CS 232:
Artificial Intelligence

Fall 2023

Prof. Carolyn Anderson

Wellesley College

Announcements

- Francesca Lucchetti will be giving a guest lecture in CS 333 today.

- My remaining help hours:

  - Friday: 3:30-4:30

  - next Monday: 3:30-5

  - next Thursday: 3-4

Study Recruitment

Liz Maylin, class of 2024:

I am looking to hold short focus group sessions with students who are regular Tik-Tok users. The purpose of the focus group session is to understand search behavior on Tik-Tok. The session will be approximately 20 minutes long and participants will receive a $5 CVS gift card for participation. Your perspectives as regular Tik-Tok users would contribute significantly to our research!

If you're interested, please fill out this interest form. If you have any questions, please don't hesitate to reach out to me at ld102@wellesley.edu

AI Art Generation Competition
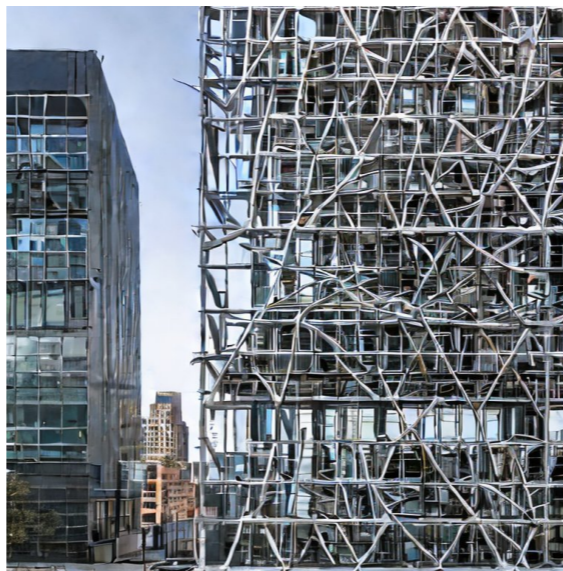
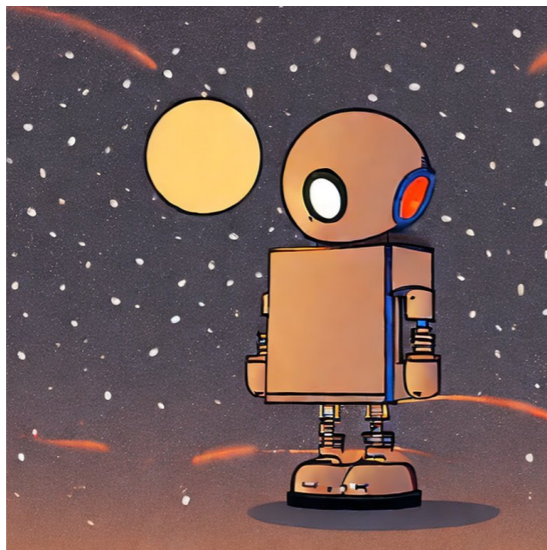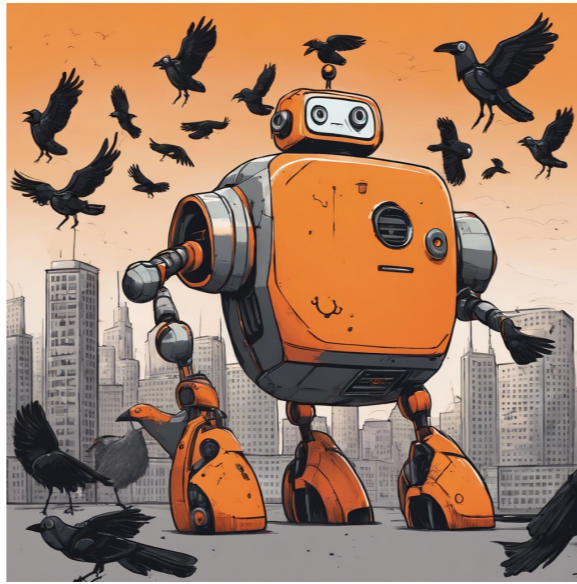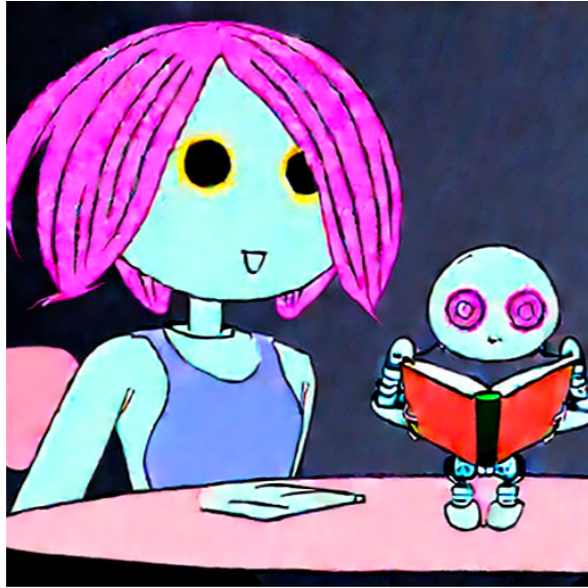# Best Image In A Particular Aesthetic
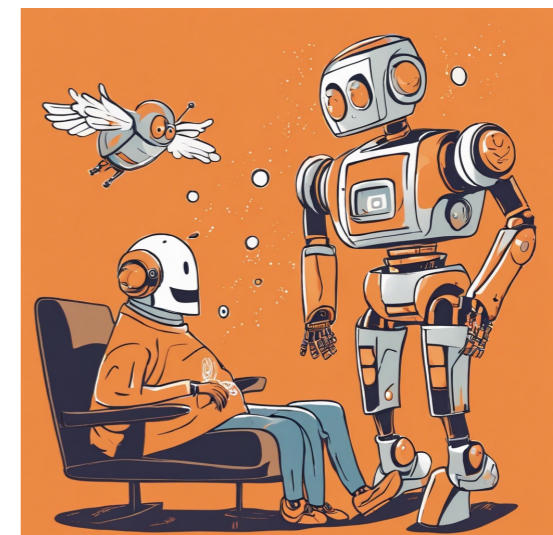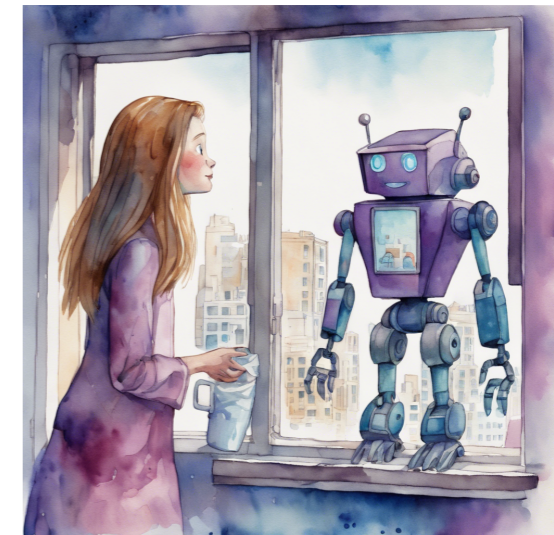
Best Image In A Particular Aesthetic

Winner: Johanna Lee

cyberpunk, impressionist style, dystopian, robots everywhere, animation, nighttime lights, street

# Best Representation Of The Story's Atmosphere

Best Representation Of The Story's Atmosphere

Winner: Sana Khan

Create an evocative and visually compelling illustration depicting a key scene from 'When Robot and Crow Saved East St. Louis.' Capture the collaboration and dynamic interaction between Robot and Crow as they navigate challenges in East St. Louis. Highlight the setting's atmosphere, the characters' expressions, and any notable elements that convey the essence of their heroic endeavors in the narrative
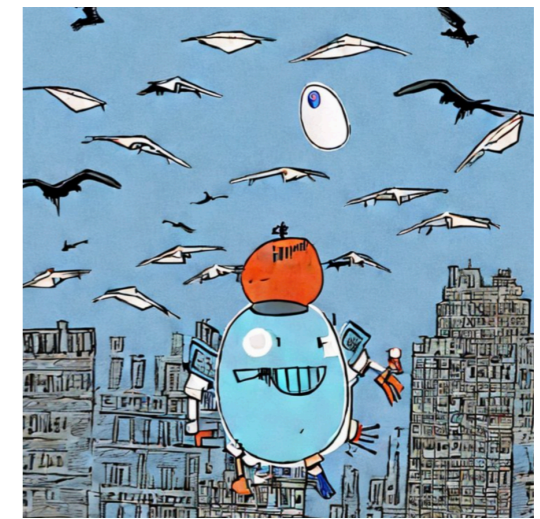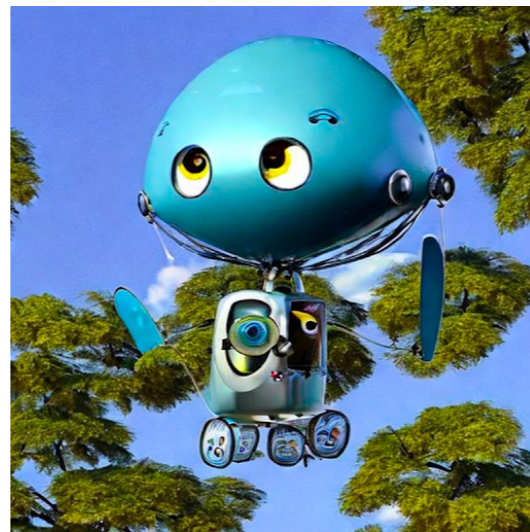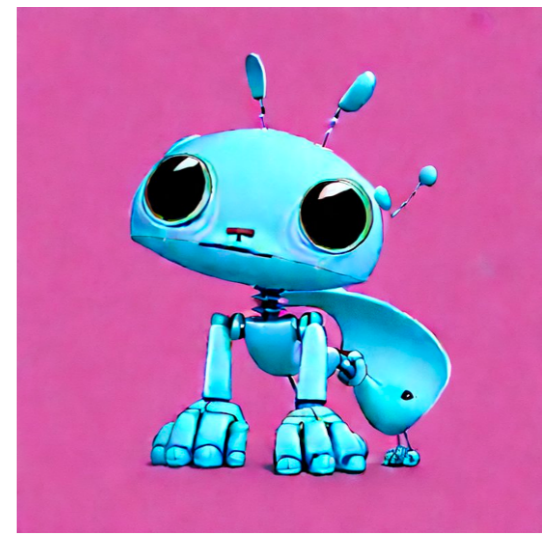
Best Background Illustration
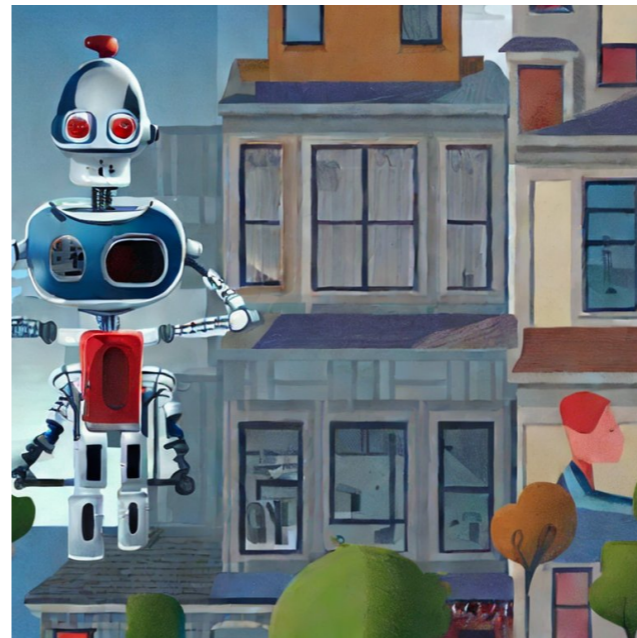
Best Background Illustration

Winner: Emily Hou

cul-de-sac street view with many mansions, oak trees, warm dreamy lighting, smart cars

Best Illustration Of A Character

Best Illustration Of A Character

Winner: Johanna Lee

cute oval shape robot, with colorful propeller on head, cartoon style

Best Illustration Of An Event In The Story
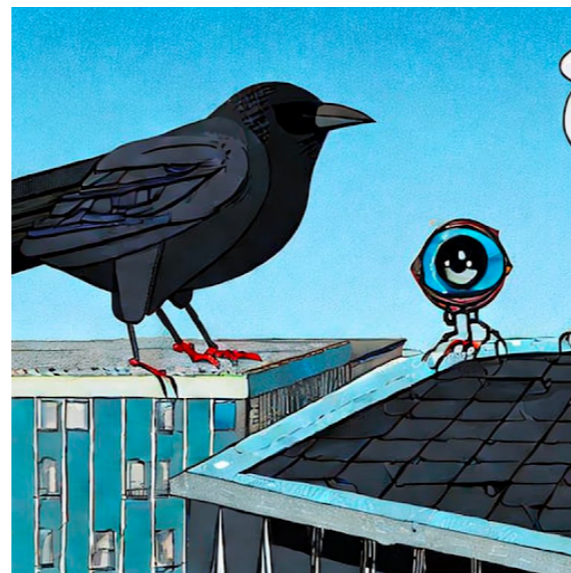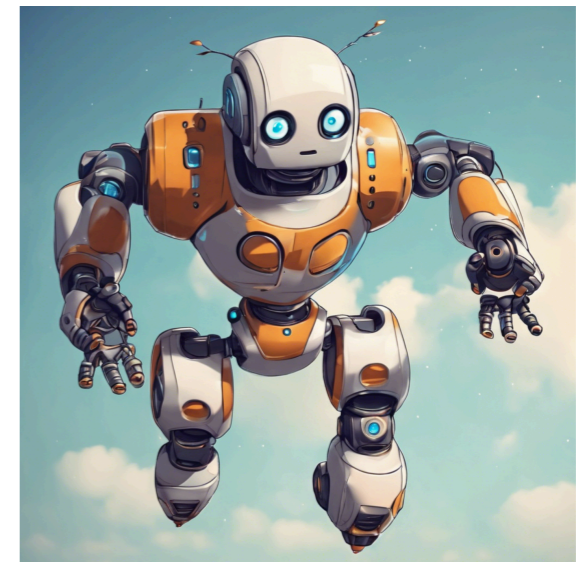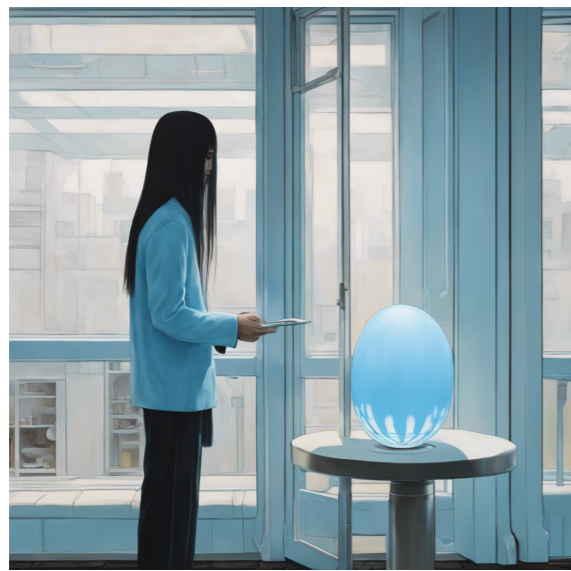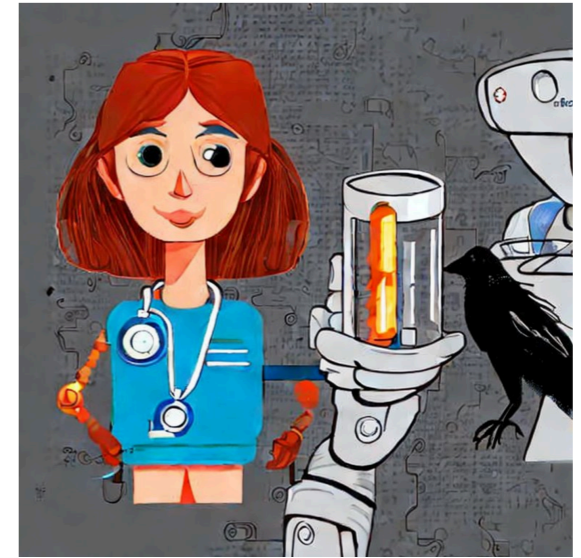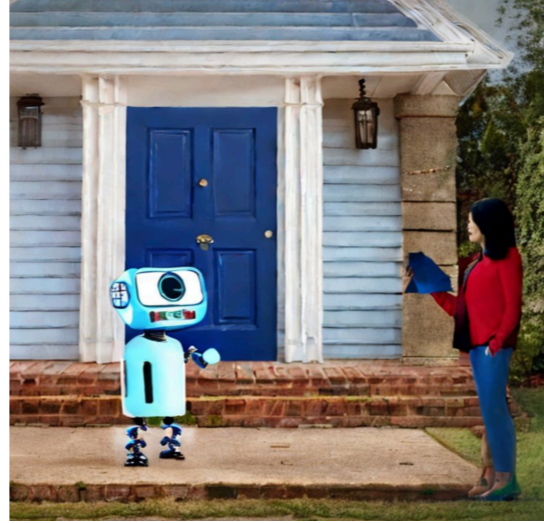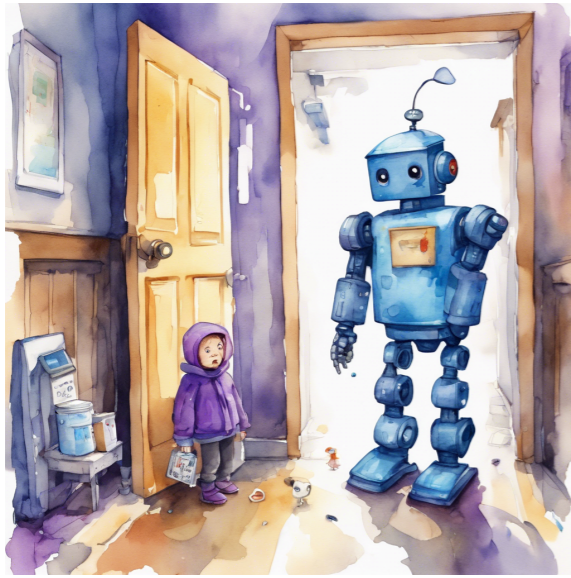
Best Illustration Of An Event In The Story

Winner: Hsi-Er Liu

A cute robot (blue body, purple mouth, beanie on top) is going through a door to a hallway with small rooms on the side with a crow perched on the railings behind the robot. Inside some of the rooms is bedding. Some people crying, children and adults, lots of clothing all around, lots of plastic containers on the floor, a person in a blanket, a person coughing. Robot gets the tissues ready to test. Modern city. Watercolor.
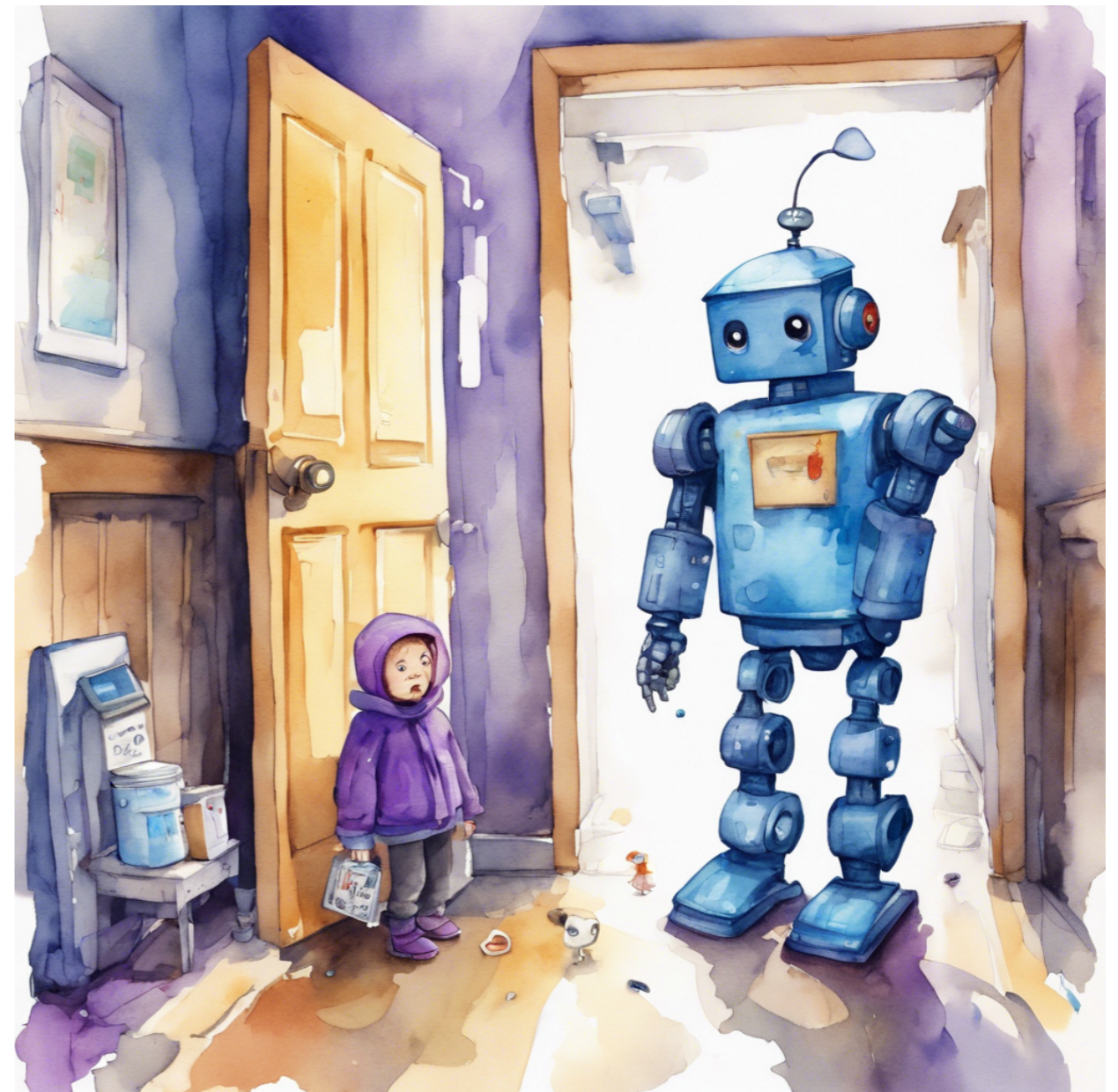
Best Foreground Illustration
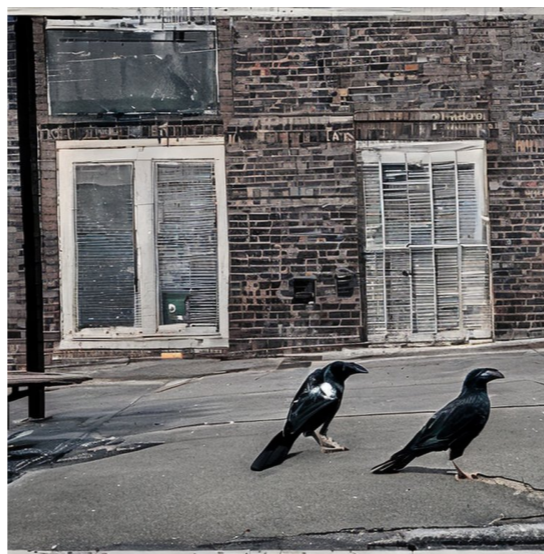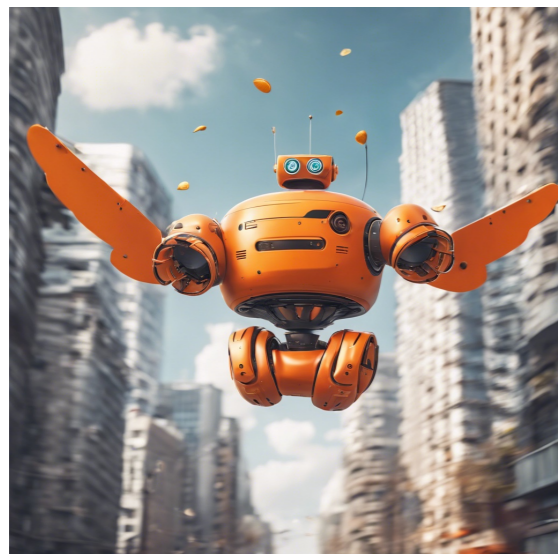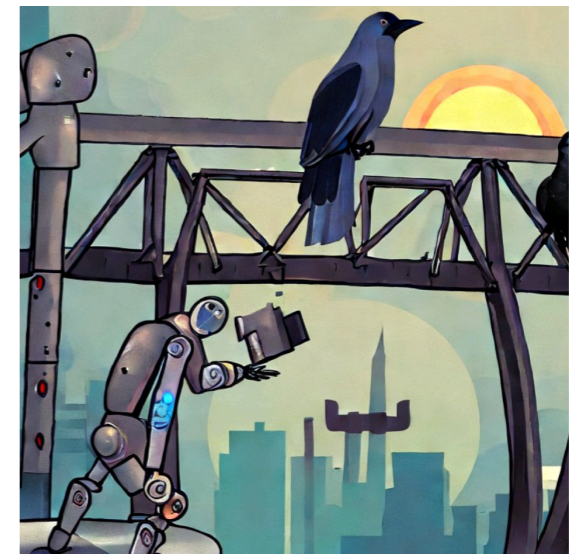
Best Foreground Illustration
Winner: Fridah Ntika

Best Photo-Realistic Illustration
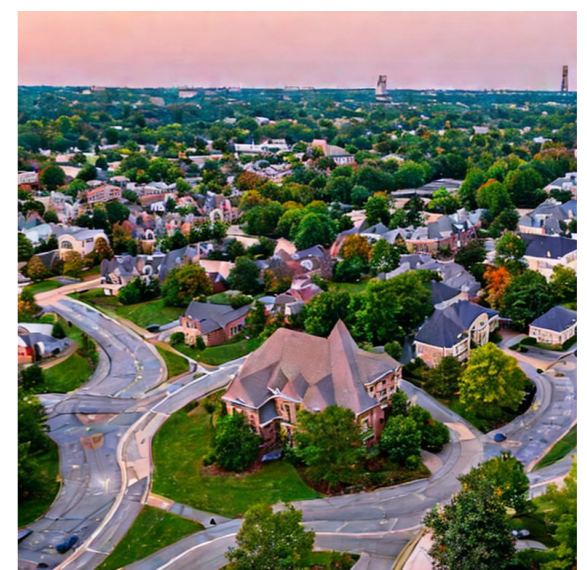
Best Photo-Realistic Illustration

Winner: Bella Furst

Tent city under freeway, cloudy day

Best Tepresentation Of Technology
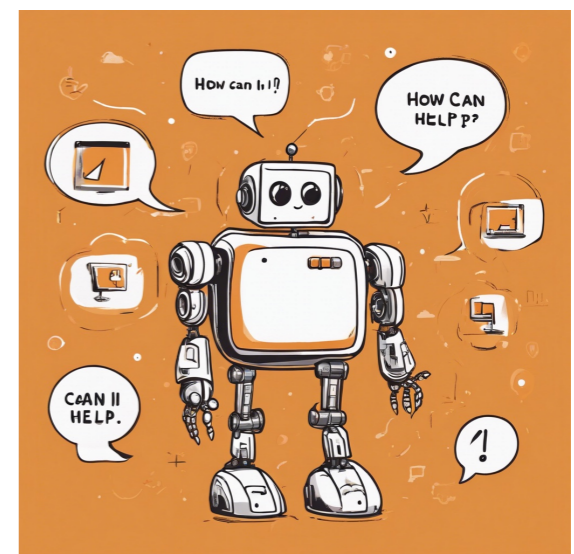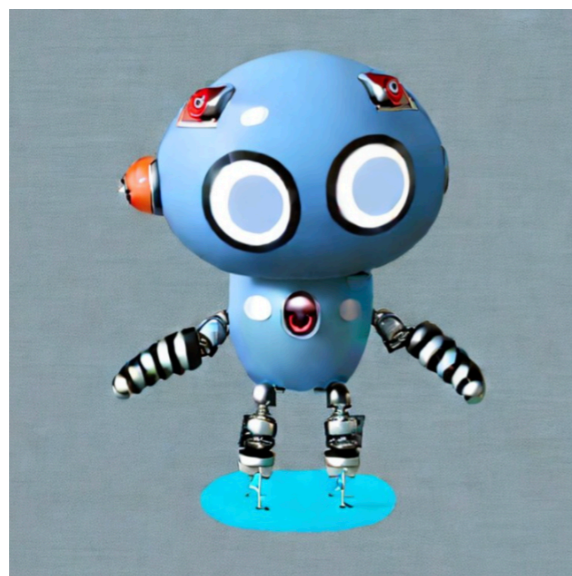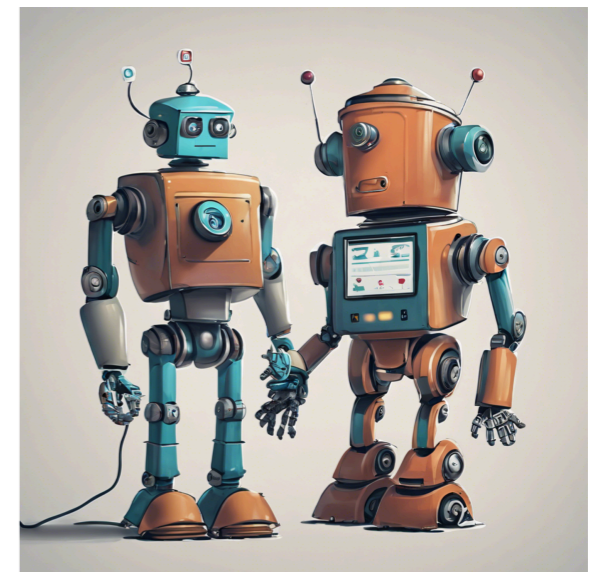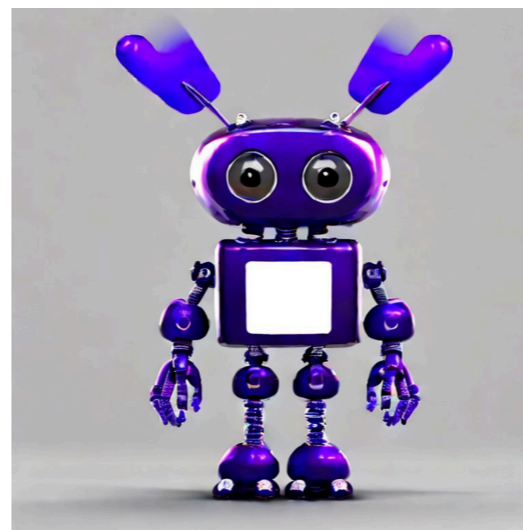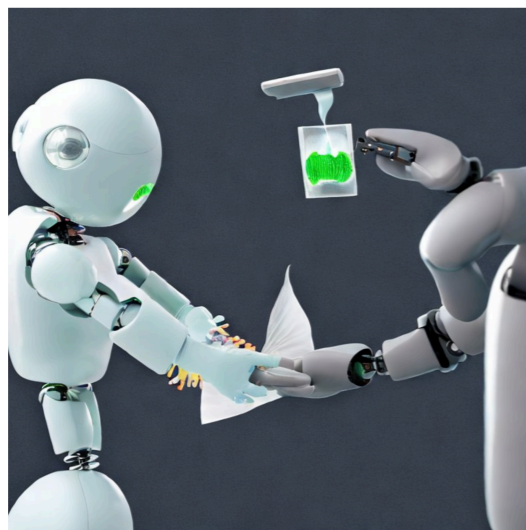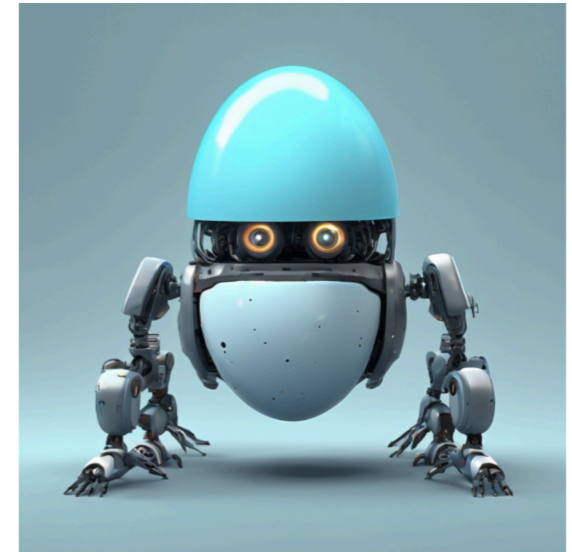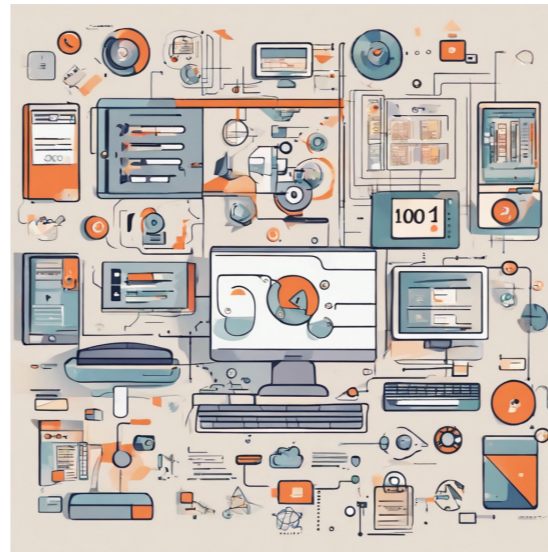
Best Representation of Technology

Winners: Maha Khan and Annie Chen

a real life wide shot photograph of 4 persons, robot and a crow all happily working together, background lighting and robot is in the spotlight. atmosphere is hopeful and there is unity amongst people

robots being helpful
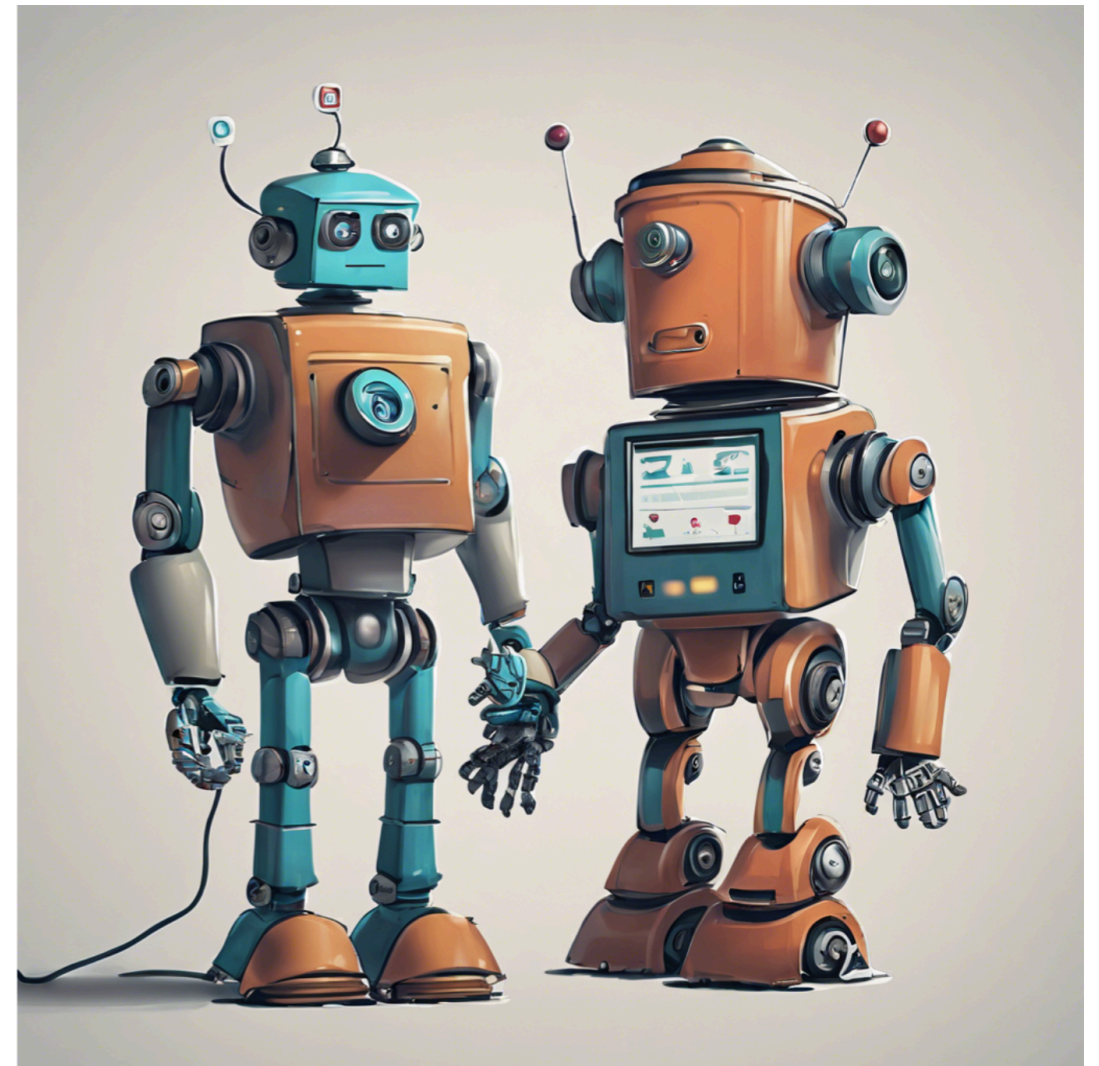
Most Creative Prompting Technique

Cartoon robot whose outward appearance is a mix of symmetrical poodle and cartoon turtle. Has a short, soft snout. Has a purple, dog-turtle mouth and cute cheeks. Overall cute. Consists of a ventral service trunk that is filled with sterile sheets that can be pulled out with a gripper contraption. Has a cute face and legs which makes the Robot seem friendly. Disney style. Negative Prompt: ugly box, scary, photograph

a human girl with short hair, woven into pink extensions in bright stripped clothes, writing notes on her notebook. Next to her is a cute, blue, small, egg-shaped robot with a tiny rainbow propeller on the top. A tiny crow is on top of the trash can. Ghibli ANIMATION style, night, cyberpunk

A cute robot (blue body, purple mouth, beanie on top) is on the ledge, smiling, steam going out from the top of the robot, confused about its identity, words "friendship" and "representation" pop out. A young person (short and pink hair) is looking through chemistry notes, complicated molecular structures, many textbooks. A crow nearby looking at the robot and the young person. Modern city. Watercolor.
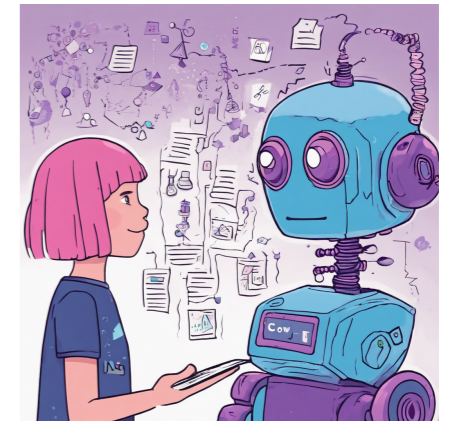
Generate a captivating illustration portraying the exhilarating moment from 'When Robot and Crow Saved East St. Louis' where the Robot and Crow are soaring through the sky. Showcase the tandem flight, emphasizing their unique characteristics, the cityscape of East St. Louis below, and the emotions conveyed during this heroic and dynamic aerial scene. Capture the sense of unity, freedom, and purpose in their flight

pale blue egg shaped robot perched on roof of History Museum in Forest Park in st louis missouri

Half of the screen on the left is at night, stars in the sky, 3 crows perched on the tree. On the right half of the screen is during the day, bright sun in the sky, crows bringing back food and some crows flying off in a group in the distance. Right in the middle is the back of a cute robot (blue body, purple mouth, propeller on top) looking at the crows with a dialogue box outside of the robot that says "Food here!" and "Enemy". Modern city. Watercolor.

Generate an image depicting a little girl arguing with a cute crow because they cannot understand each other. A cute pale blue egg-shaped small drone with a propeller beanie on top is in the air and watching them. The style should be futuristic

Most Creative Prompting Technique Winner: Hsi-Er Liu

Half of the screen on the left is at night, stars in the sky, 3 crows perched on the tree. On the right half of the screen is during the day, bright sun in the sky, crows bringing back food and some crows flying off in a group in the distance. Right in the middle is the back of a cute robot (blue body, purple mouth, propeller on top) looking at the crows with a dialogue box outside of the robot that says "Food here!" and "Enemy". Modern city. Watercolor.

Most Surprising Prompt

In the old strip mall with gleaming counters and windows duct taped. On three chairs, a cute blue robot with a purple mouth and propeller on the top, a crow, and a person with short pink hair are sitting down. The person has a tube of medicine on one hand and has a black feather on the other hand, the person is sticking the black feather on top of the robot. The robot, crow, and person are happy, laughing, smiling. In the distant background, a person with purple hair and a prosthetic arm and three other people are putting things in test tubes and packaging syringes. Modern city. Watercolor.

Create a wideview high resolution image of two beings, Robot and crow bird, looking down the decaying streets of East St. Louis
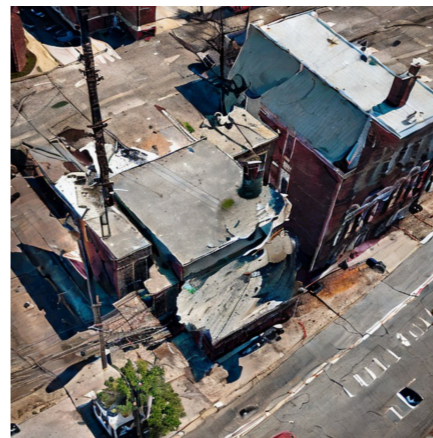
23 flying crows birds in a blue sky and pale robin blue robot titanium oval flying egg with propellers on top of head flying
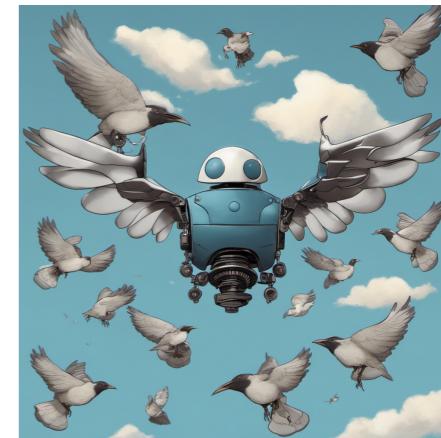
show me a photoshopped picture of a human with short hair woven into pink extensions holding a kindle. The human is coughing on a blue robot
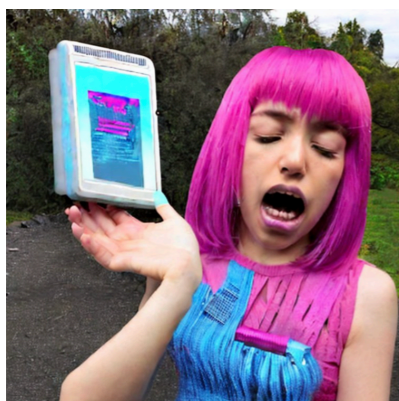
robot, crows, human, talking

1 cartoon crow bird, 1 cartoon robot, 1 cartoon little girl with short hair Negative prompt-scary

Depict a side angle shot of a cute pale blue egg-shaped small drone with a propeller beanie on top talking to a human
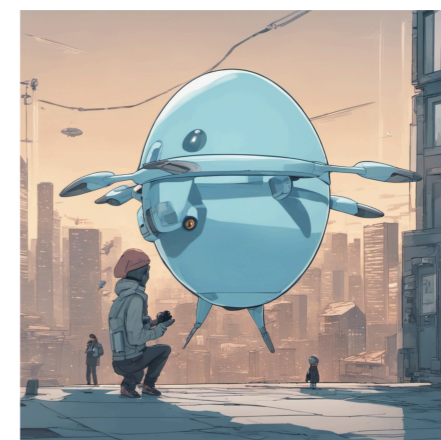
Most Surprising Prompt

Winner: Aeka Tomita

show me a photoshopped picture of a human with short hair woven into pink extensions holding a kindle. The human is coughing on a blue robot
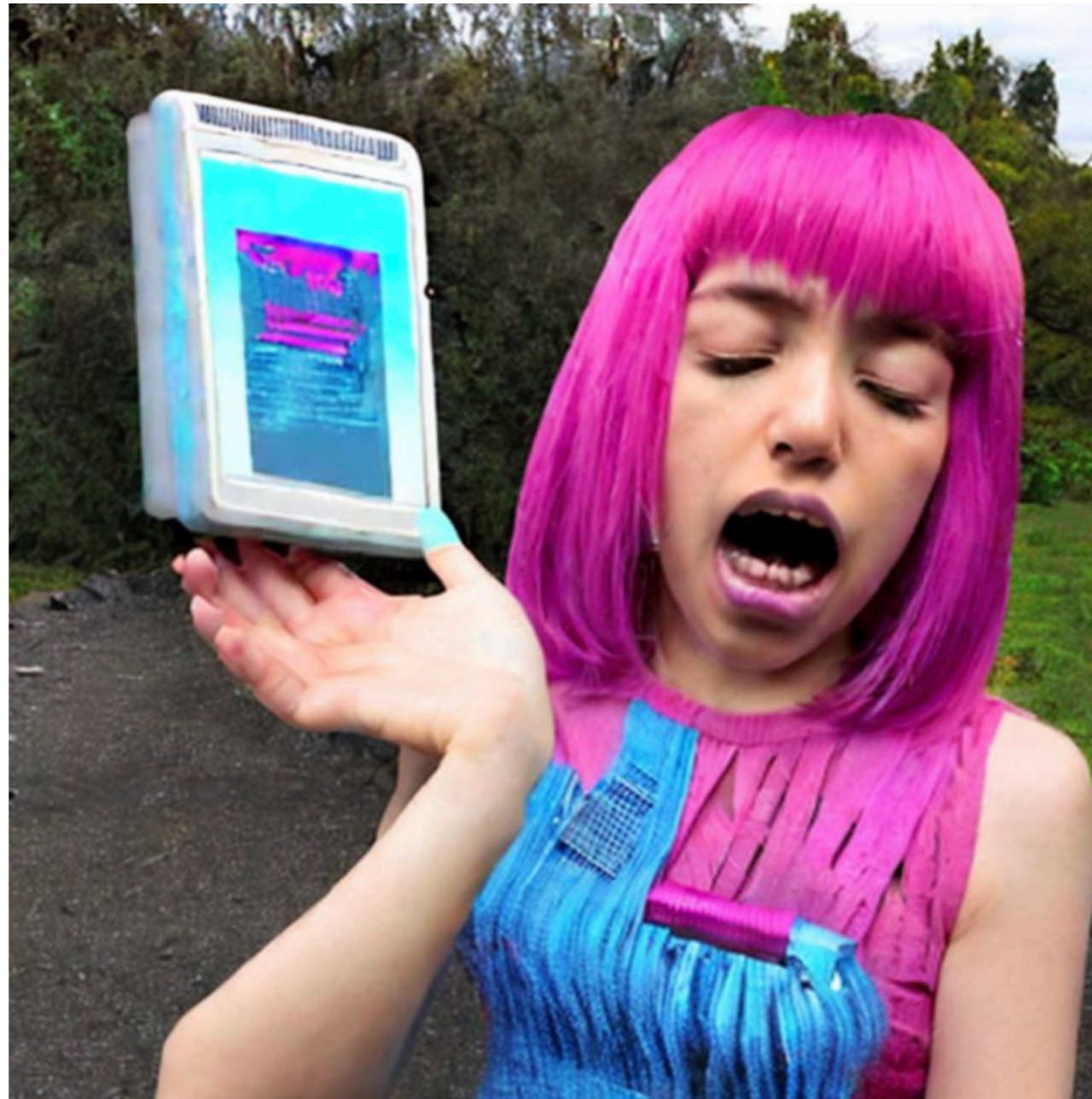
Harm Reduction

Scenario: Art Generation

Stable Diffusion releases an image generation model trained on data scraped from the internet.

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.

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."

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.

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).

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 <u>disclaimer </u>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

Algorithmic Justice League protests bias in voice AI and media coverage

Be Heard

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
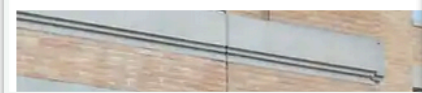
NEWS | 19 June 2020

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.

Protest

Senators Protest a Health Algorithm Biased Against Black People

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

MOTHERBOARD
TECH BY VICE

Schools Are Abandoning Invasive Proctoring Software After Student Backlash

N.Y.P.D. Robot Dog's Run Is Cut Short After Fierce Backlash
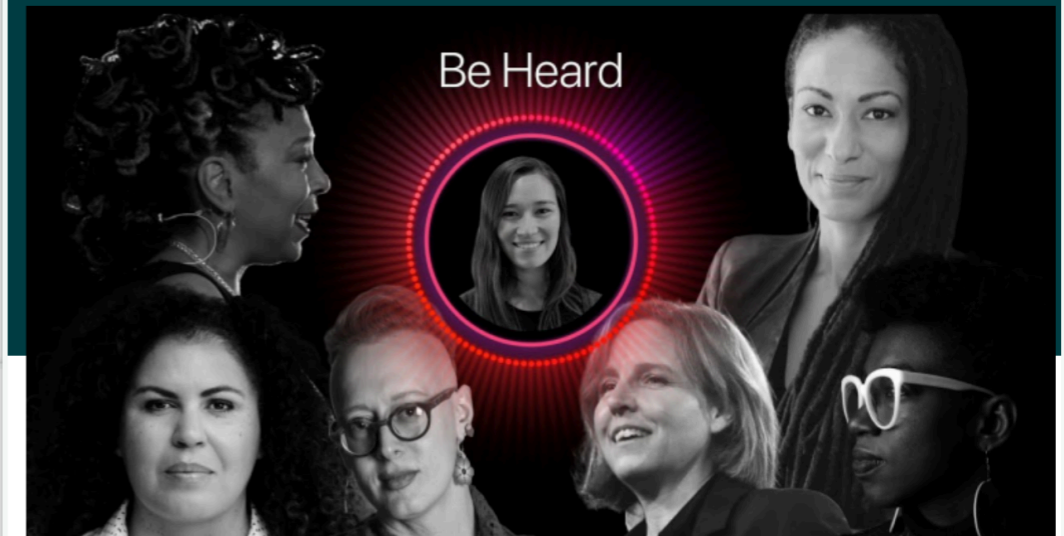
The Police Department will return the device earlier than planned after critics seized on it as a dystopian example of overly aggressive policing.
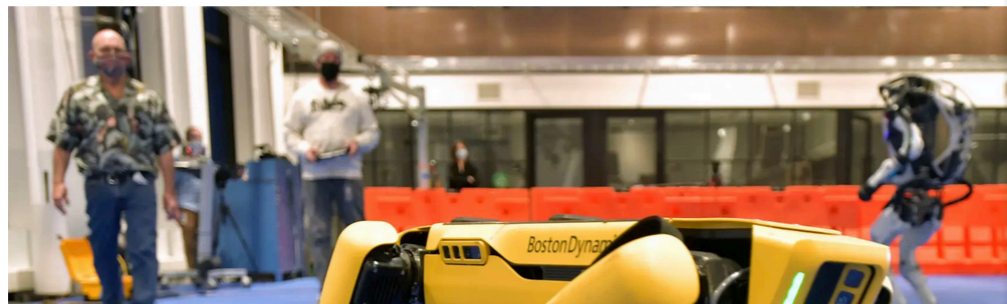
COVER STORIES

Was It Racist? Santa Cruz Bans Predictive Policing

Mayor Justin Cummings leads charge on law enforcement reform

SANTA CRUZ MAYOR JUS...
BROUGHT FORW...
PREDICTIVE POLICING BA...

Give this article

BostonDynam...

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   8 M

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

Massachusetts Pioneers Rules For Police Use Of Facial Recognition

May 7, 2021 · 6:00 AM ET

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 strips federal support for state and local

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 protecti 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

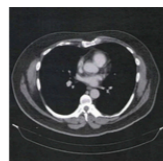✦ In-app purchases and young children

✦ 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.