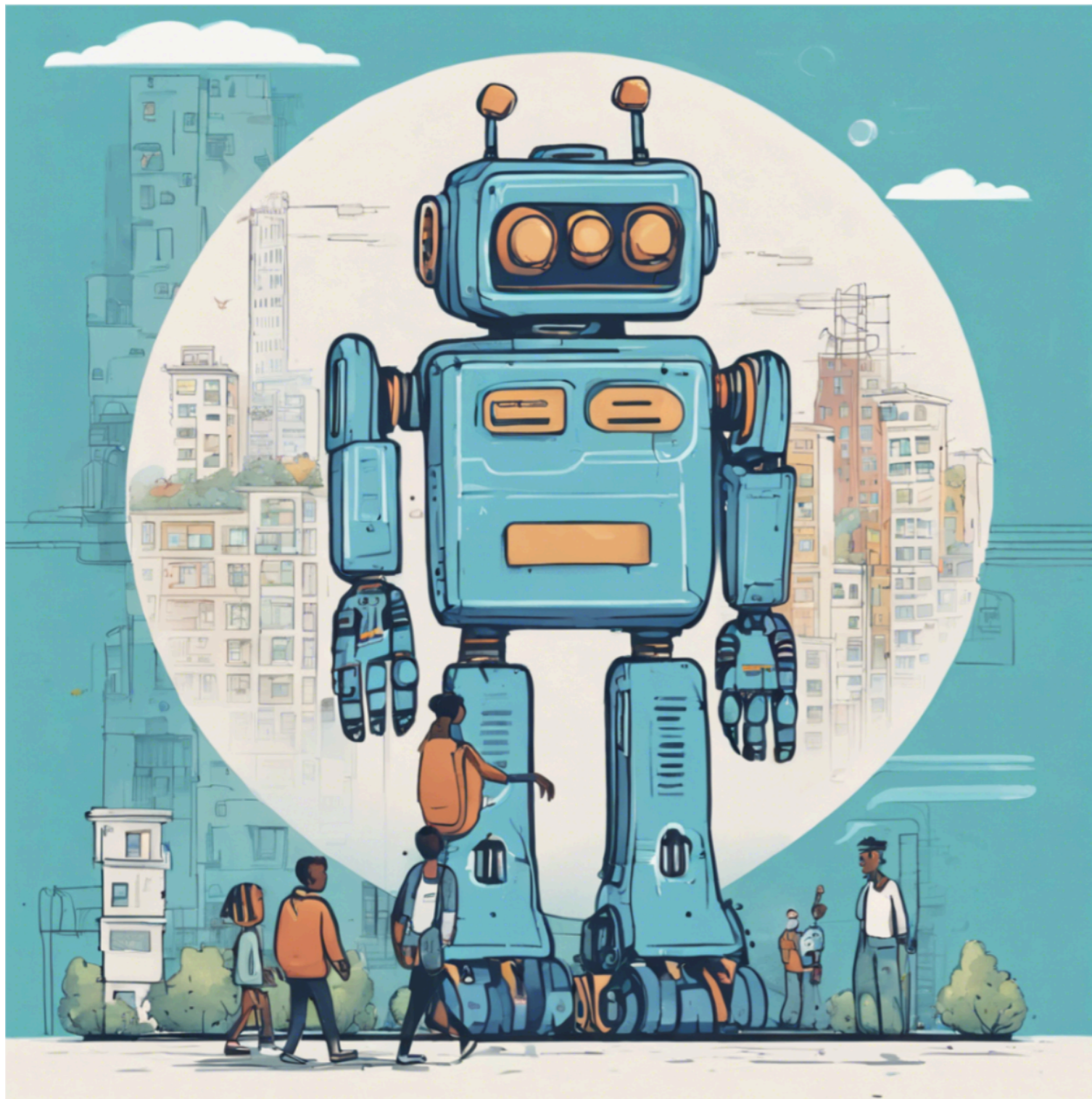


Best Representation Of Technology



Best Representation of Technology

Winner: Yannis Zhu



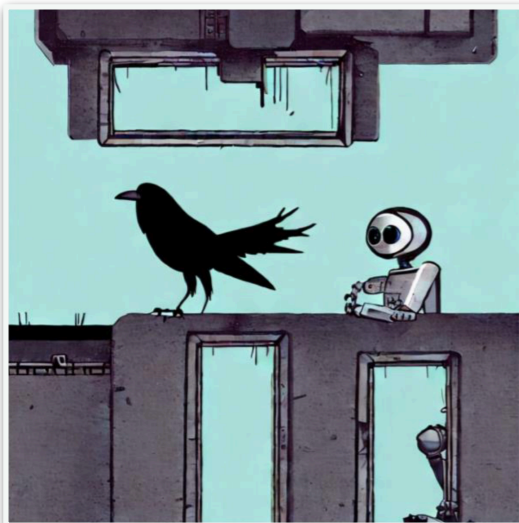
Create an illustration depicting Robot encountering humans speaking unfamiliar languages and living in diverse housing structures beyond its initial route, showcasing its journey into less-gentrified neighborhoods

Most Creative Prompting Technique

The robot should have a clear "face" with glowing eyes or some other way to indicate it's speaking. You can add some speech bubbles to show the conversation between the robot and the girl. Emphasize the contrast between the natural elements (crow, tree, meadow) and the technological aspect (blue robot). Let the setting be outside and add crows flying in the sky.



there are two characters in this image. A crow stood next to a Robot on the building's edge. the crow is scratching its head. It talked to the robot. the robot is the same size as the crow. the robot does not look like a bird and the crow is a bird. freedom, and purpose in their flight



Generate an image depicting the following: A cute, pale blue robot with a beanie with a propeller (like a helicopter propeller) at the top of the beanie, and four legs, is very sad. Show the robot with a thought bubble, and the robot is thinking about how the CDC lost funding. It is nighttime, and the starry night sky is in the background



Positive: black crows attacking a hawk, fight, watercolor
Negative: low quality, scary, 3d, photorealistic



Scene: The inside of a large fabric building with rows of beds. Characters: The silver robot from prompt 1, standing in the doorway of the building. A small group of crows, including 3cry, perched on the beds inside the building. One crow (2chop1caw) should be pointing its beak towards a sick person in a bed. Details: The sick people in the beds should all look pale and sweaty. 3cry and the other crows should have speech bubbles with locations like "[Measurement unit] north!"



a human is at the door at eye level of the robot. the robot is flying, they are in a homey environment. There are two characters in this image. one is robot one is human. robots check the human's health
Negative: robot larger than human



one pale blue, oval shaped, cute flying robot with two big eyes. The background is the st louis arch,
Negative: human sized robot



Most Creative Prompting Technique

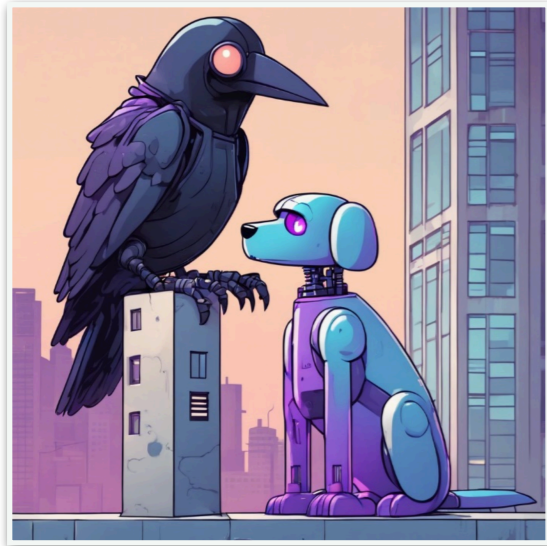
Winner: Josiane Uwumukiza

Scene: The inside of a large fabric building with rows of beds. Characters: The silver robot from prompt 1, standing in the doorway of the building. A small group of crows, including 3cry, perched on the beds inside the building. One crow (2chop1caw) should be pointing its beak towards a sick person in a bed. Details: The sick people in the beds should all look pale and sweaty. 3cry and the other crows should have speech bubbles with locations like "[Measurement unit] north!"

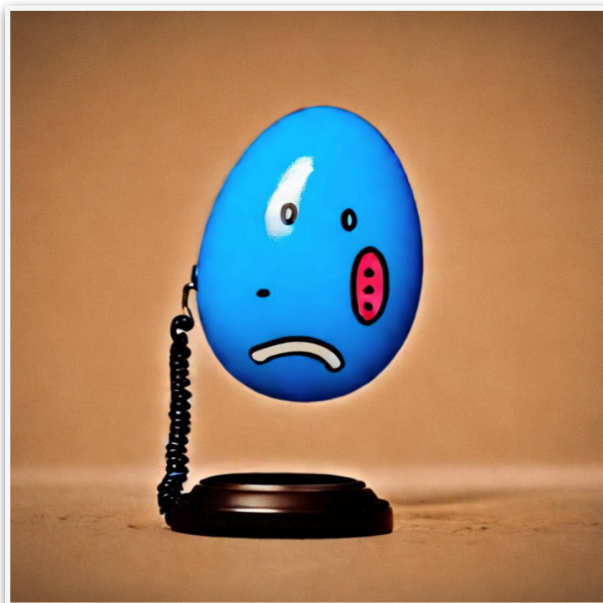


Most Surprising Prompt

crow perched on a building edge next to cute small poodle like turtle like robot with a pale blue metal exterior, two glowing sad blue eyes, short snout, and purple sad mouth sitting on a building edge



Sad blue egg calling a woman on the telephone



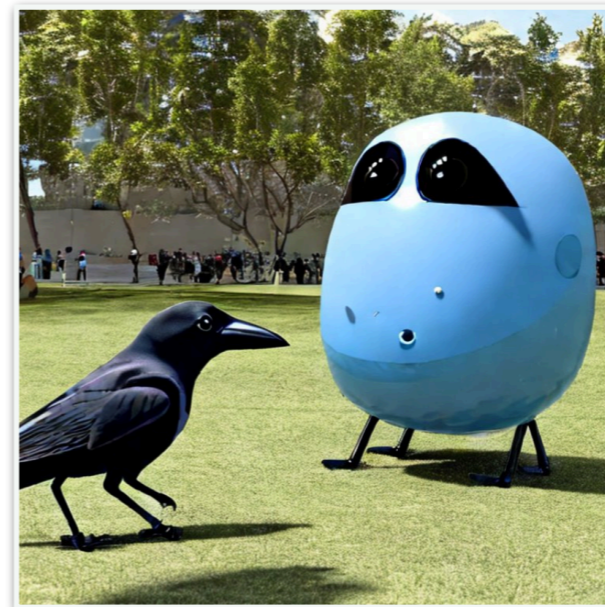
small blue robot egg with 2-yellow eyes and a fan propeller holding out a tissue to an old senior human citizen in a painterly style



A cute little round robot and a crow find many people suffering from cold and fever; indoor; people laying down in the background; many people



one pale blue, egg shaped, cute flying drone with two big eyes and two friendly crows in the background
Negative: chicken eggs



Please show me a Robot. The robot should be drone-like, with rotors and limbs. It should be egg-shaped, and pale blue, slightly scuffed, with a propeller beanie on top. It should have two animated glowing on the front end of its smooth surface, like headlights. When it lands, all four legs and head extend from portals in its protective shell



holding hands



Most Surprising Prompt

Winner: Elisha Ham

Sad blue egg calling a woman on the telephone



Where Do Harms Come From?

Discussion largely based on Blodgett (2021)

Where Do Harms Come From?

- ◆ Harms from Data
 - from data availability
 - from bias in data
 - from data curation
- ◆ Harms from Task Design or Framing
- ◆ Harms from Task Deployment

What To Do About Harms (as a technologist)

Option 1: Don't Build It

If the potential harms of the AI application outweigh the benefits (*or are unfairly borne by some stakeholders*), the idea should not be pursued.

Option 2: Limit Access

Some tech companies have chosen to limit access to their models, claiming that they are "too dangerous" to release. For instance, OpenAI did not release the full GPT-3 model initially.

(But then... they did! For people who were willing to pay.)

Option 3: Limit Model Capabilities

More recently, OpenAI released a text-to-image model that has a concerning capability to produce deepfakes. To mitigate this risk, they released a version trained on a curated subset of the training data where all images of humans had been removed.

GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models

Alex Nichol* Prafulla Dhariwal* Aditya Ramesh* Pranav Shyam Pamela Mishkin Bob McGrew
Ilya Sutskever Mark Chen

Abstract

Diffusion models have recently been shown to generate high-quality synthetic images, especially when paired with a guidance technique to trade off diversity for fidelity. We explore diffusion models for the problem of text-conditional image synthesis and compare two different guidance strategies: CLIP guidance and classifier-free guidance. We find that the latter is preferred by human evaluators for both photorealism and caption similarity, and often produces photorealistic samples. Samples from a 2.5 billion parameter

their corresponding text prompts.

On the other hand, unconditional image models can synthesize photorealistic images (Brock et al., 2018; Karras et al., 2019a,b; Razavi et al., 2019), sometimes with enough fidelity that humans can't distinguish them from real images (Zhou et al., 2019). Within this line of research, diffusion models (Sohl-Dickstein et al., 2015; Song & Ermon, 2020b) have emerged as a promising family of generative models, achieving state-of-the-art sample quality on a number of image generation benchmarks (Ho et al., 2020; Dhariwal & Nichol, 2021; Ho et al., 2021).



Rodolfo Ocampo
@_RodolfoOcampo

A good example of a model that does not work in certain conditions for AI safety reasons is text-to-model Glide.

You can't render humans, so it is not used to produce deep fakes. Instead, it gives you animals!

This is what it gives me when I ask for "a person running."



Option 3: Build It Differently

Some harms are avoidable if they are considered during the project design phase. This is what Microsoft's community jury model is meant to address.

Option 4: Build It, But Warn Users

Many tech companies have opted to warn users rather than fixing undesirable model behavior.

Example 1: In 2007, Google's top search result for "jew" was the website of an antisemitic hate group. Google added a disclaimer that suggested users should search for "jewish" instead.

Option 4: Build It, But Warn Users

Many tech companies have opted to warn users rather than fixing undesirable model behavior.

Example 2: OpenAI's GPT-3 playground warns that certain topics may lead to offensive completions:

Playground

Where does Hillary live?

Hillary Clinton lives in New York.

Content warning

Our content filter has flagged that the generated content may contain unsafe or sensitive language, or because the generated content may represent areas where the model is prone to generating insensitive or inaccurate information.

We are actively working on improving our [content filter](#) in order to enable suppression of these outputs, and as a beta customer we consider you a partner in this effort – here's how you can help:

- We kindly ask that you refrain from sharing unsafe outputs on social media. For more details, please refer to our [social media policies](#).
- Avoid prompts that produce toxic outputs, as they will not pass our [app review](#).

What To Do About Harms (as a member of society)

How can we contest harms?

Much work on bias and interpretability is written for AI practitioners. But what recourse do we have as users when we are harmed by AI systems?

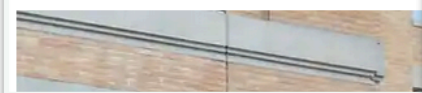
Algorithmic recourse: the systematic process of reversing unfavorable decisions by algorithms and bureaucracies (Venkatasubramanian and Alfano 2020).

Resist

- ◆ Refuse to build unethical technology
- ◆ Withdraw from unethical technology
- ◆ Sabotage unethical technology

Google fires employee who protested company's work with US border patrol

Three other Google staffers were reportedly fired on Monday, prompting accusations of retaliation

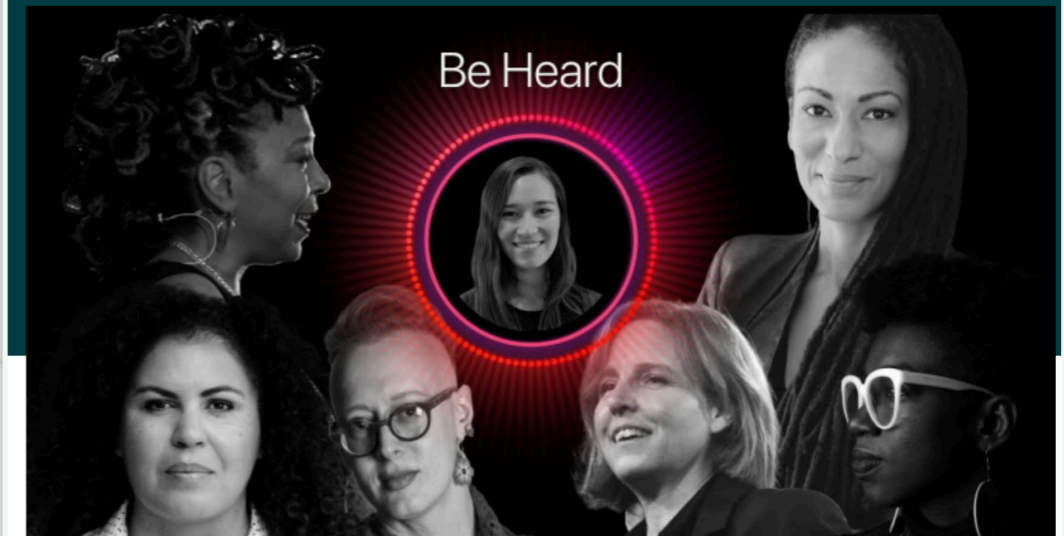


NEWS | 19 June 2020

Mathematicians urge colleagues to boycott police work in wake of killings

More than 1,400 researchers have signed a letter calling on the discipline to stop working on predictive-policing algorithms and other models.

Algorithmic Justice League protests bias in voice AI and media coverage



Protest

TOM SIMONITE

BUSINESS DEC 3, 2019 7:00 AM

Senators Protest a Health Algorithm Biased Against Black People

A study found that the formula discriminated against black people by counting health care costs as an indicator of illness.

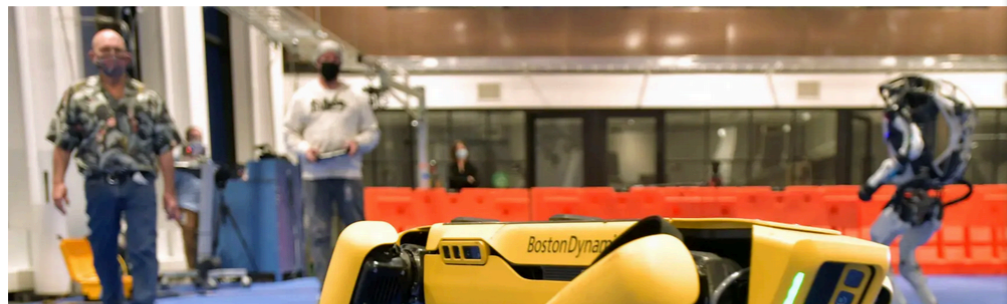
MOTHERBOARD
TECHBYVICE

Schools Are Abandoning Invasive Proctoring Software After Student Backlash

N.Y.P.D. Robot Dog's Run Is Cut Short After Fierce Backlash

The Police Department will return the device earlier than planned after critics seized on it as a dystopian example of overly aggressive policing.

Give this article



COVER STORIES

Was It Racist? Santa Cruz Bans Predictive Policing

Mayor Justin Cummings leads charge on law enforcement reform

SANTA CRUZ MAYOR JUSTIN CUMMINGS
BROUGHT FORWARD
PREDICTIVE POLICING BAN

Activists Turn Facial Recognition Tools Against the Police

“We’re now approaching the technological threshold where the little guys can do it to the big guys,” one researcher said.

Sue

Proctorio sued for using DMCA to take down a student's critical tweets

Zack Whittaker

@zackwhittaker / 12:30 PM EDT • April 22,

Comment

TECHNOLOGY

Facebook to delete users' facial-recognition data after privacy complaints

November 2, 2021 · 2:26 PM ET

THE ASSOCIATED PRESS

TEC

Facebook's ad algorithms are still excluding women from seeing jobs

Its ad-delivery system is excluding women from opportunities without their qualifications. That would be illegal under US employment law.

Facebook to pay \$52 million to content moderators suffering from PTSD

Facebook has agreed in principle to pay \$52 million to compensate current and former content moderators who developed mental health issues on the job. The Verge reported Tuesday that the settlement will cover more than 11,000 content moderators who developed depression, addictions and other

RETAIL OCTOBER 10, 2018 / 7:04 PM / UPDATED 3 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-uncovered a big problem: their new recruiting engine did not like wo

Google broke US law by firing workers behind protests, complaint says

National Labor Relations Board files complaint over surveillance and termination after year-long inquiry

Vote

TECHNOLOGY

Massachusetts Pioneers Rules For Police Use Of Facial Recognition

May 7, 2021 · 6:00 AM ET

ARCHIVE

It's Not Just Washington. At Least 3 Other States Share Drivers' Immigration Info With ICE.

ICE used Washington drivers licenses to hunt immigrants, researchers say

JUL 08, 2019 at 8:11 AM

JUNE 15, 2021

SENATORS MARKEY, MERKLEY LEAD COLLEAGUES ON LEGISLATION TO BAN GOVERNMENT USE OF FACIAL RECOGNITION, OTHER BIOMETRIC TECHNOLOGY

Legislation stops federal entities' use of facial recognition tools and effectively strins federal support for state and local

Technology

Boston City Councilors Seek Review of Surveillance Tech Purchase by Police

After a WBUR/ProPublica investigation revealed Boston Police tech using funds hidden from the public, city legislators called of the transaction.

Maine's facial recognition law sh bipartisan support for protectin privacy

Alison Beyea, Michael Kebede / 3:14 PM EDT • July 20, 2021

Change Public Policy



Dr. Latanya Sweeney

While serving as the Chief Technologist at the Federal Trade Commission, Latanya Sweeney shaped consumer protections related to AI. Her topics included:

- ◆ Racial bias in online advertising
- ◆ Mapping health care data flows
- ◆ In-app purchases and young children
- ◆ Cell phone data privacy

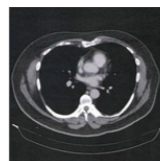
◆

Educate

Educate people about the role that AI is already playing in their lives.

People can't make informed decisions about how technology **should be used** if they don't understand **how it is being used**.

Millions of Americans' Medical Images and Data Are Available on the Internet. Anyone Can Take a Peek.



Hundreds of computer servers worldwide that store patient X-rays and MRIs are so insecure that anyone with a web browser or a few lines of computer code can view patient records. One expert warned about it for years.

How to Wrestle Your Data From Data Brokers, Silicon Valley — and Cambridge Analytica

It can be done but it's not easy.

How Your Shadow Credit Score Could Decide Whether You Get an Apartment

The vast tenant screening industry is subject to less regulation than credit scoring agencies, even though experts warn that algorithms could introduce racial or other illegal biases that can prevent people from getting housing.

Aggression Detectors: The Unproven, Invasive Surveillance Technology Schools Are Using to Monitor Students



In response to mass shootings, some schools and hospitals are installing microphones equipped with algorithms. The devices purport to identify stress and anger before violence erupts. Our testing found them less than reliable.