

COMP322: Computer Vision & Image Processing

What is Computer Vision and Image Processing?

Modified by Hong Xu,
mostly taken from
slides by Betsy
Sanders, in turn
modified from James
Tompkin's class.



Image Credit: Pixar Animation Studios

Prof Hong Xu



Rhodes College



Introductions

1. Preferred Name (optional: pronouns)
 2. Sophomore/junior/senior
 3. What did you do this summer?
 4. Are humans alone in the universe?
5. What sci fi or fantasy tech/magic/power would you wish existed, if any?

Jitendra Malik, UC Berkeley

Three 'R's of Computer Vision



“The classic problems of computational vision:
reconstruction
recognition
(re)organization.”

Pixar's Wall-E



The Three R's of Computer (Wall-E) Vision



The Three R's of Computer (Wall-E) Vision



The Three R's of Computer (Wall-E) Vision

(Re-)Organization



Have you ever used computer vision?

How? Where?

Have you ever used computer vision?

How? Where?

Is it Reconstruction? Recognition? (Re)organization?

Think-Pair-Share

Smartphones: QR codes, computational photography (Android Lens Blur, iPhone Portrait Mode), panorama construction (Google Photo Spheres), face detection, expression detection (smile), face filters/tracking, FaceID (iPhone), Night Sight (Pixel), iPhone 12 Pro (LiDAR), body workout form detection

Smartwatches: Heart rate detection, proximity detection

Security: Fingerprint/iris/face scanning (offices, airports), CCTV monitoring

Laptops/Desktops: Biometrics auto-login (face recognition, 3D)

Web: Image search, Google photos (face recognition, object recognition, scene recognition, geolocalization from vision), Facebook (image captioning), Google maps aerial imaging (image stitching), YouTube (content categorization), Photoshop, PowerPoint (captioning, design suggestions)

Virtual Worlds: VR/AR head tracking (Oculus, HTC Vive), simultaneous localization and mapping, person tracking (Kinect), gesture recognition, virtual try-on, digital humans

Telepresence: Virtual backgrounds (Zoom, Google Meet), webcam person/face following

Media: Visual effects for film/TV, virtual sports replay, semantics-based auto edits

Transportation: Assisted driving (cruise control, self-driving), face tracking/iris dilation for safety

Supermarkets: Cashier-less checkout, theft detection (Walmart), fruits/vegetables sorting, packaging, manufacture

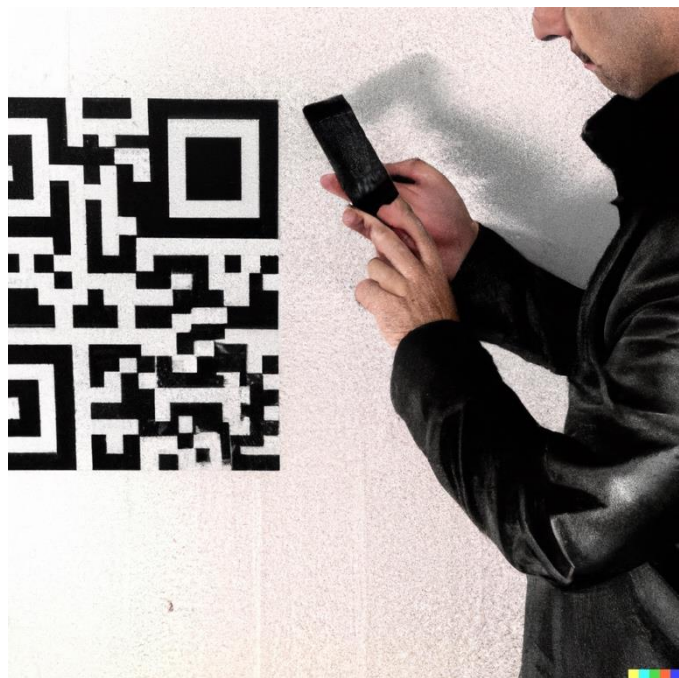
Medical imaging: CAT / MRI reconstruction, assisted diagnosis, automatic pathology, connectomics, endoscopy

Space Exploration: Mar rovers, space telescopes (Hubble, James Webb)

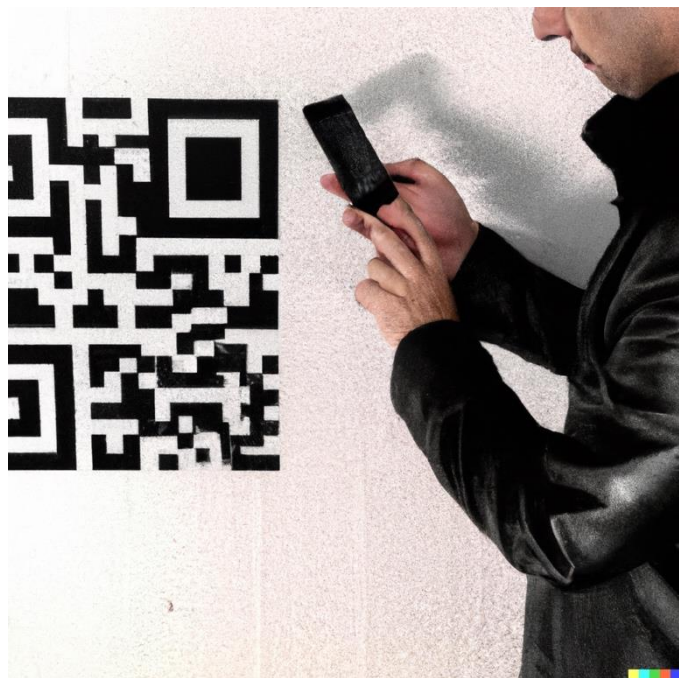
Industry: Vision-based robotics, online shopping (Amazon, Walmart), machine-assisted router (jig), OCR (USPS), ANPR (number plates for tolls), drones

Creative: Photoshop, vision-language models for image generation (Dall-E), video editors

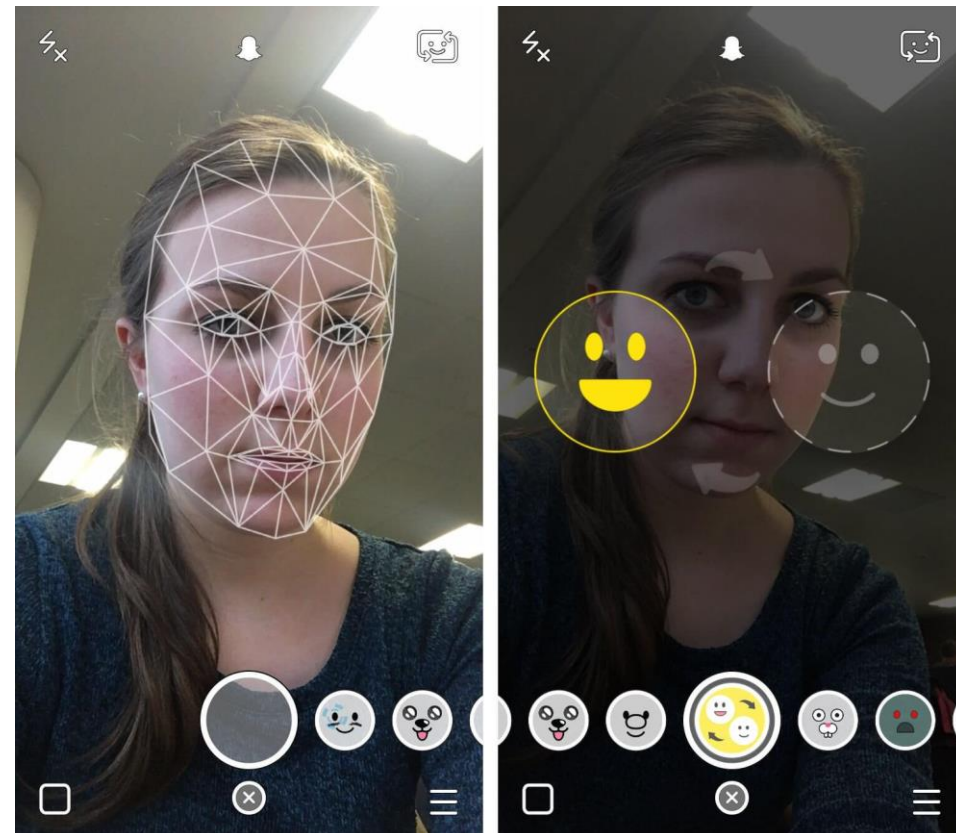
Smartphones: QR Codes, Panorama, Emotion



Smartphones: QR Codes, Panorama, Emotion



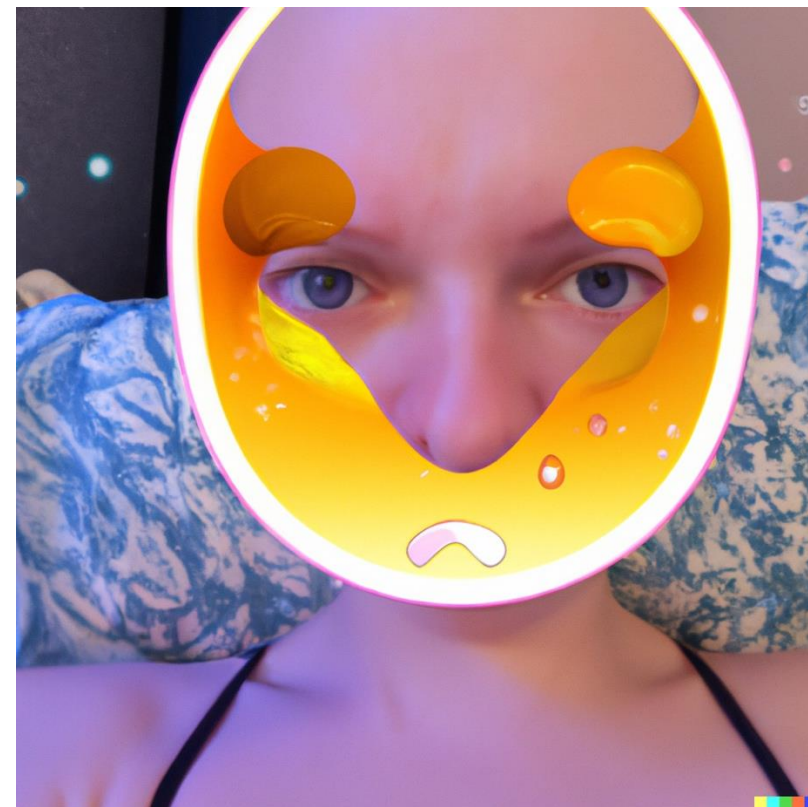
Smartphones: QR Codes, Panorama, Emotion



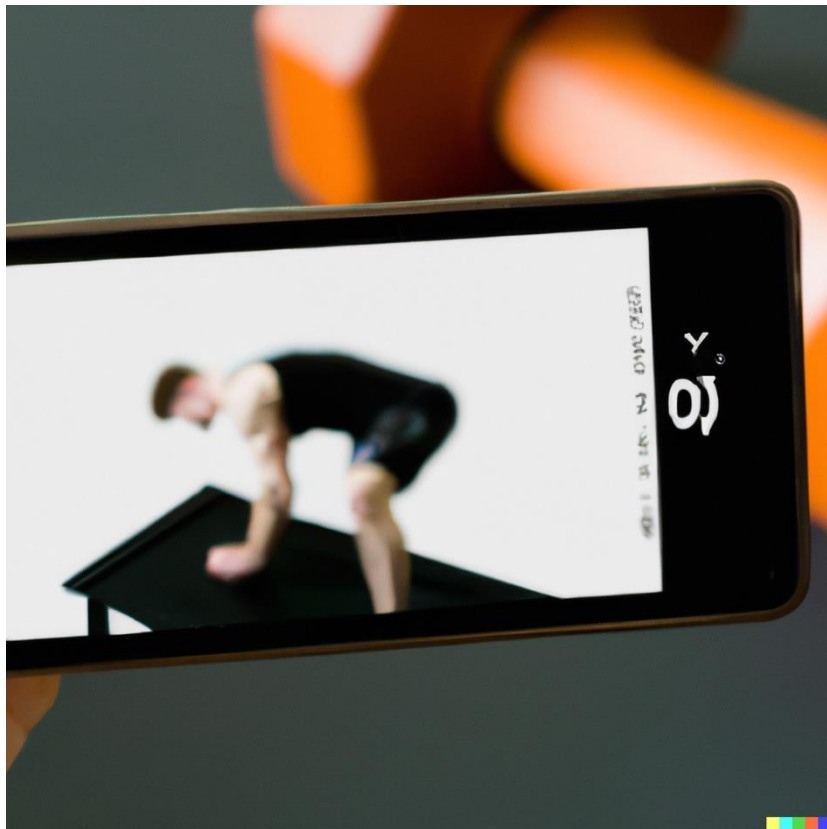
Smartphones: Face Filters



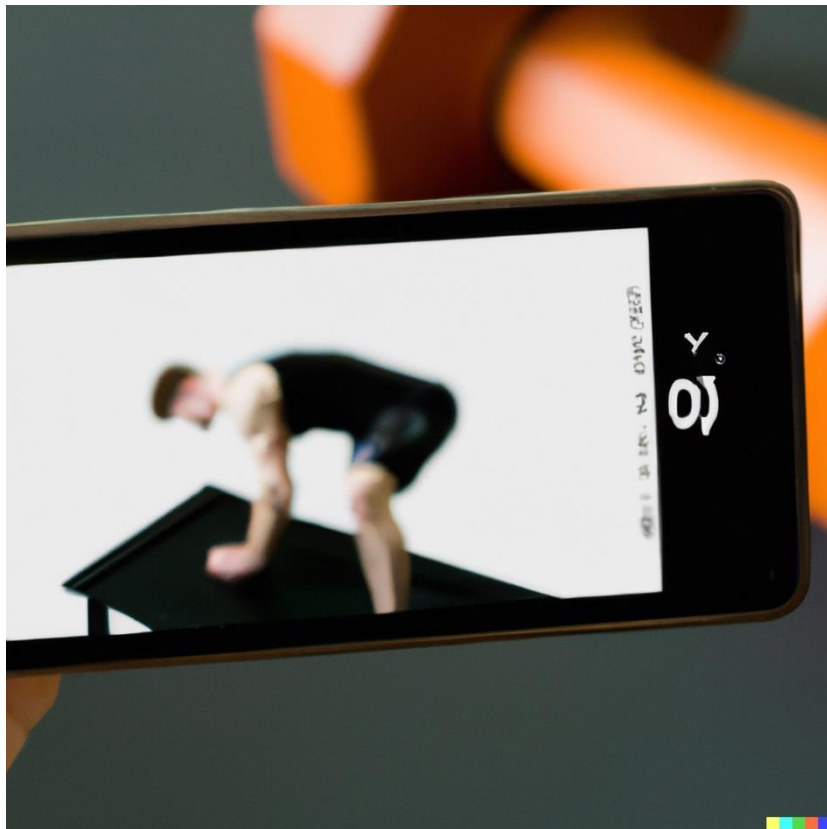
Smartphones: Face Filters



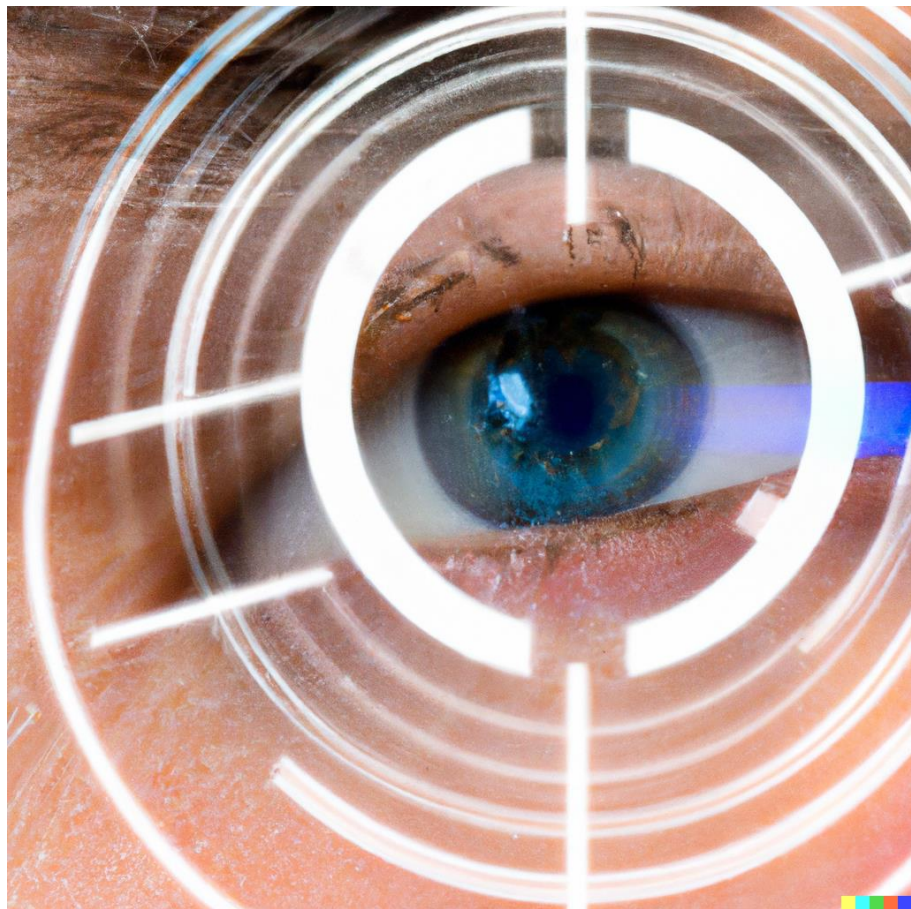
Smartphones: Biometrics, Security



Smartphones: Biometrics, Security

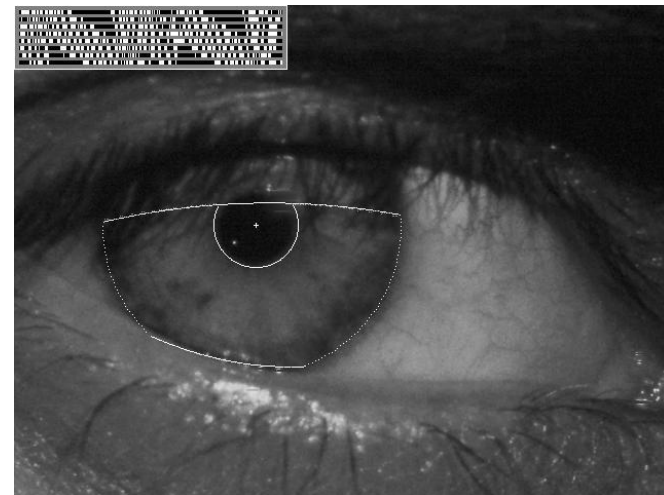
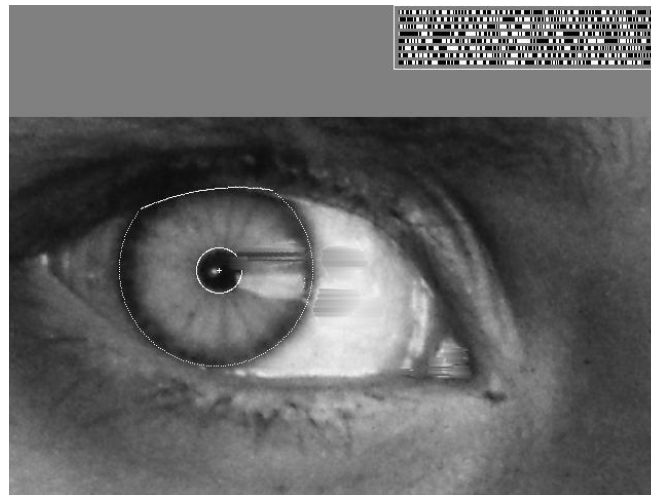


Security | Biometrics

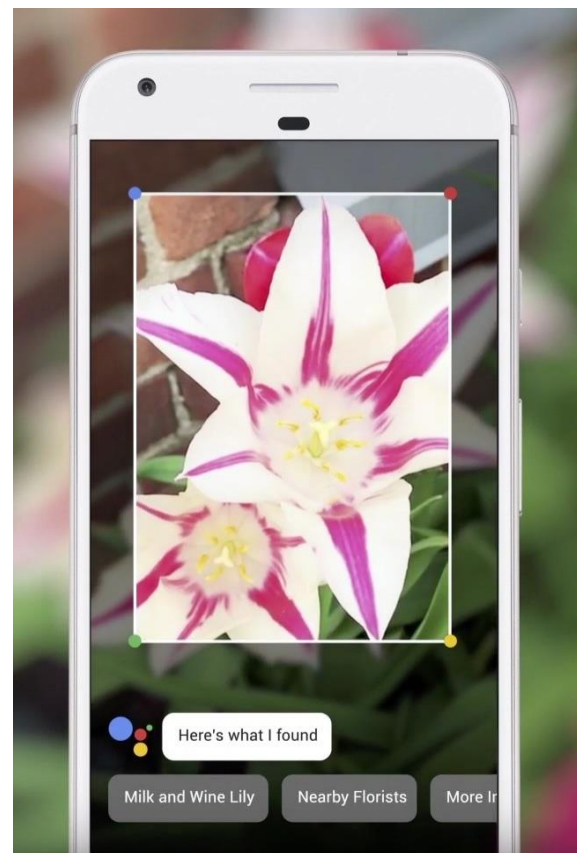
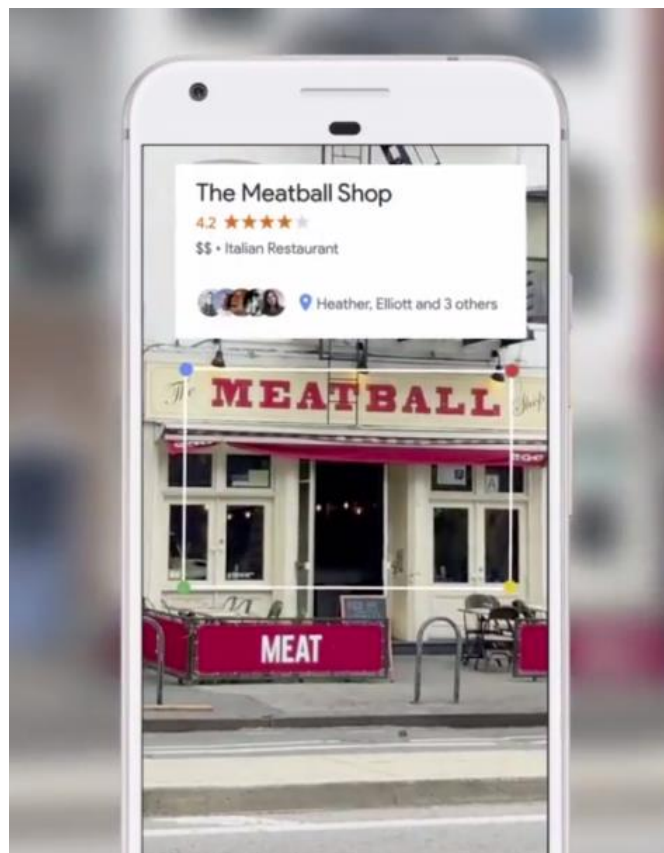


“How the Afghan Girl was Identified by Her Iris Patterns”

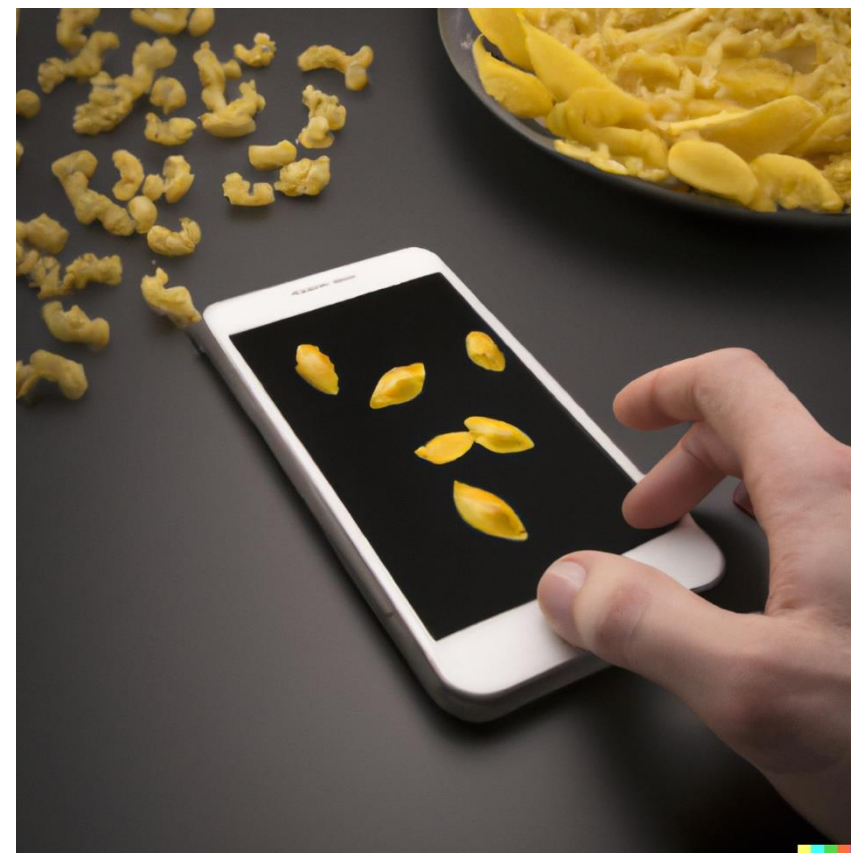
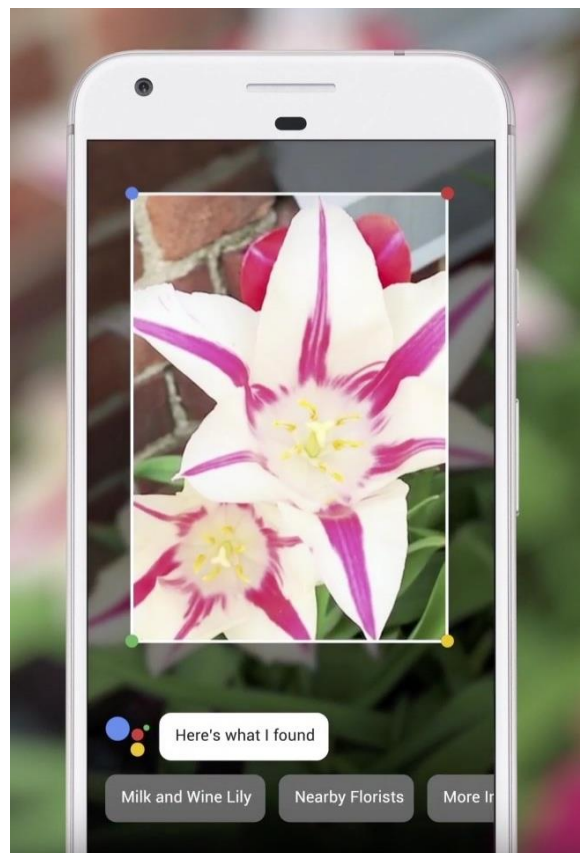
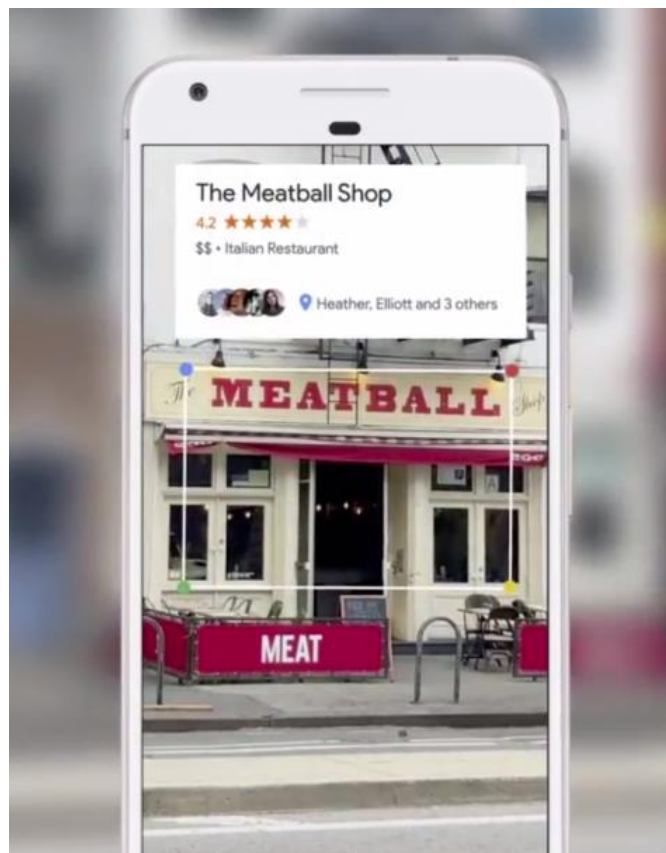
Read the [story](#) ([Wikipedia](#))



Web: Place | Thing Recognition



Web: Place | Thing Recognition



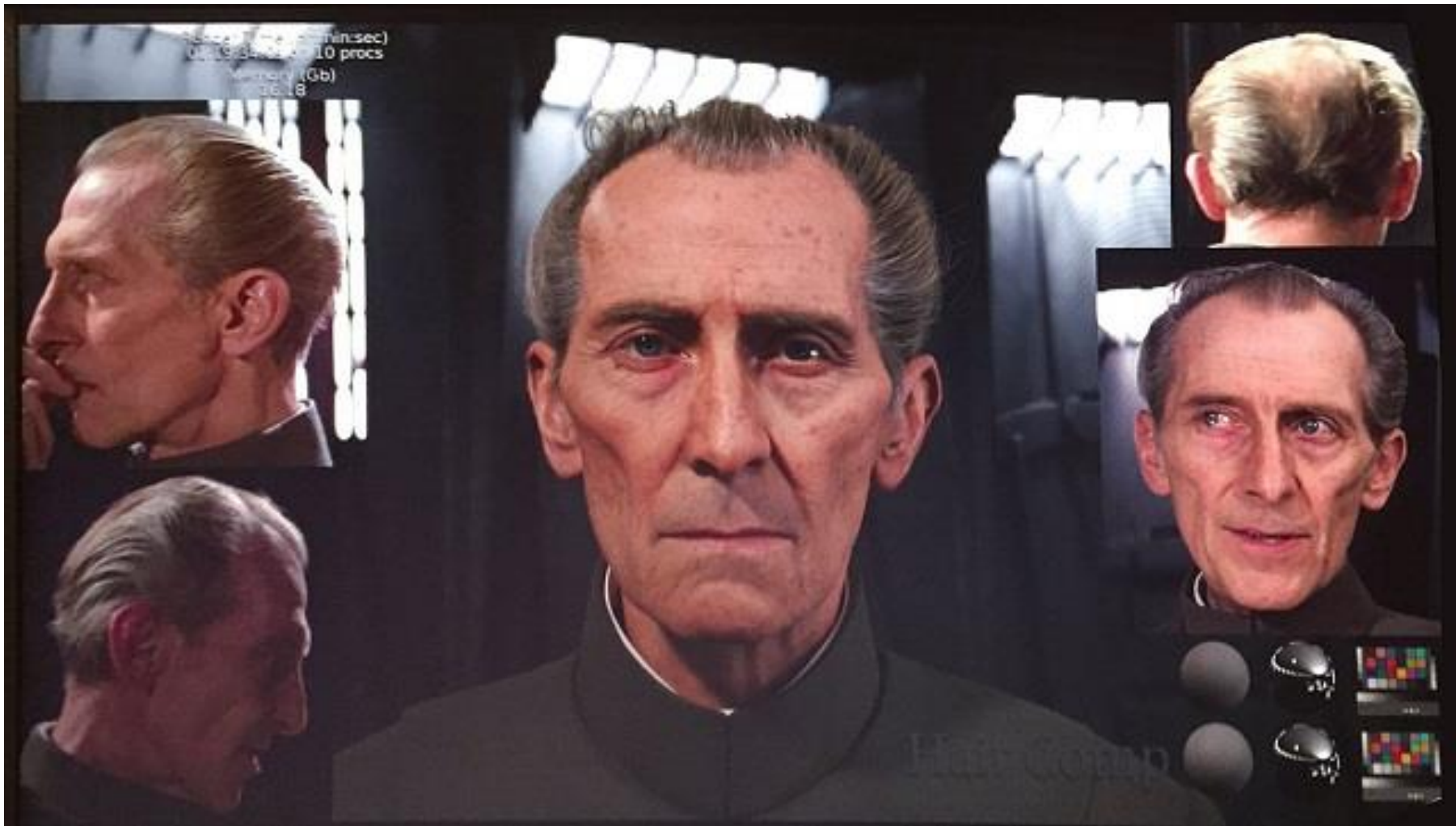
Virtual Worlds: Digital Humans



Virtual Worlds: Digital Humans



Visual Effects



Star Wars. Tarkin played by the late Peter Cushing.

Visual Effects



Dawn of the Planet of the Apes

Transportation: Self-Driving Cars



Supermarkets



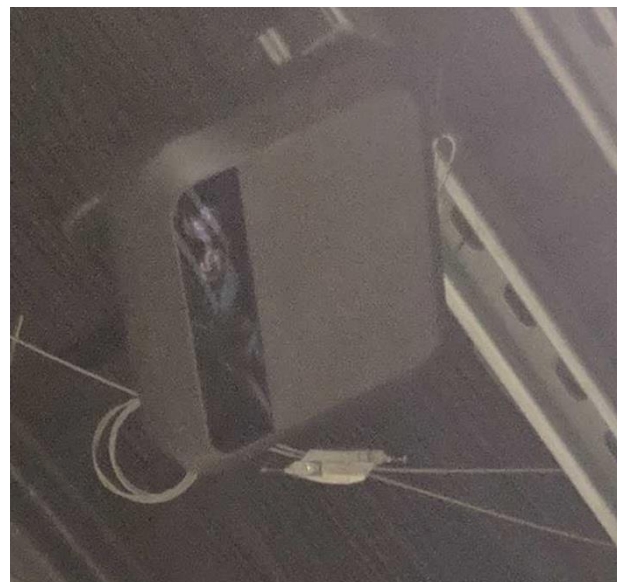
How does it work?

Think-Pair-Share

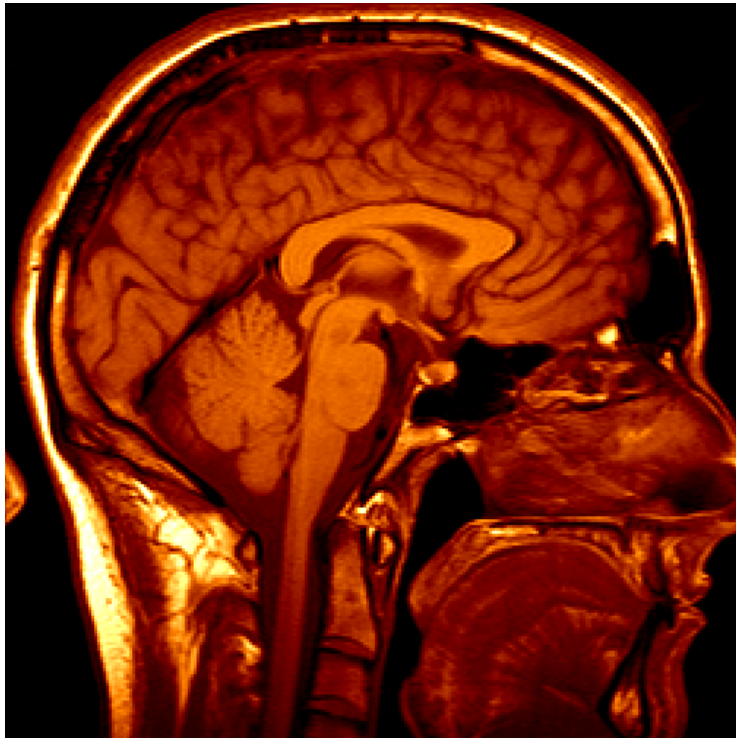


How does it work?





Medical Imaging

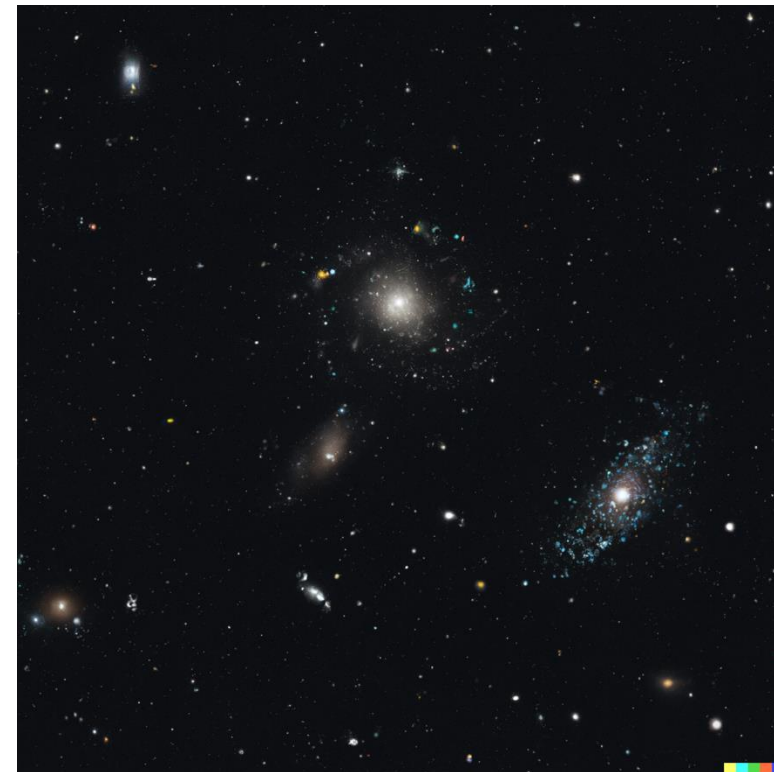
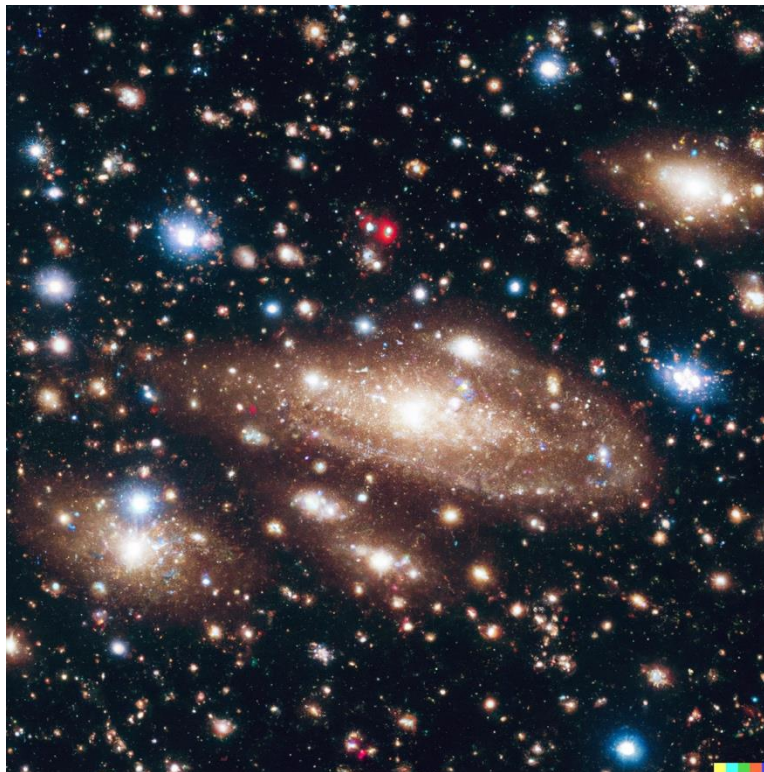
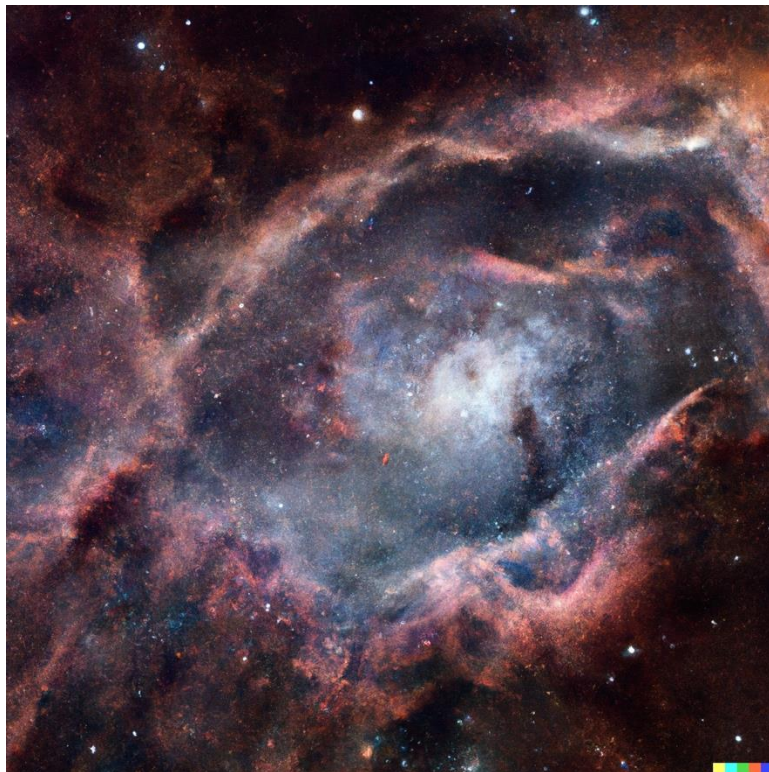


3D imaging
MRI, CT

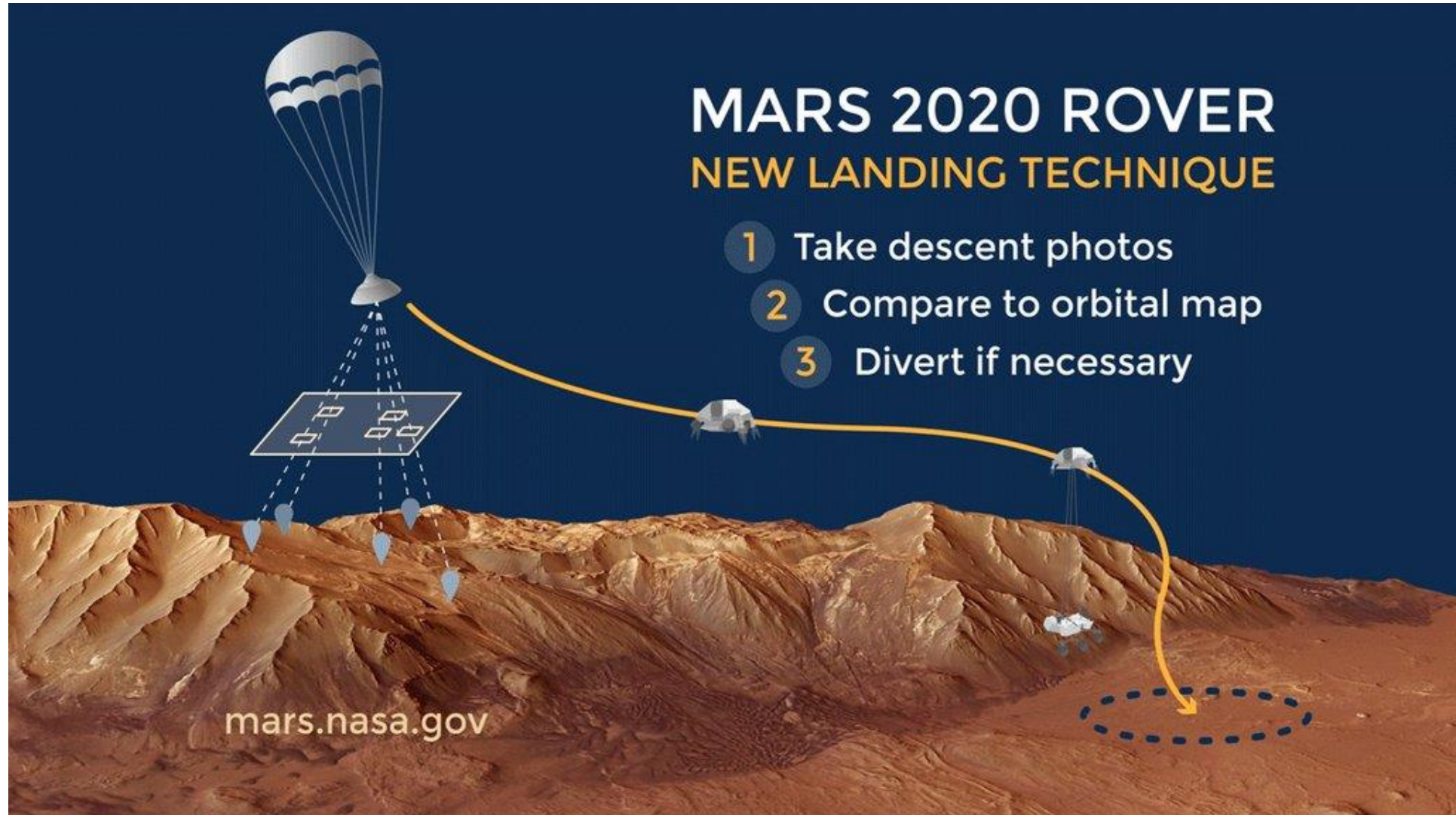


Ultrasound

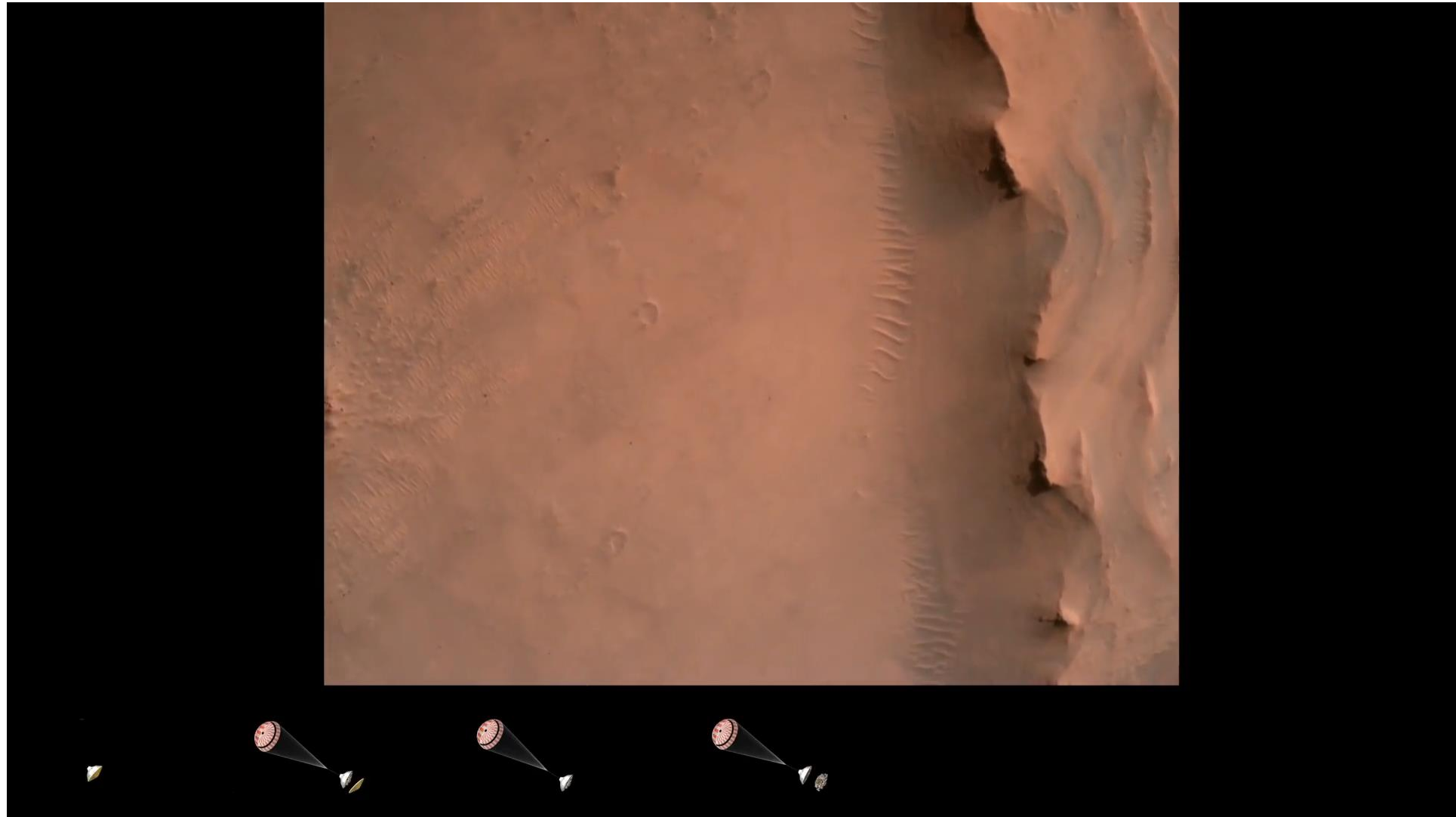
Space Exploration



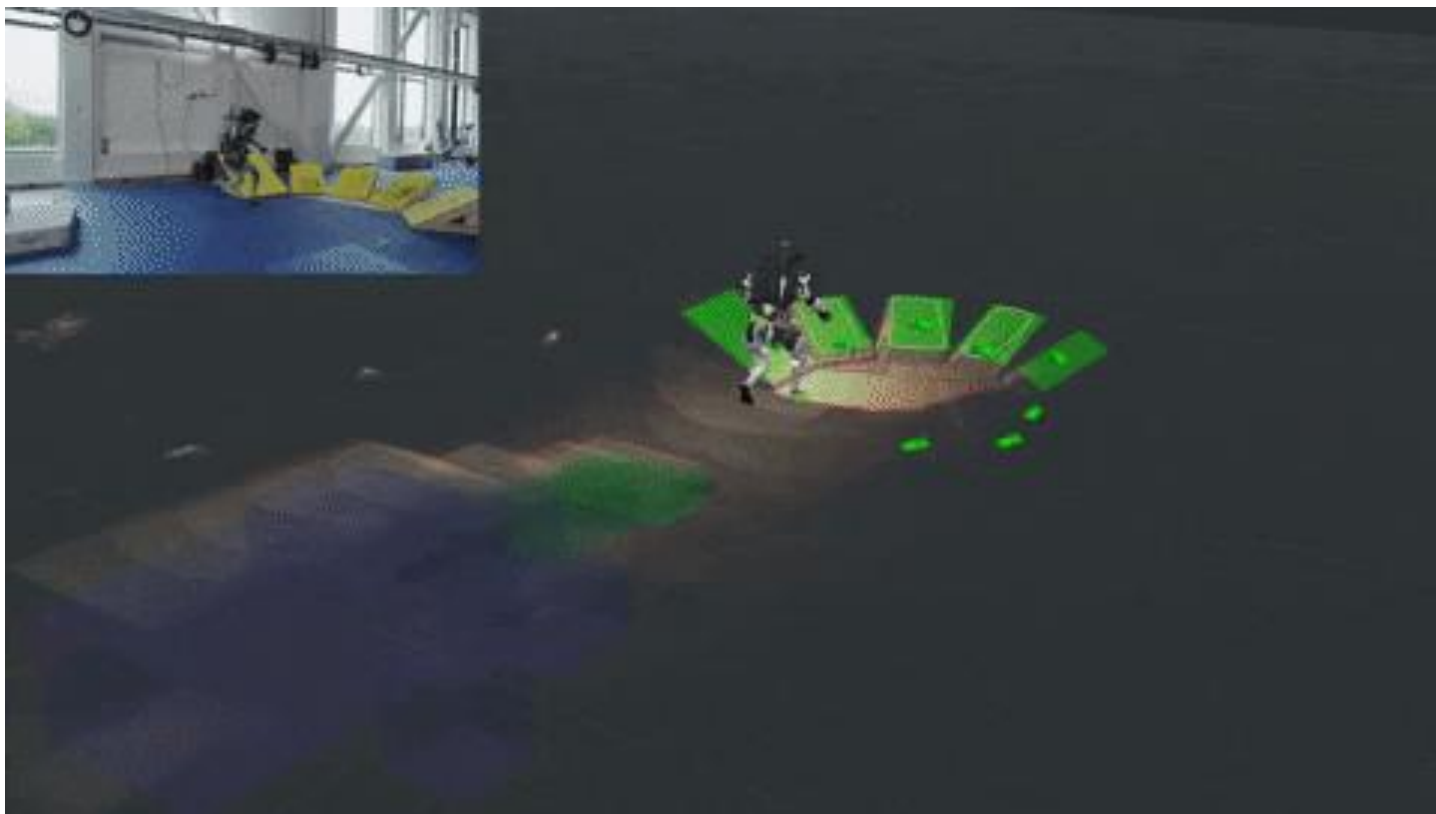
Space Exploration



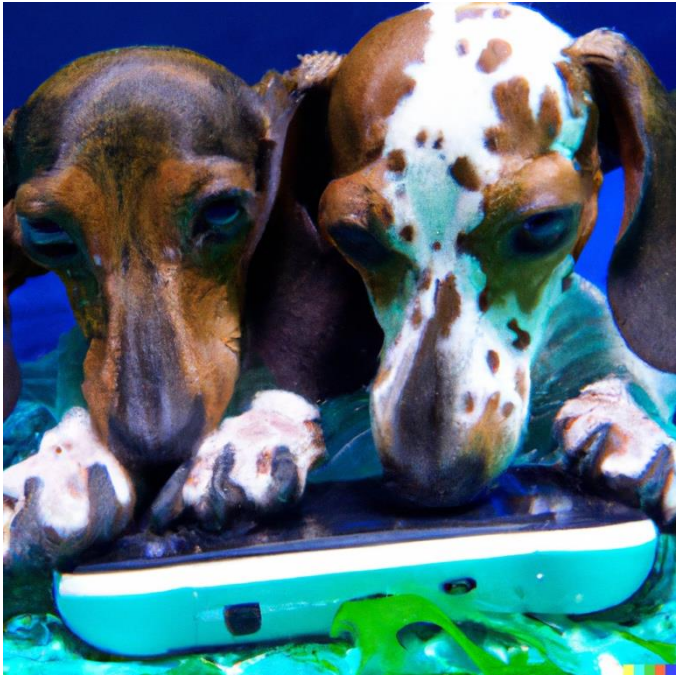
Space Exploration



Industry: Robots

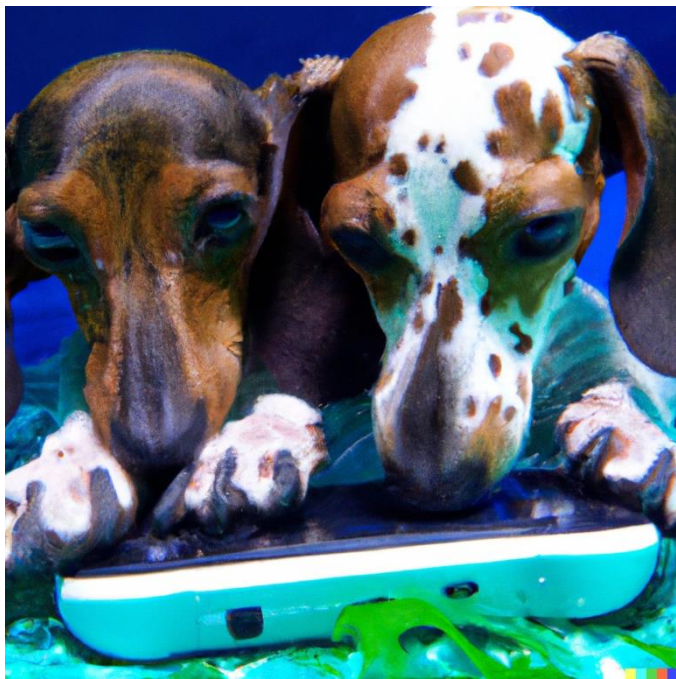


Creative: Text-to-Image Synthesis



“Piebald dachshunds underwater playing video games”

Creative: Text-to-Image Synthesis

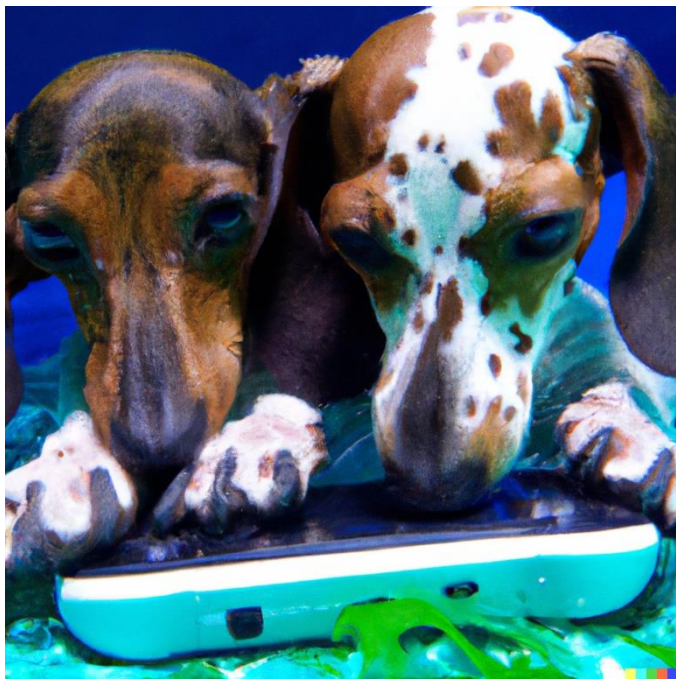


“Piebald dachshunds underwater playing video games”



“Long Haired Dachshunds playing chess”

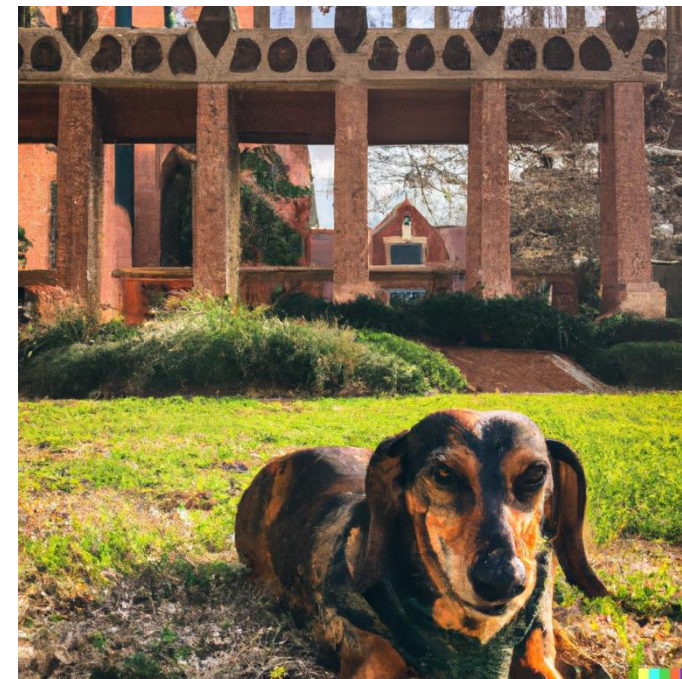
Creative: Text-to-Image Synthesis



“Piebald dachshunds underwater playing video games”



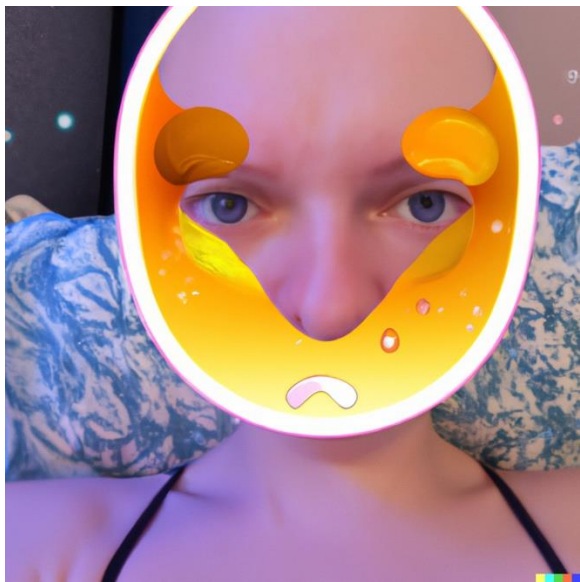
“Long Haired Dachshunds playing chess”



“A dapple dachshund at Rhodes College”

Wait a minute...

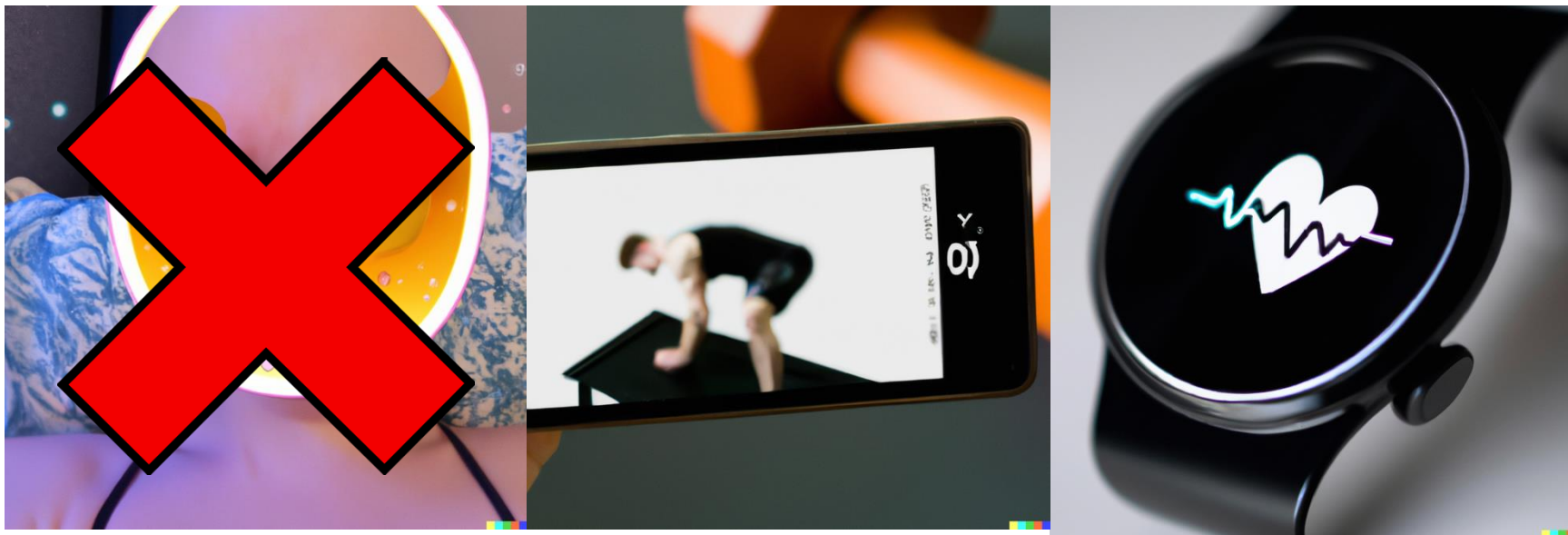
Wait a minute...



Wait a minute...



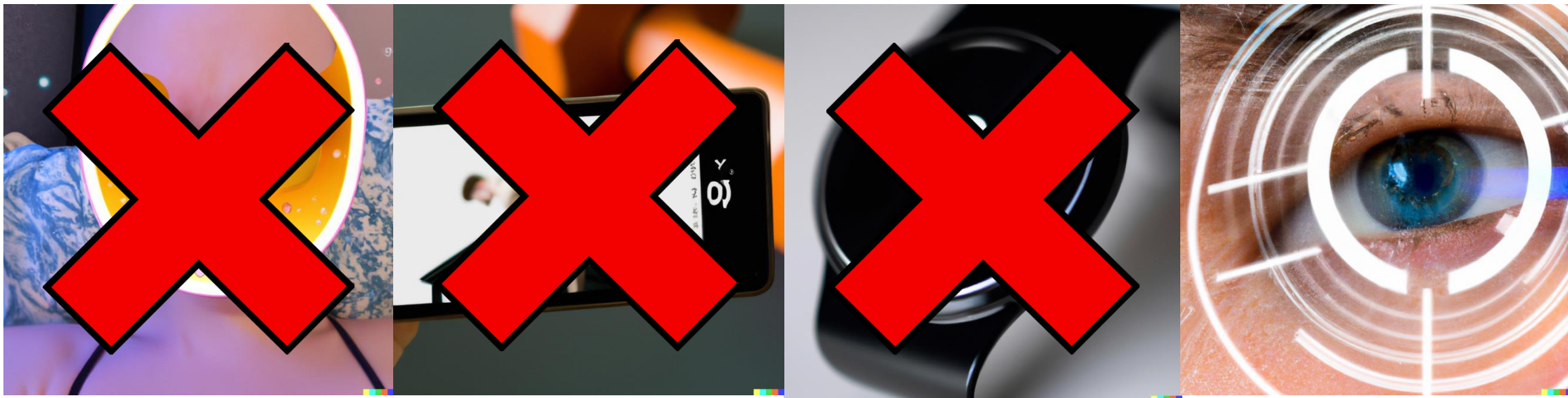
Wait a minute...



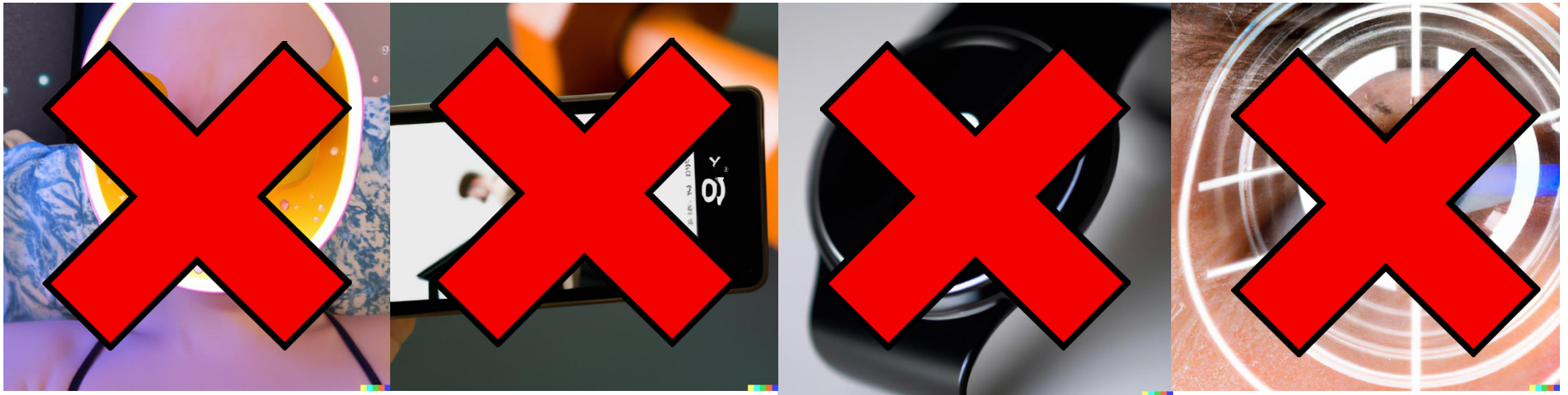
Wait a minute...



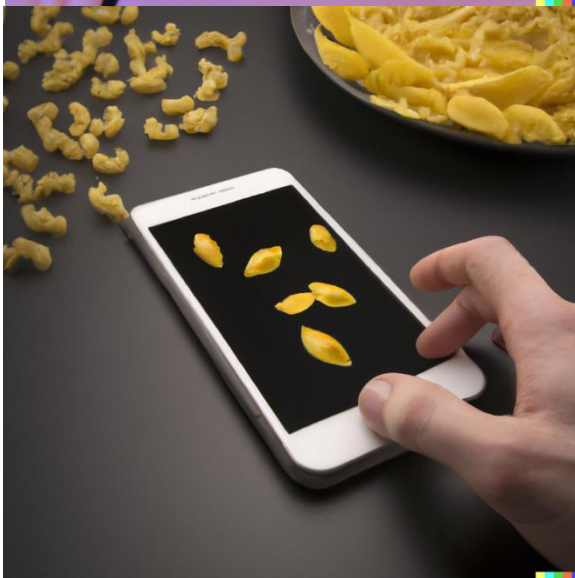
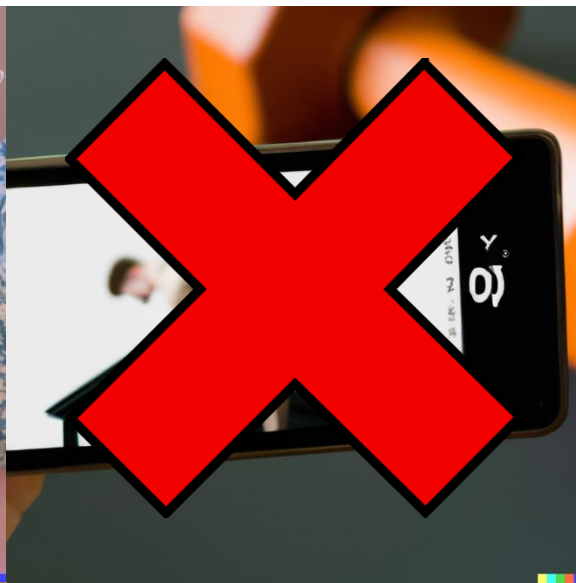
Wait a minute...



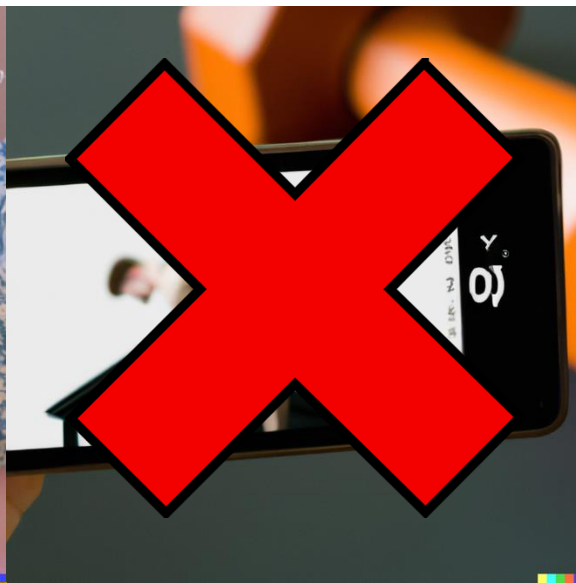
Wait a minute...



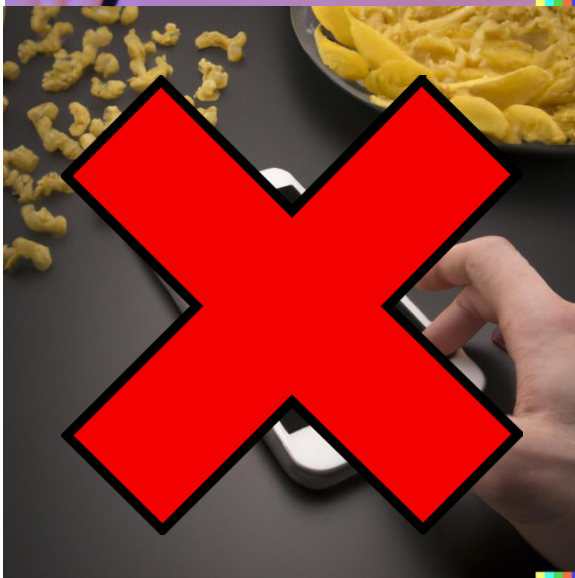
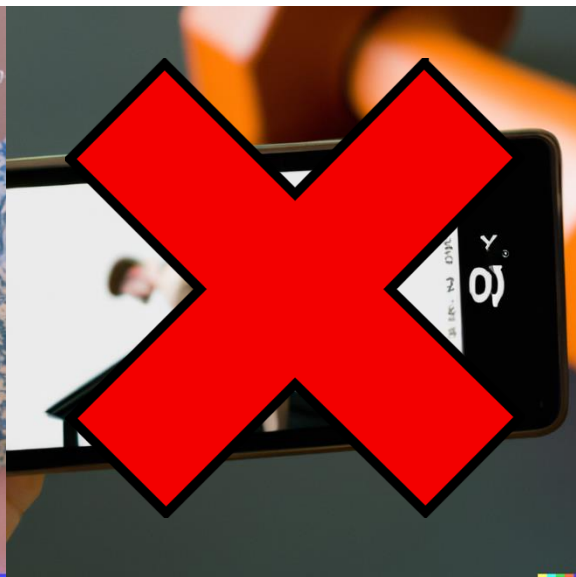
Wait a minute...



Wait a minute...



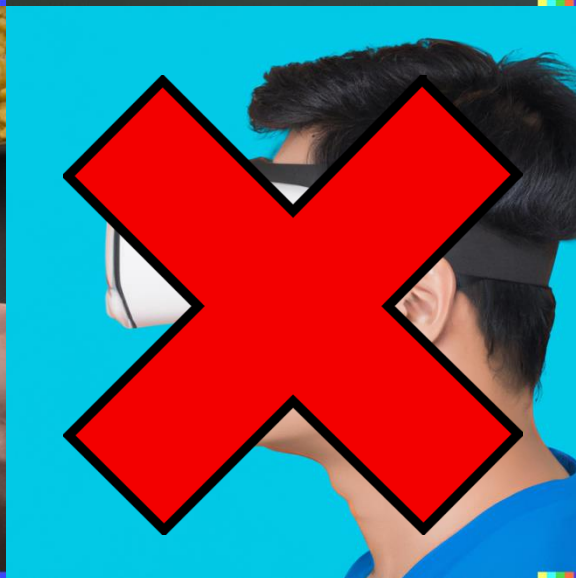
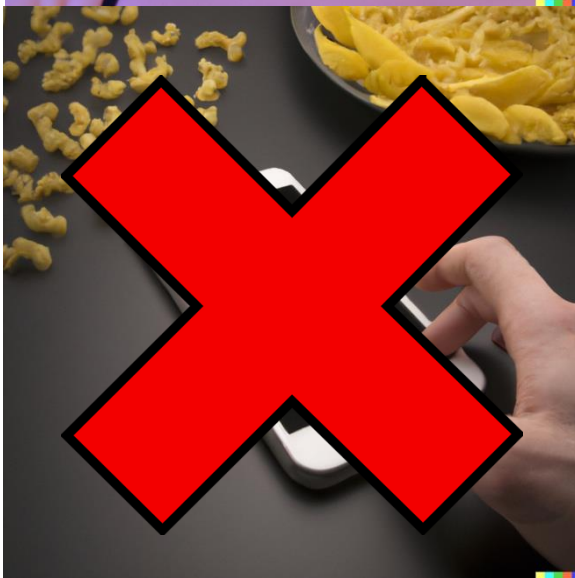
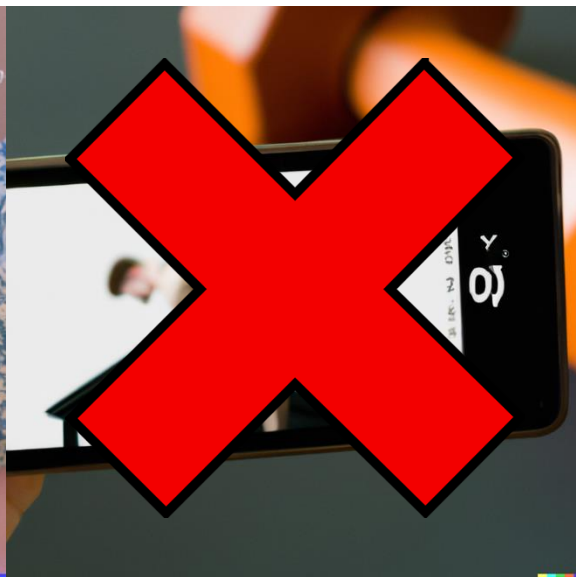
Wait a minute...



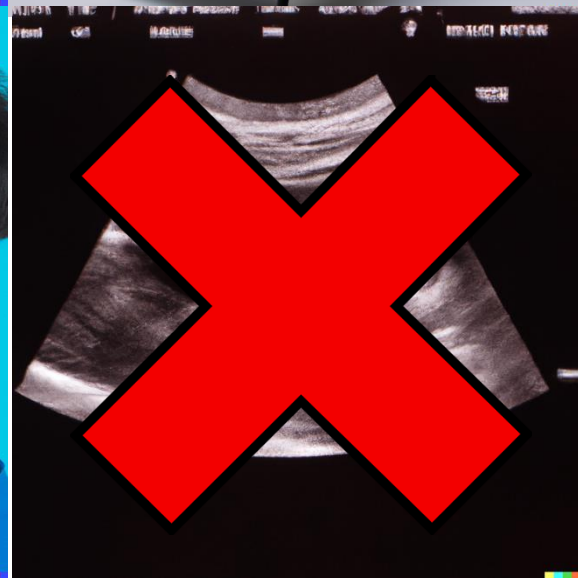
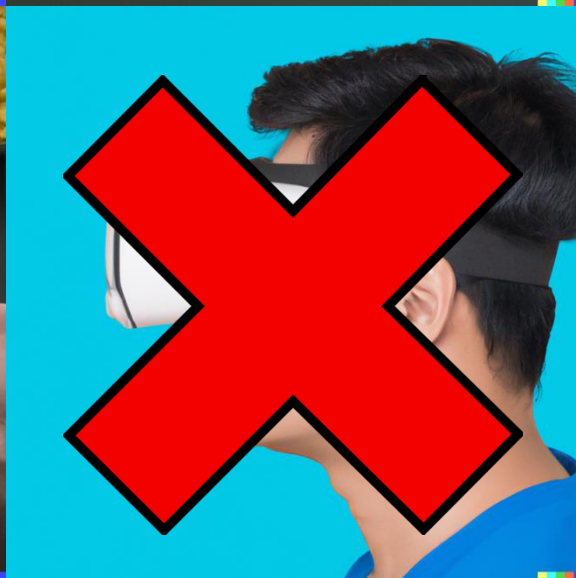
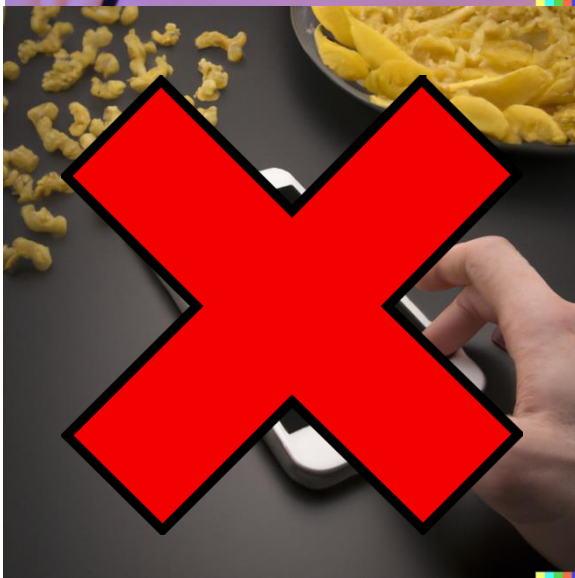
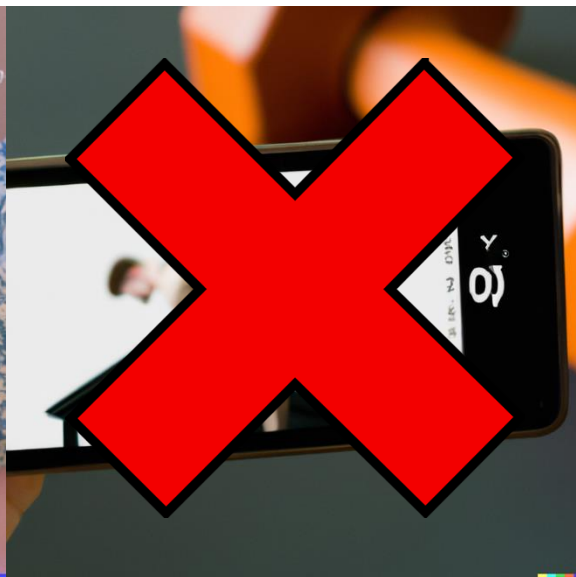
Wait a minute...



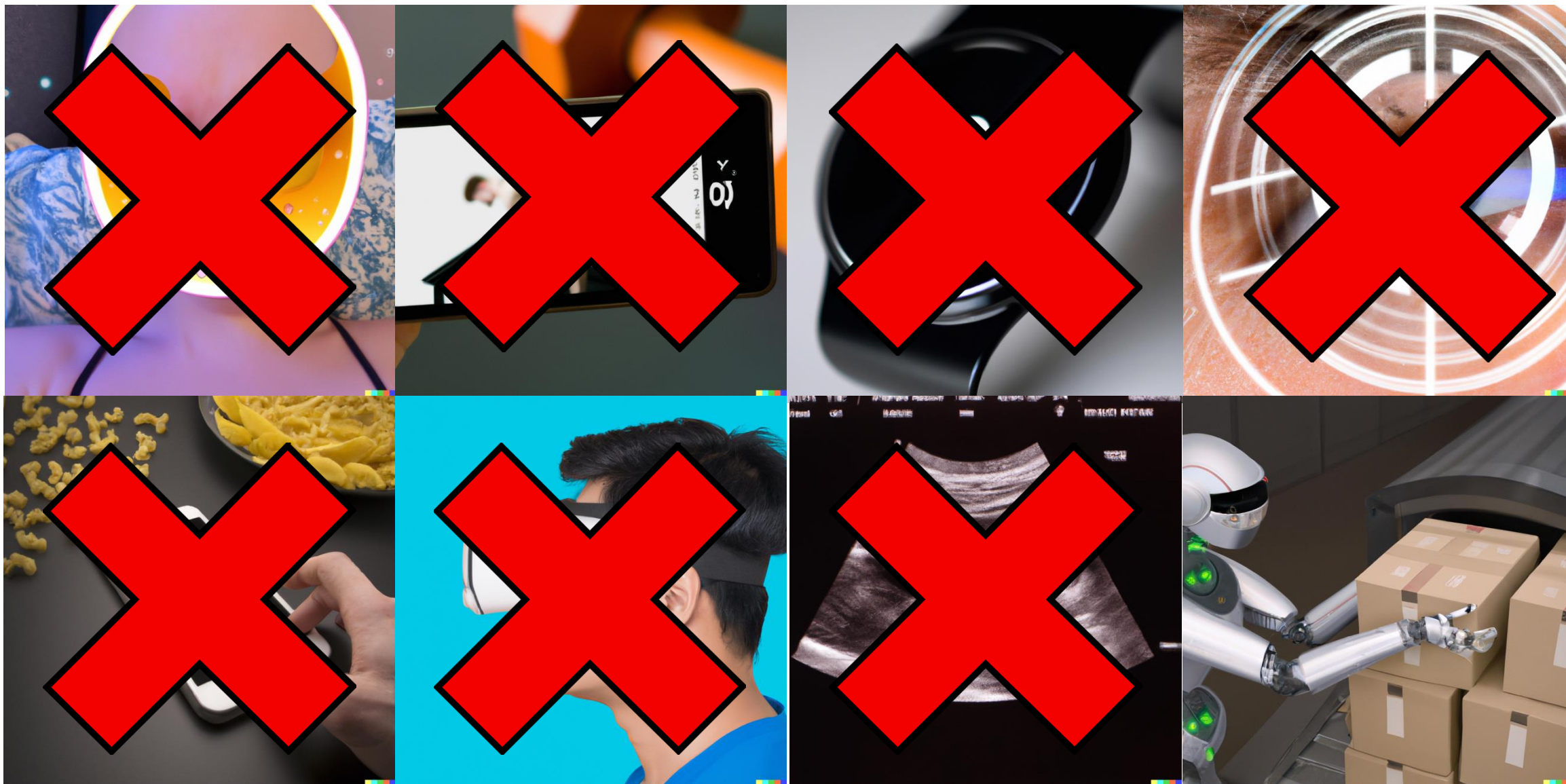
Wait a minute...



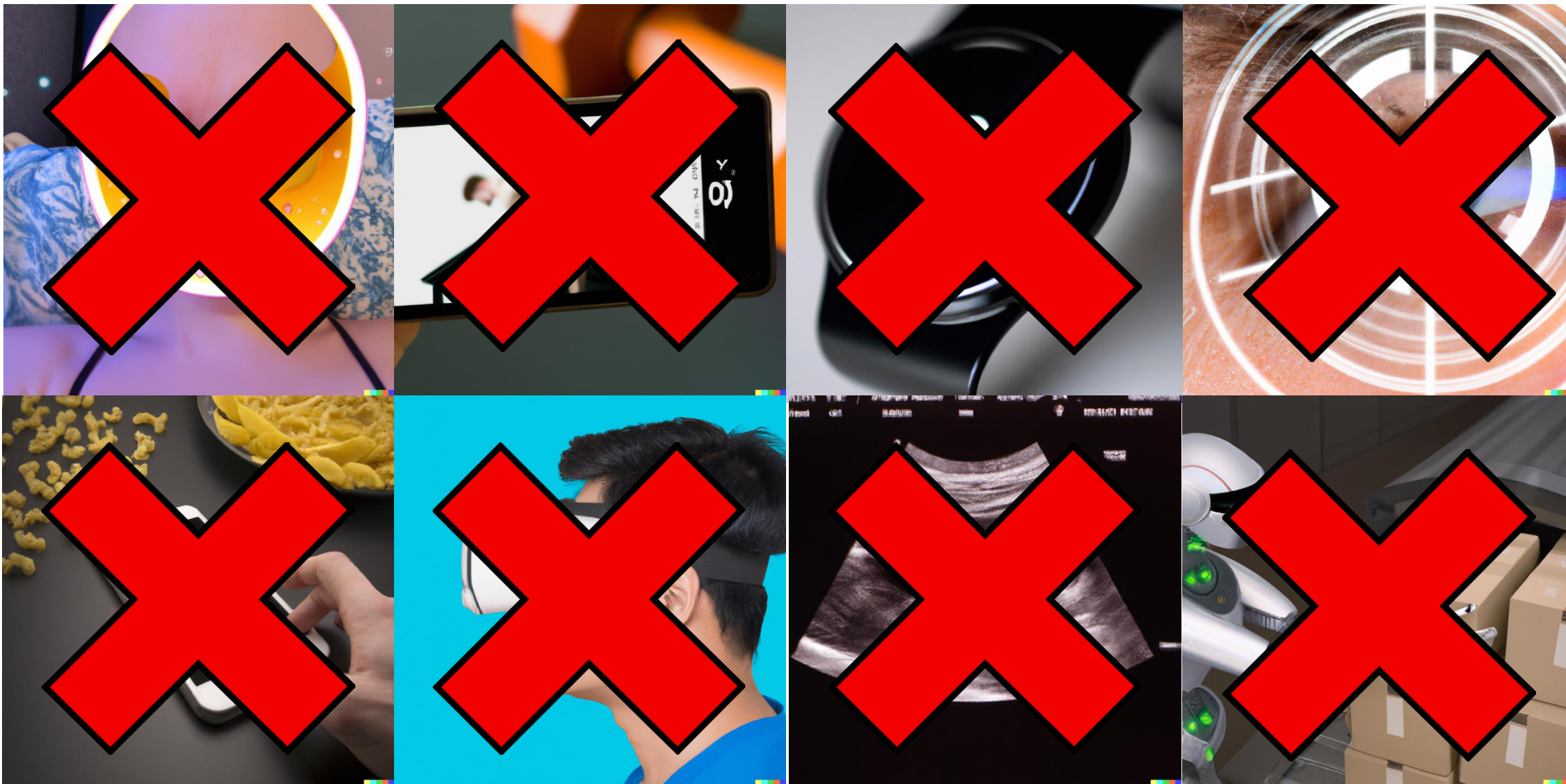
Wait a minute...



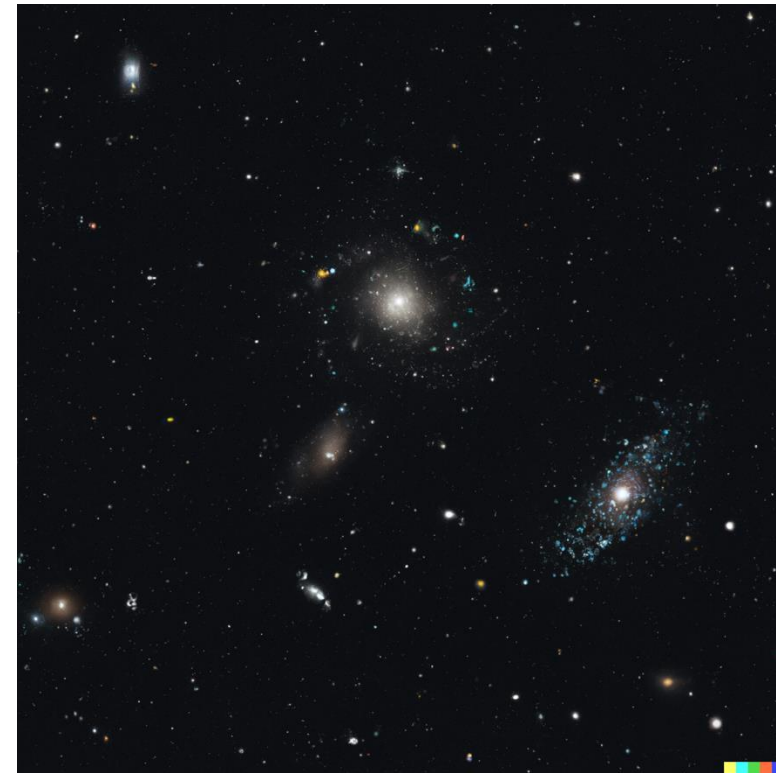
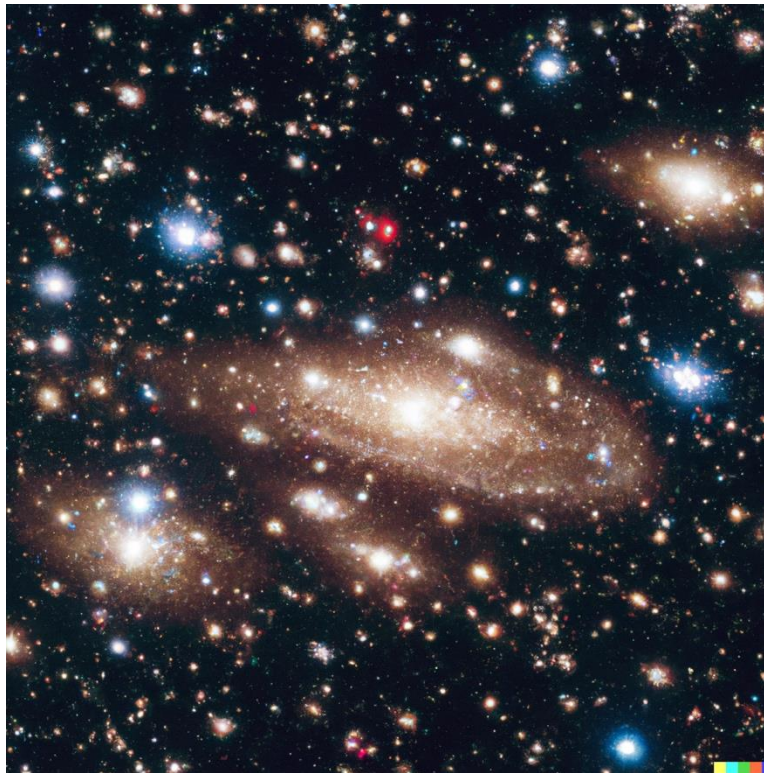
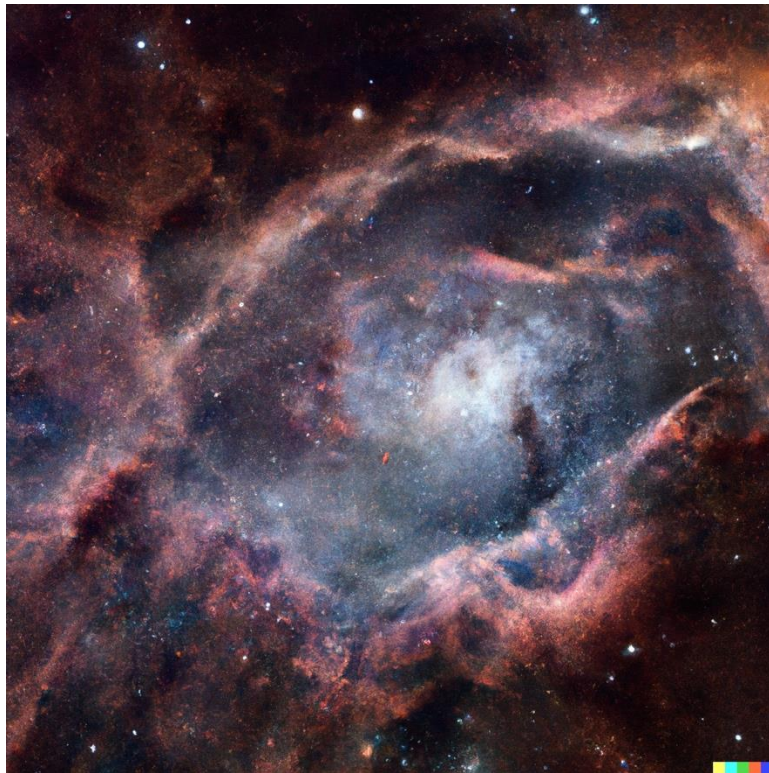
Wait a minute...



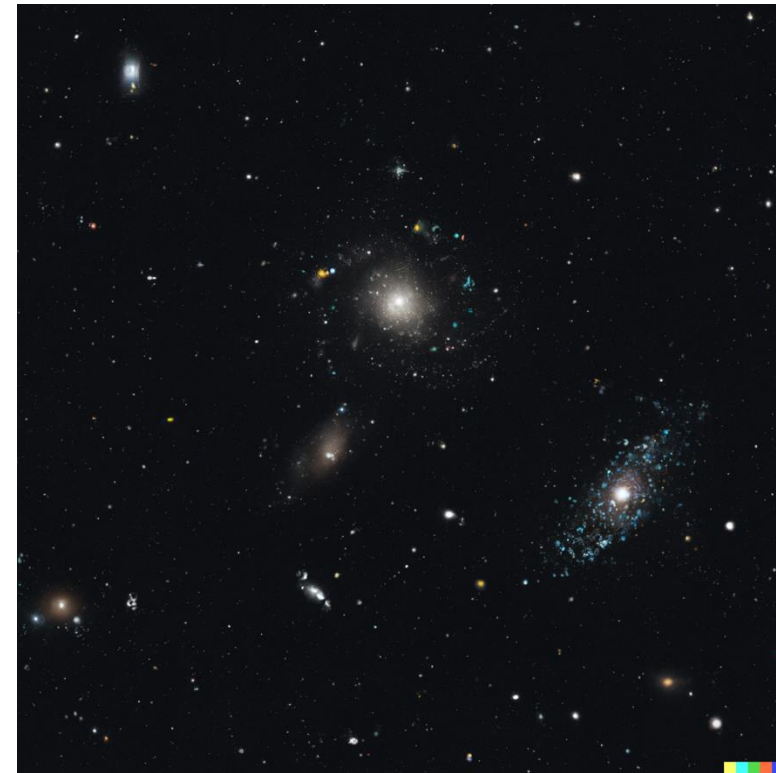
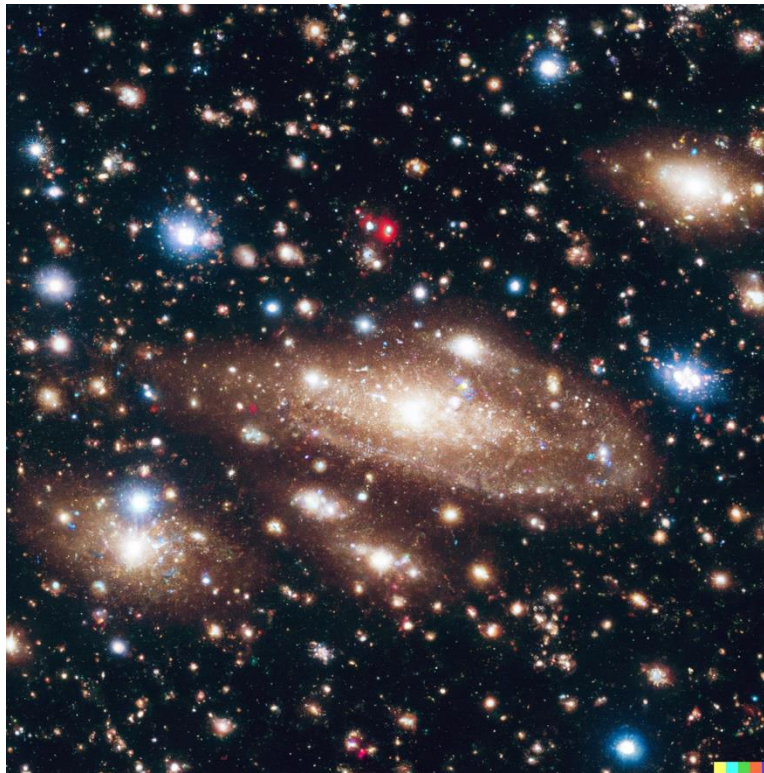
Wait a minute...



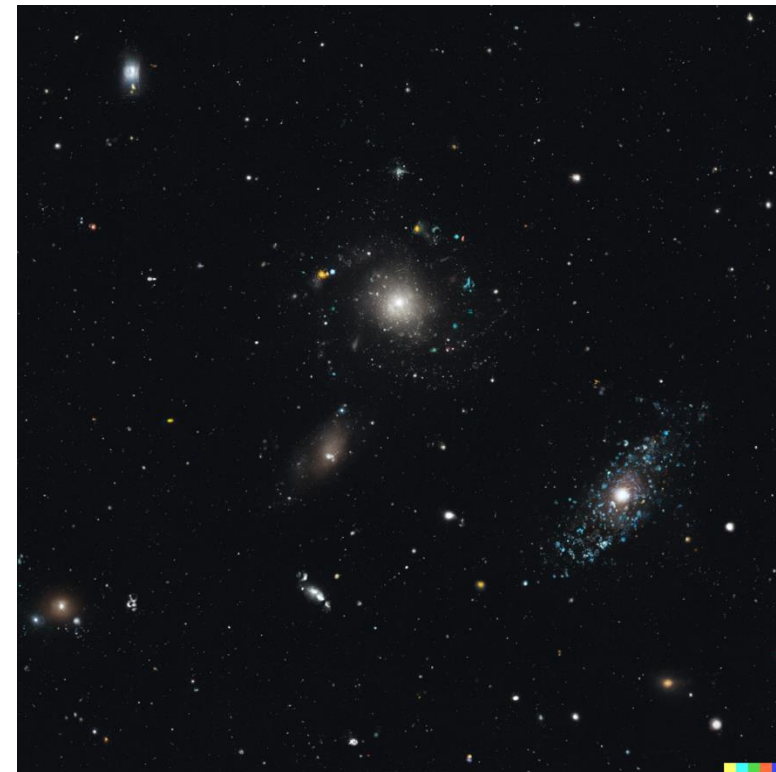
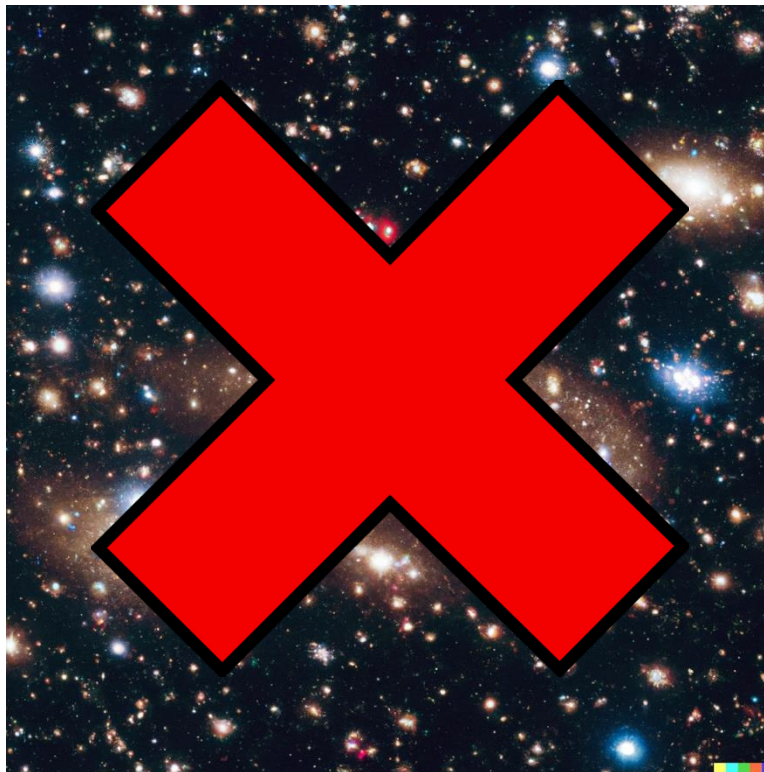
... these?



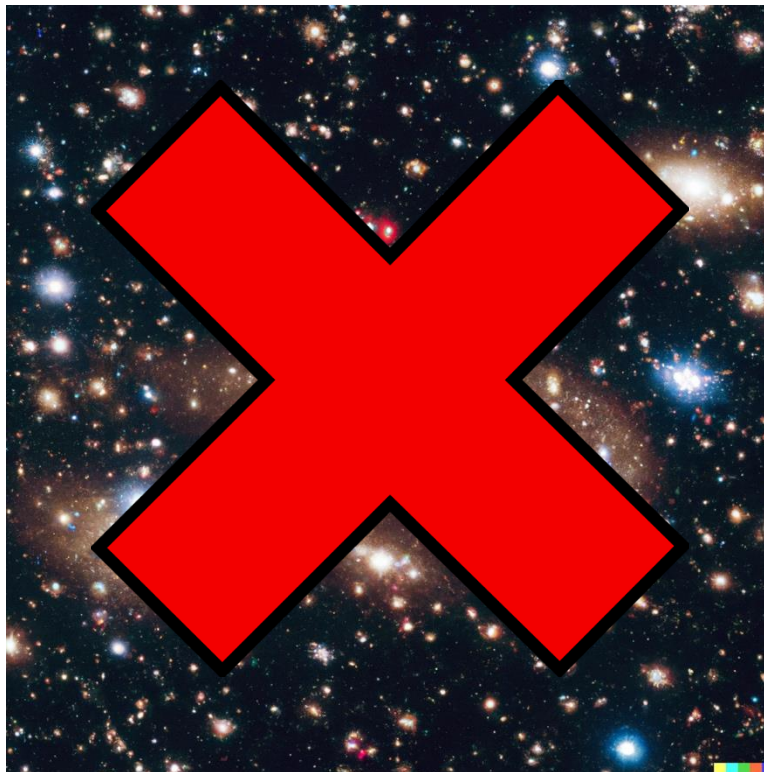
... these?



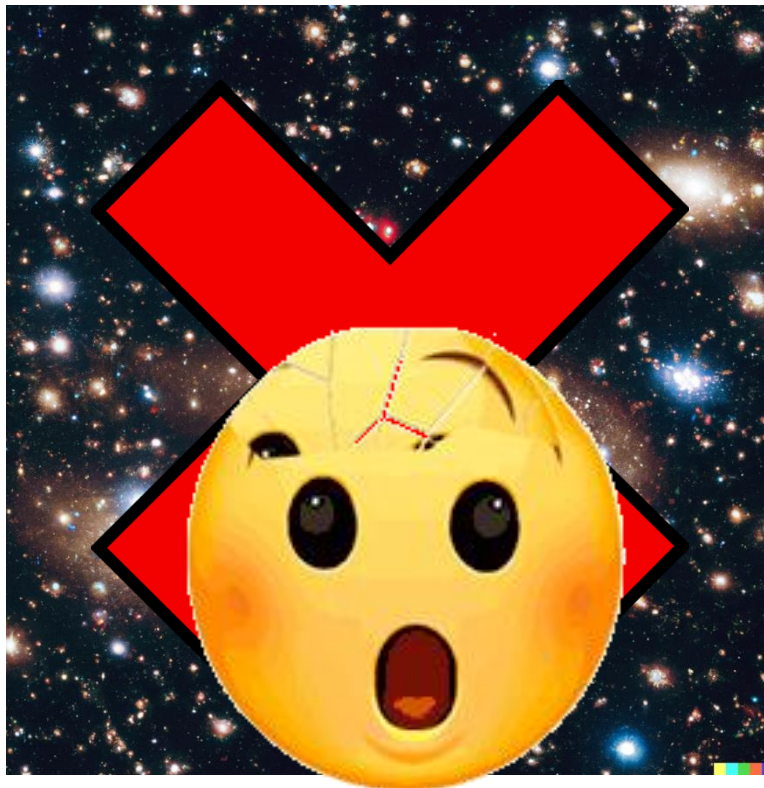
... these?



... these?



... these?



Jitendra Malik, UC Berkeley

Three 'R's of Computer Vision



“[Further progress in] the classic problems of computational vision:

reconstruction

recognition

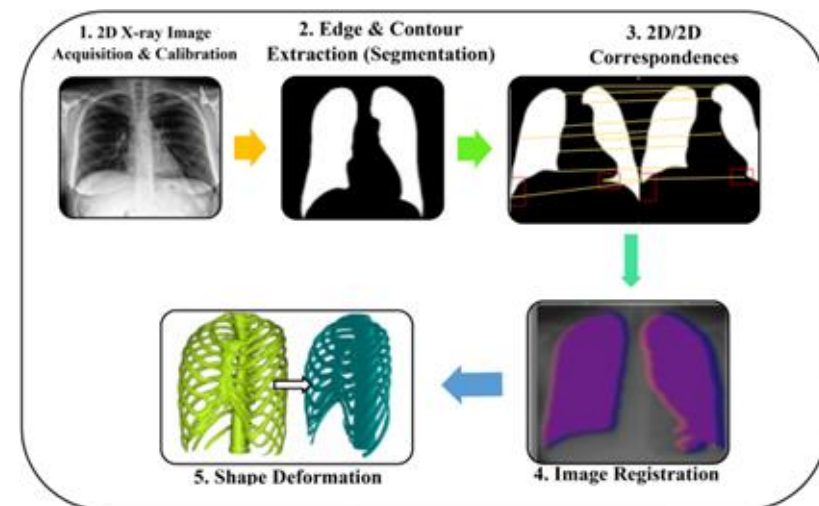
(re)organization

[requires us to study the interaction among these processes]”

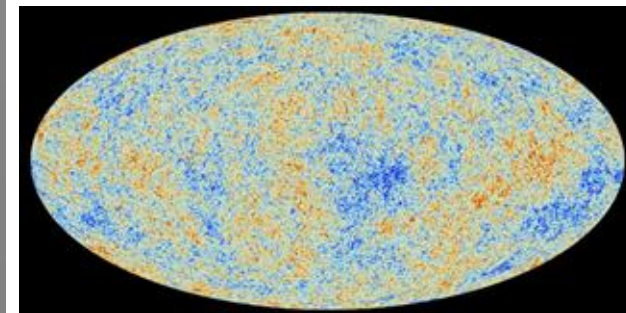
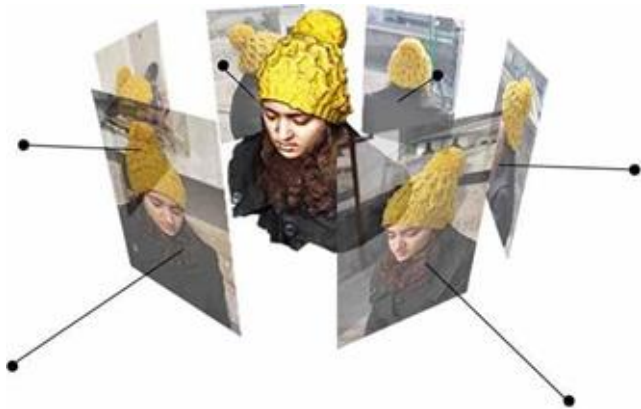
Vision as a poster child for success in AI



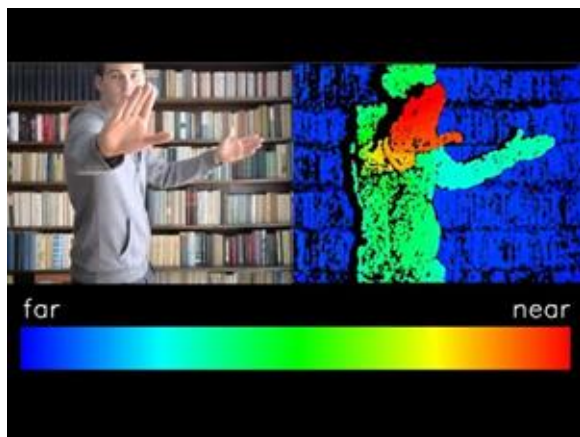
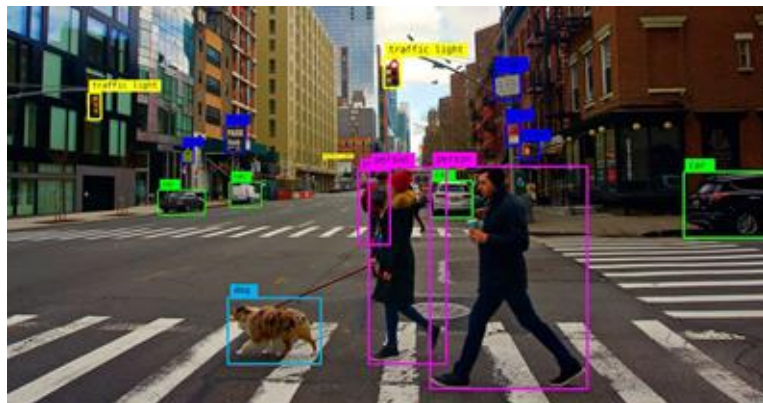
Boston Dynamics (2017)



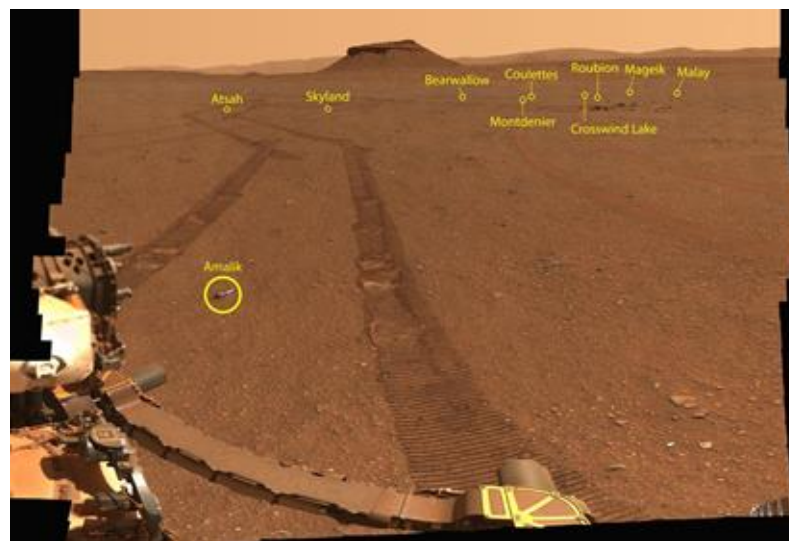
What is IP and why?



Lots more!



What is CV and why?



Lots more!

COMP322: COMPUTER VISION & IMAGE PROCESSING

Welcome!



Questions?



Dall-E: "wall-e answering questions at a press conference"

COMP322: Computer Vision & Image Processing

Vision and Society



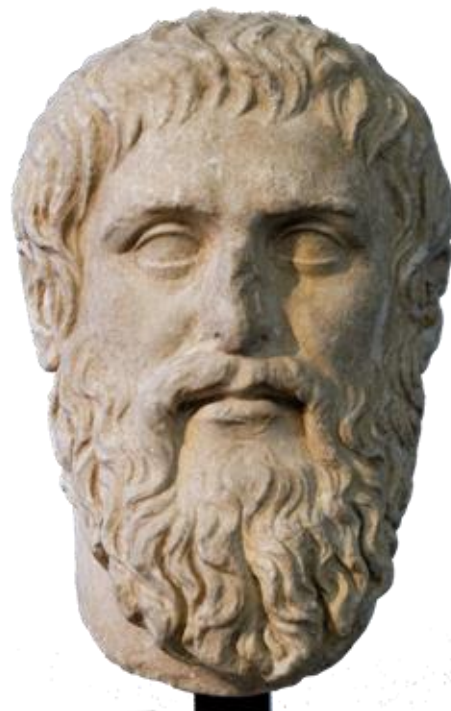
Image Credit: Pixar Animation Studios

Vision and Society

“World-view” vs. “World-sense”

“Vision, in my view, is the cause of the **greatest benefit** to us, inasmuch as none of the accounts now given concerning the Universe would ever have been given if men had not seen the stars or the sun or the heavens.”

- Plato (Timeus, 360 BC)



Vision and Society

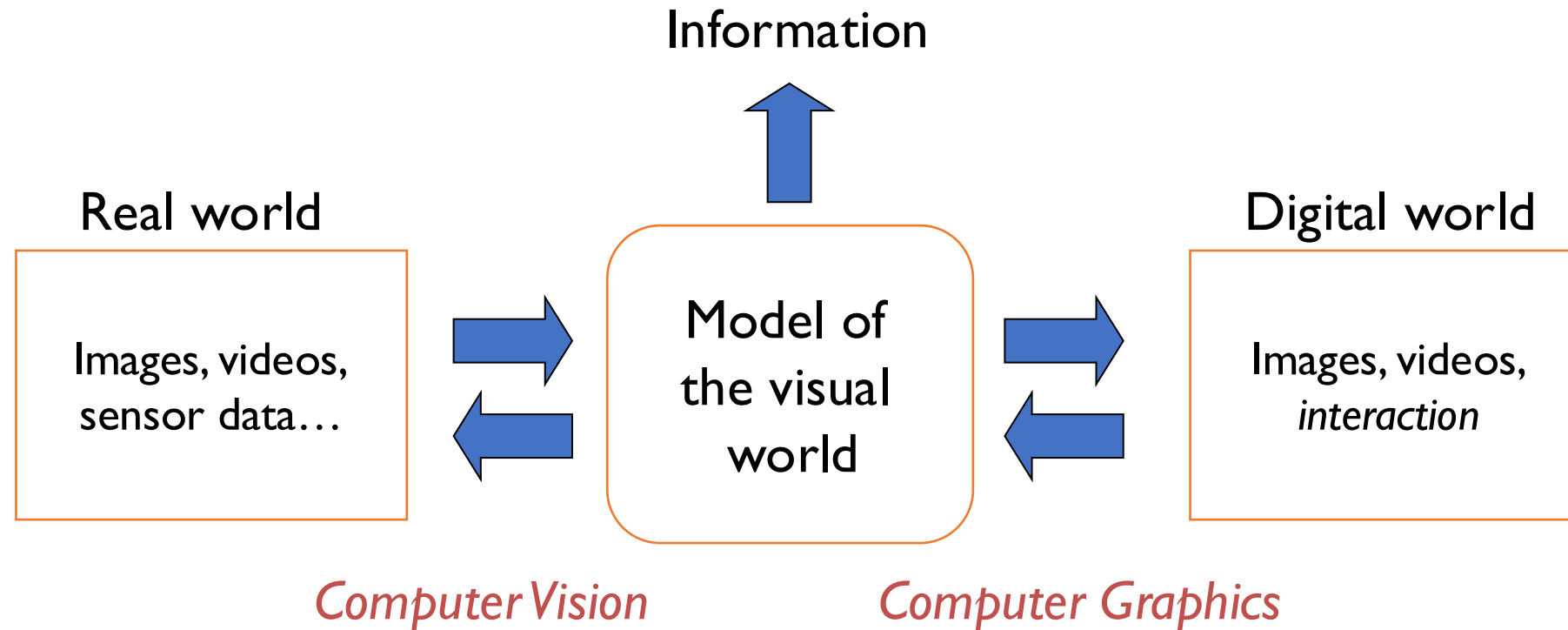
“World-view” vs. “World-sense”

“... In the West, [...] the world is primarily perceived by sight. The **differentiation** of human bodies in terms of sex, skin color, and cranium size is a testament to the powers attributed to "seeing." It is believed that just by looking at it [the body] **one can tell a person's beliefs and social position or lack thereof.**”

- Oyeronke Oyewumi
(The Invention of Women, 1997)

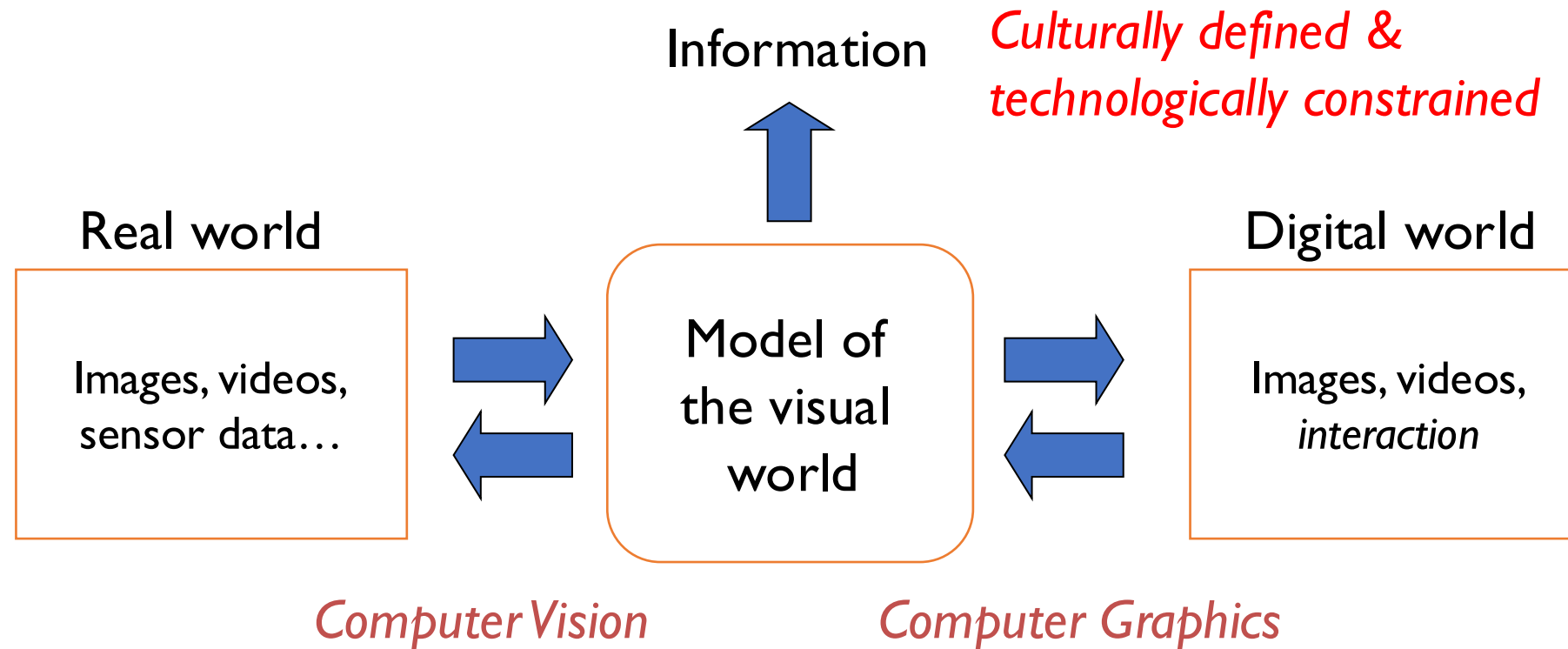


Computer Vision and Nearby Fields



“Machine learning applied to visual data.”

Computer Vision and Nearby Fields



~~“Machine learning applied to visual data.”~~

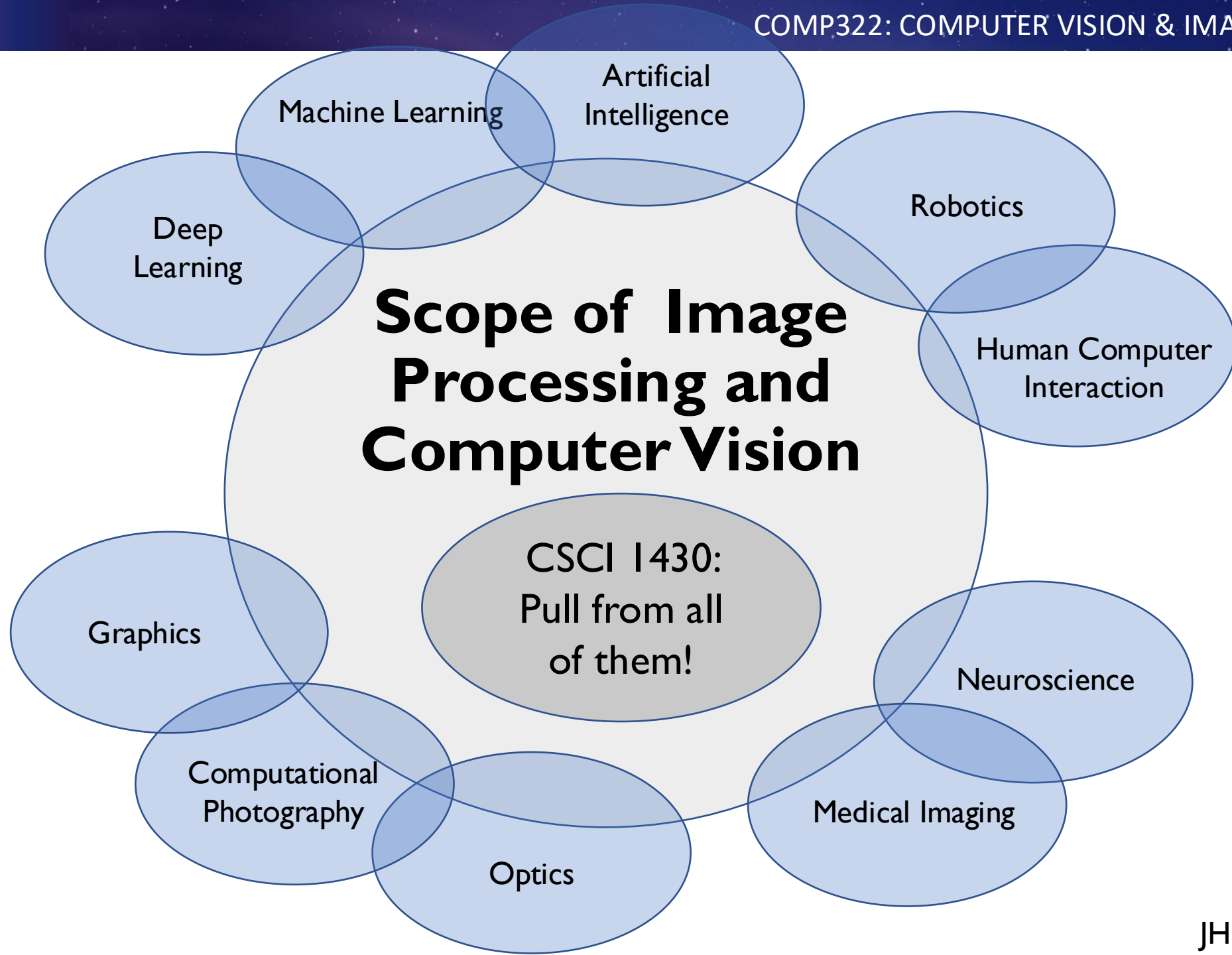
Superhuman state of the art?

Deep learning is an enormous disruption to the field. Since 2012,
rapid expansion and commercialization.

Why?

Reckless statement: “With enough data, computer vision matches
or even outperforms human vision at most recognition tasks.”

WHAT. Missing ‘world-sense’.



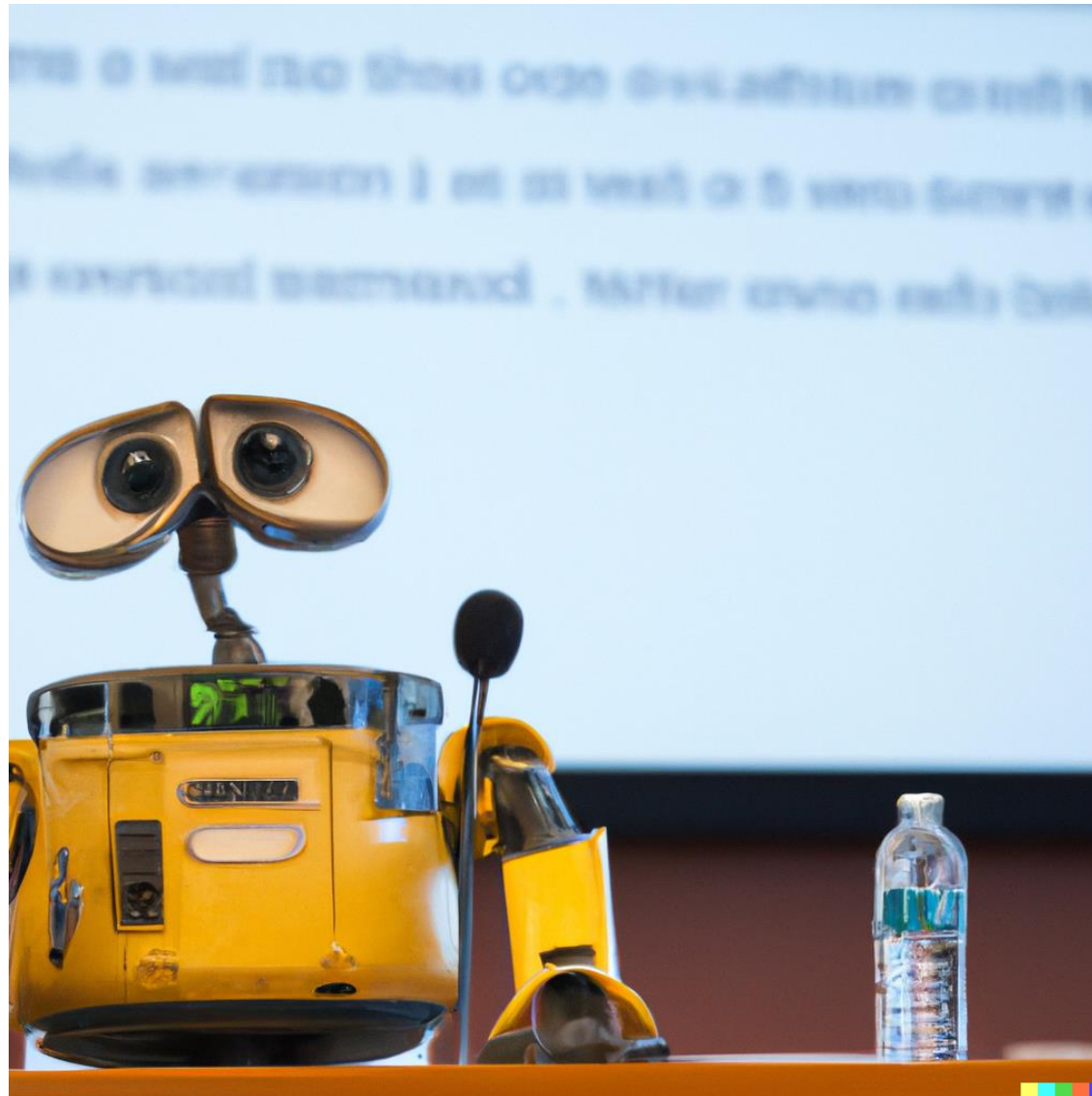
COMP322: Computer Vision & Image Processing

Course Logistics: Syllabus



Image Credit: Pixar Animation Studios

Questions?



Dall-E: "wall-e answering questions at a press conference"