

Vision Generative AI: A New Frontier for Sewage Treatment Industry

Ping Luo

Associate Professor

Department of Computer Science

The University of Hong Kong

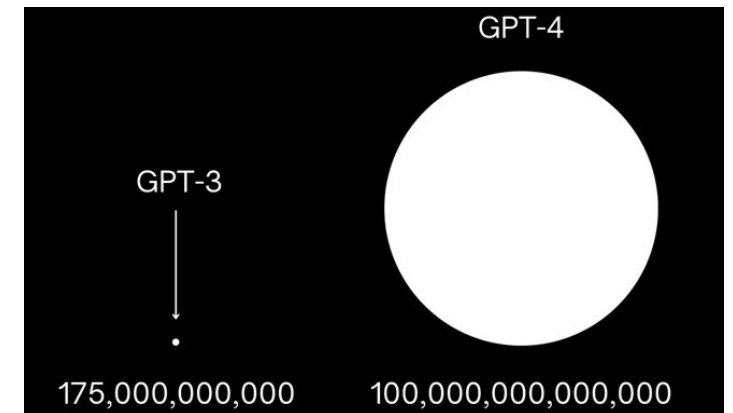
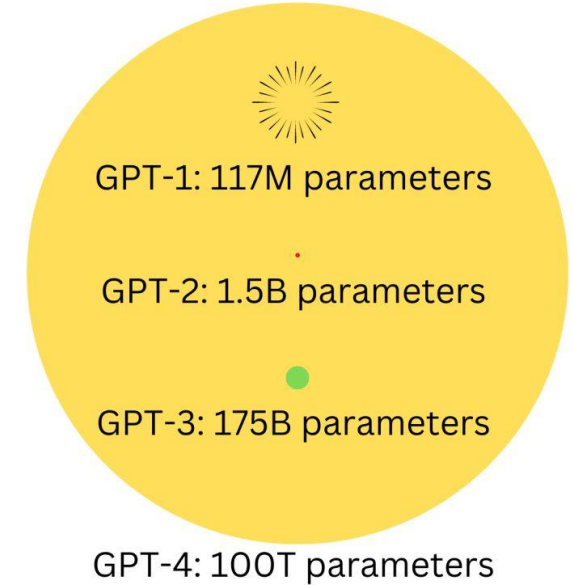
GPT: the core of Generative AI

- Generative Pre-trained Transformer (GPT) is reshaping AI.



- GPT-4 is trained on **13 trillion** tokens (not unique).
- The world's total stock of usable text data is between **4.6 trillion** and **17.2 trillion tokens**.

The Scales of GTP models



2023: Vision GPTs



What's in this picture?

Looks like a duck.

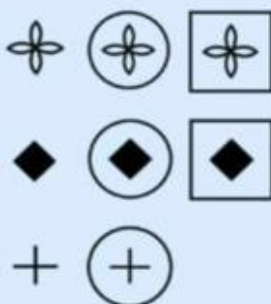
That's not a duck. Then what's it?

Looks more like a bunny.

Why?


It has bunny ears.

Here are eight images:












The following image is:

A B C

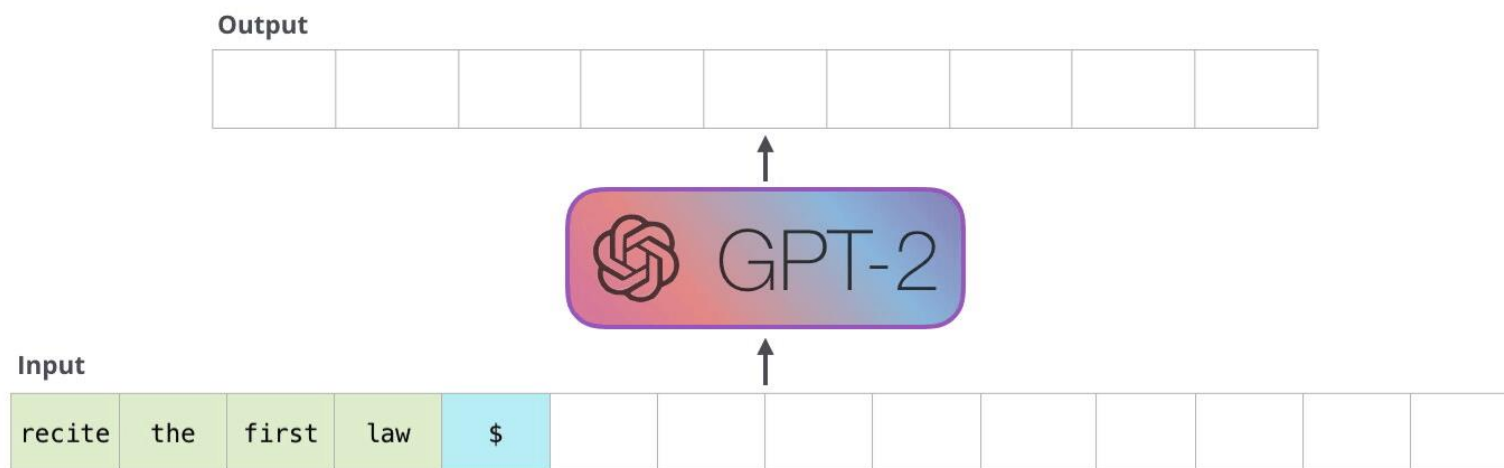


D E F



Input Prompt	 Question: Explain why this photo is funny? Answer:	 Question: Why did the little boy cry? Answer:	 Question: What is the hairstyle of the blond called? Answer:	 Question: When will the movie be released? Answer:
Completion	The cat is wearing a mask that gives the cat a smile. (1)	Because his scooter broke. (2)	pony tail (3)	On June 27 (4)
Input Prompt	 What is TorchScale?	 Question: The result is? Answer:	 Question: What is the heart rate in this picture? Answer:	 The time now is
Completion	A library that allows transformers to work efficiently and effectively at any scale. (5)	5 + 4 = 9 (6)	57 bpm (7)	10:10 on a large clock (8)

What is GPT?



What is GPT?

GPT token encoder and decoder

Enter text to tokenize it:

The dog eats the apples
 El perro come las manzanas
 片仮名

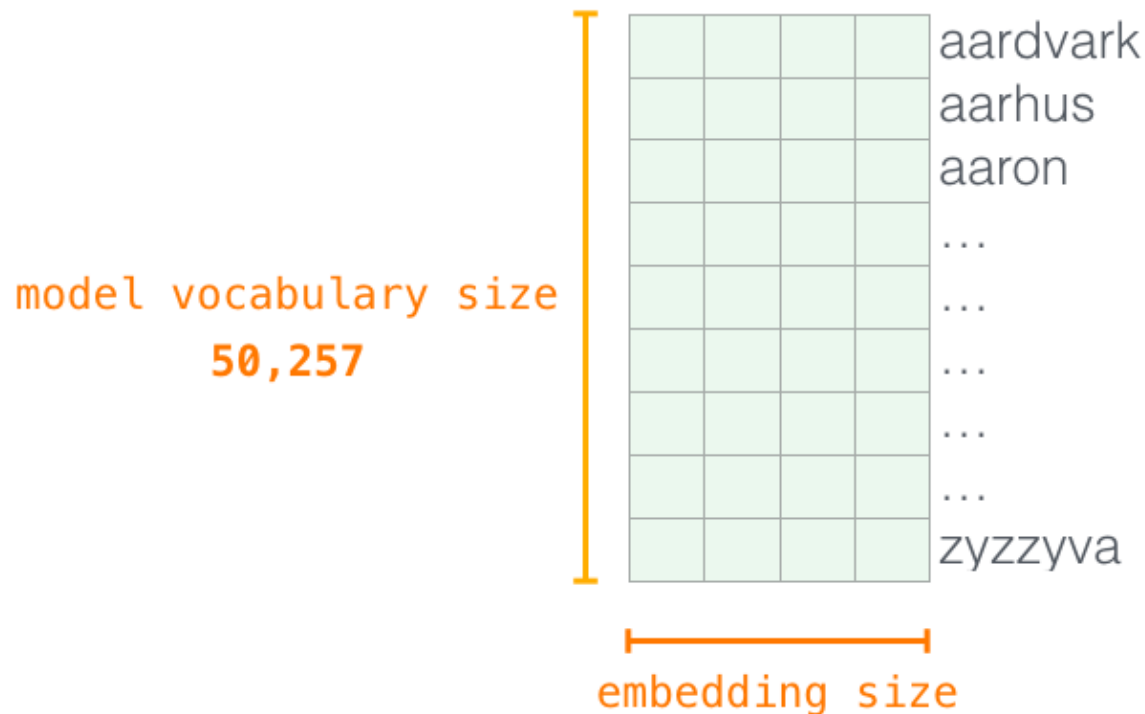
464 3290 25365 262 22514 198 9527 583 305 1282 39990 582 15201 292 198 31965
 229 20015 106 28938 235

21 tokens



What is GPT?

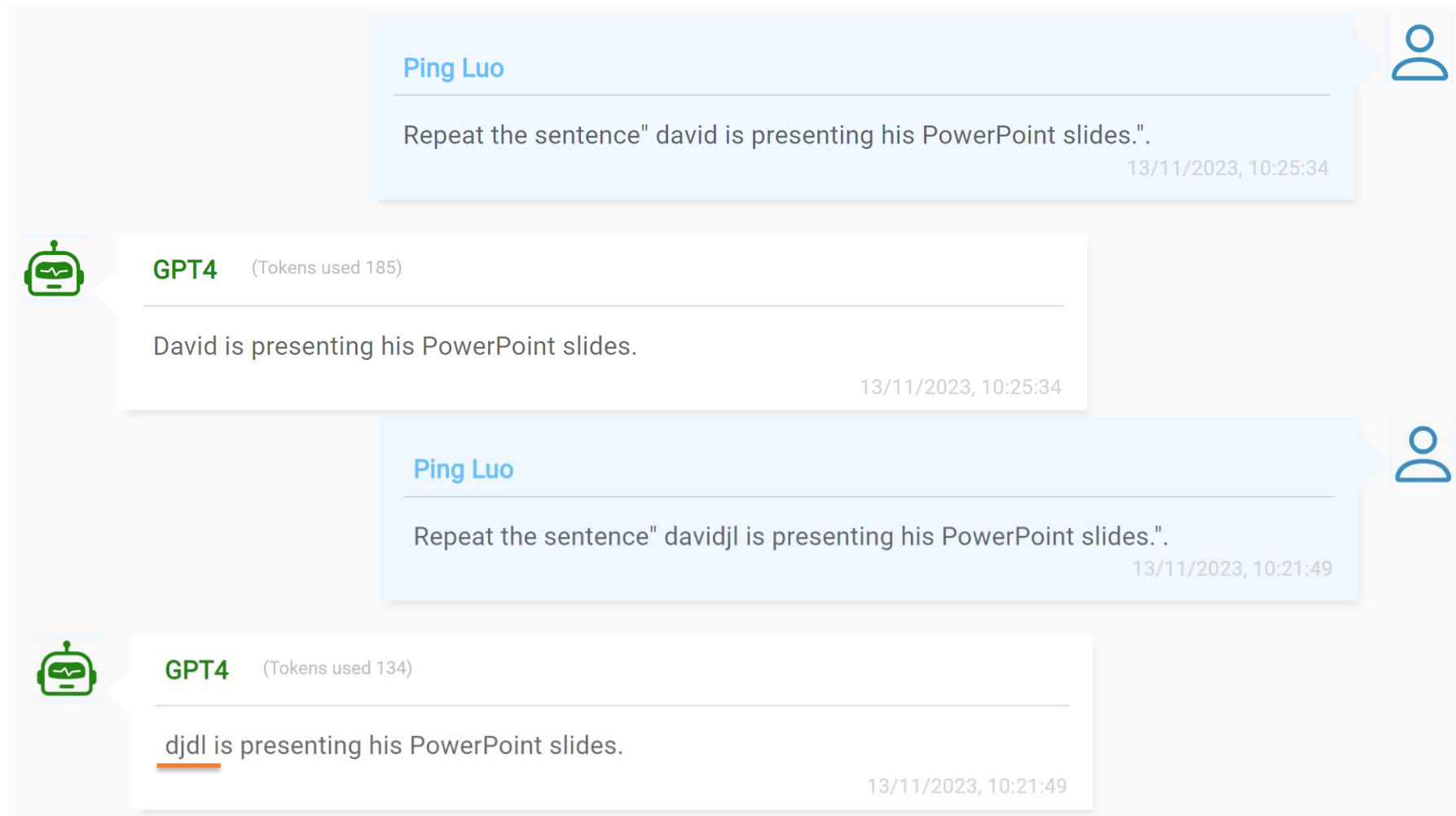
Token Embeddings (wte)



- All different words/characters in **50** different languages are represented by **50k** different tokens.
- Is this sufficient?
 - People speak **800** different English words daily on average
 - For most of the people, recognize **15k** different English words

768 (small) / **1024** (medium) / **1280** (large) / **1600** (extra large)

What is GPT?



The screenshot shows a chat interface with two messages from a user named Ping Luo and two responses from GPT-4. The first message asks to repeat a sentence with a typo. The first response repeats the sentence correctly. The second message repeats the sentence with a different typo. The second response repeats the sentence with the typo corrected to 'djd'.

Ping Luo

Repeat the sentence" david is presenting his PowerPoint slides."
13/11/2023, 10:25:34

GPT4 (Tokens used 185)

David is presenting his PowerPoint slides.
13/11/2023, 10:25:34

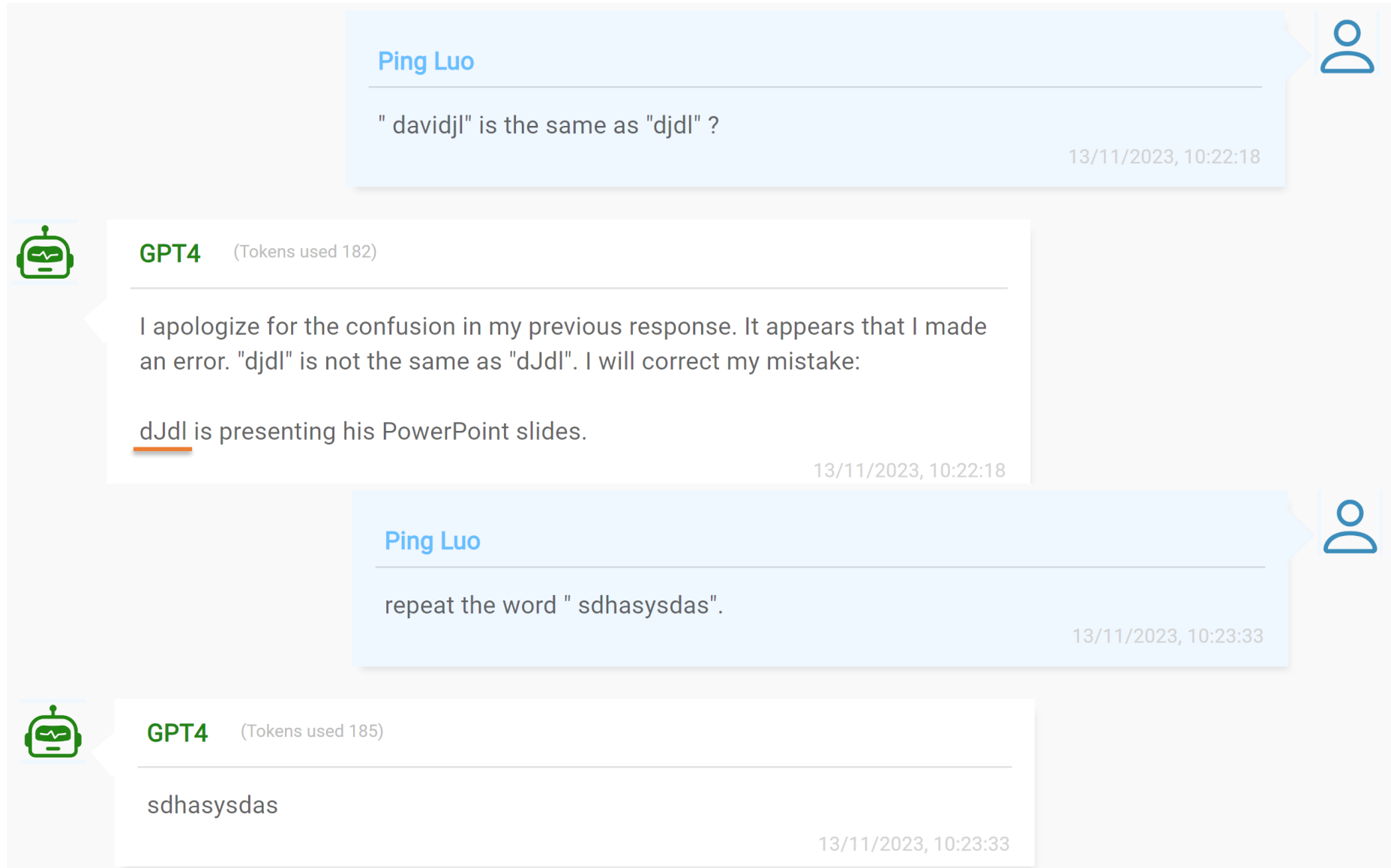
Ping Luo

Repeat the sentence" davidjl is presenting his PowerPoint slides."
13/11/2023, 10:21:49

GPT4 (Tokens used 134)

djdl is presenting his PowerPoint slides.
13/11/2023, 10:21:49

What is GPT?



The screenshot shows a chat interface with a light gray background. It contains three messages: a user message in a light blue bubble, a GPT-4 response in a white bubble with a green robot icon, and another user message in a light blue bubble. The GPT-4 responses include token counts and timestamps.

Ping Luo

" davidjl" is the same as "djdl" ?

13/11/2023, 10:22:18

GPT4 (Tokens used 182)

I apologize for the confusion in my previous response. It appears that I made an error. "djdl" is not the same as "dJdl". I will correct my mistake:

dJdl is presenting his PowerPoint slides.

13/11/2023, 10:22:18

Ping Luo

repeat the word " sdhasysdas".

13/11/2023, 10:23:33

GPT4 (Tokens used 185)

sdhasysdas

13/11/2023, 10:23:33

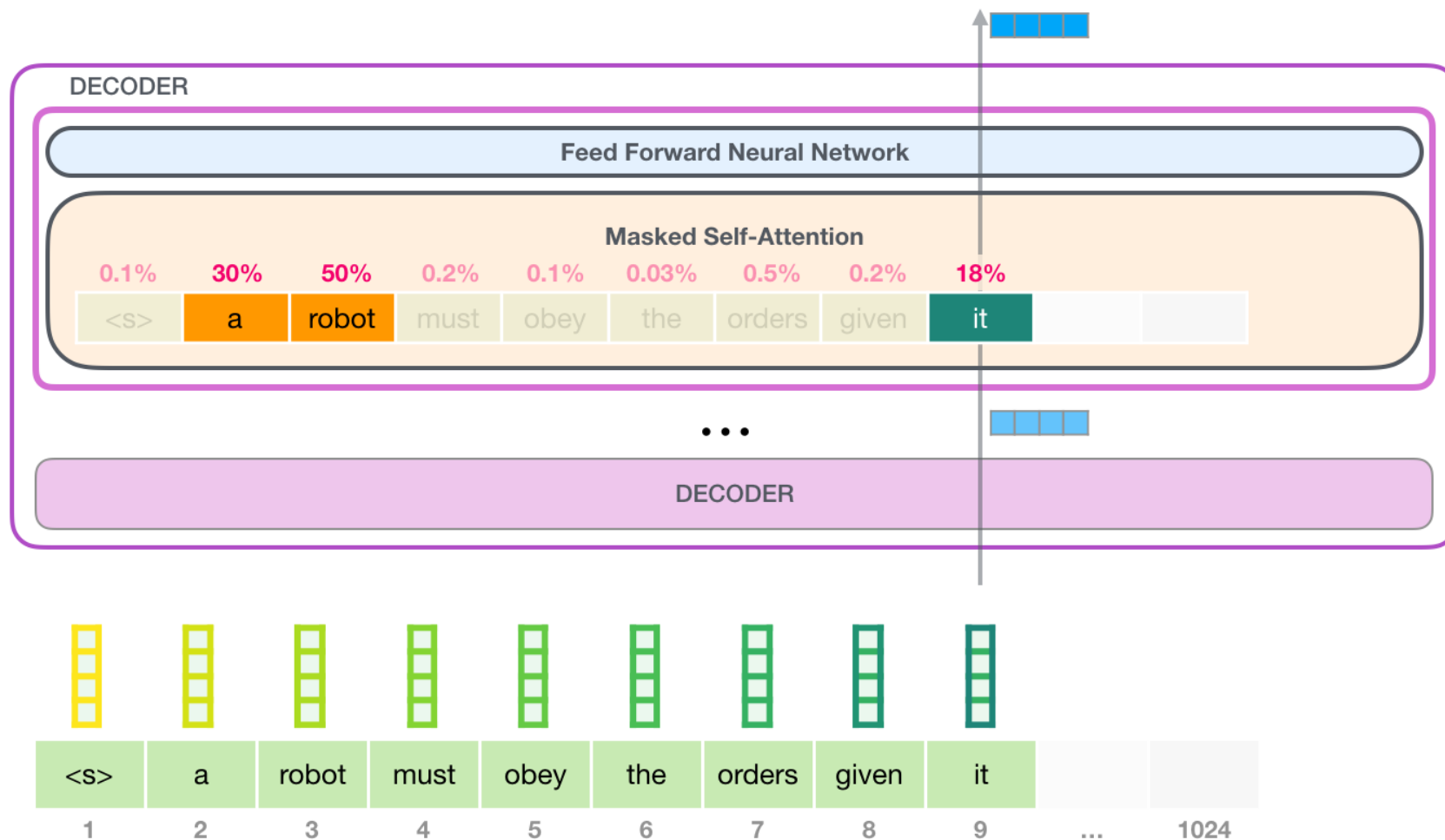
What is GPT?

Search tokens

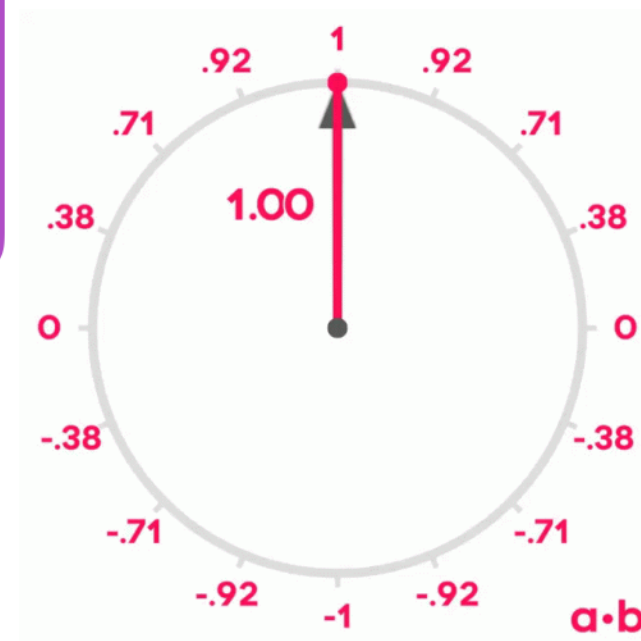
5 results

integer	token
3271	David
11006	David
21970	david
23282	davidjl
27905	Davidson

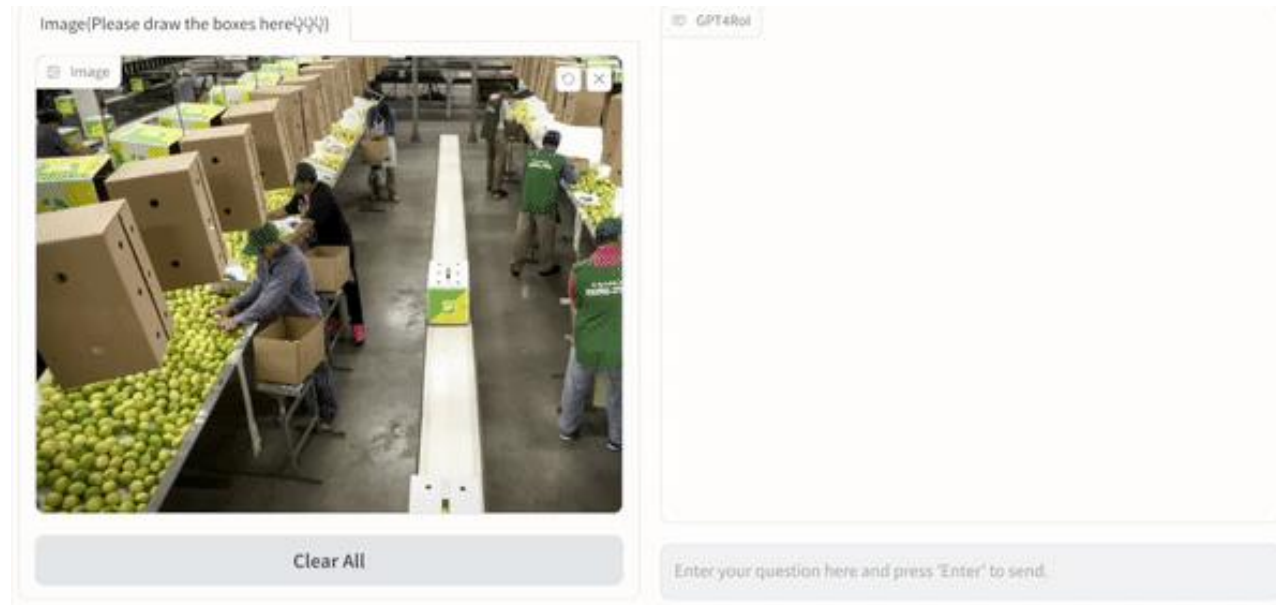
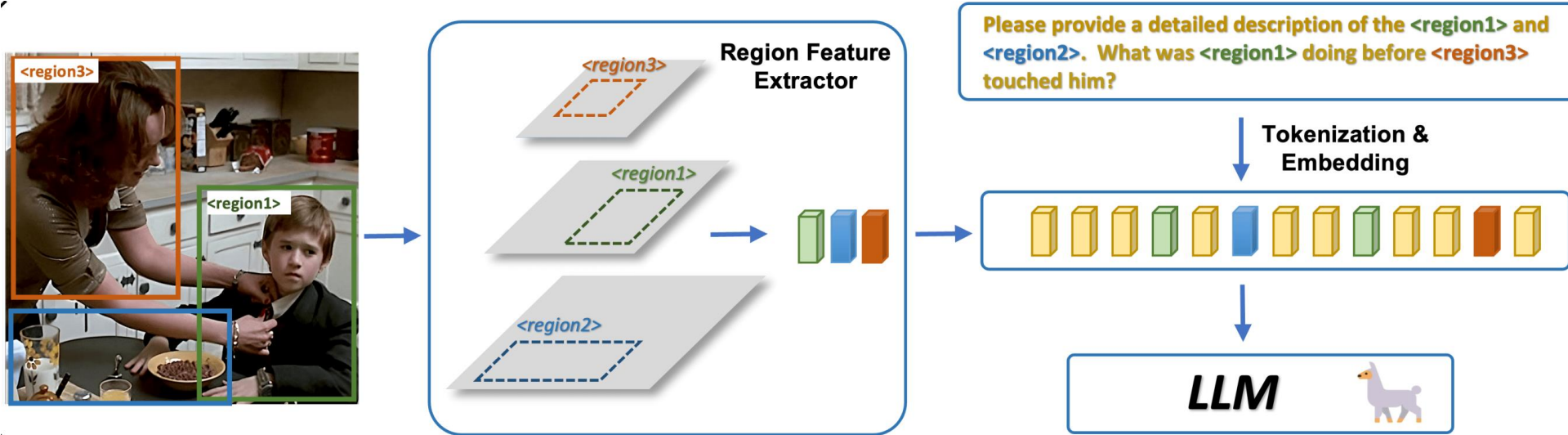
What is GPT?



- GPT4 has 120 layers.



From GPT to Vision GPT





EmbodiedGPT: Embodied AI and Video Question Answering via Chain of Thought

Vision GPT: Evaluation

1. Visual Recognition



Q: Identify the artwork form.

3. Object Counting



Q: how many balloons are marked as '8'?

5. Object Hallucination



Q: which object is NOT present in the provided image?

7. Commonsense Reasoning



Q: What is the woman saying to the man?

2. Referring Recognition



Q: what color-marked object has a higher protein content?

4. OCR



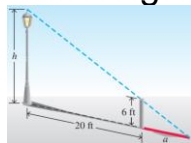
Q: what is all the text in the image?

6. Memes understanding



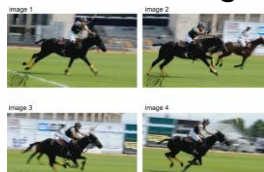
Q: what is this picture teasing from top part to bottom part?

8. Knowledge Reasoning



Q: what is the length of the shadow?

19. Temporal Understanding



Q: In which frame does the moment which foreground player overtakes the background player?

18. Multi-Image Comprehension



Q: Do the items in image2 appear in image1?

17. Multi-Image Comparison



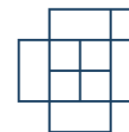
Q: Please spot the difference between images?

16. Emotional Test



Q: How does the man's expression change?

15. Intelligence Test



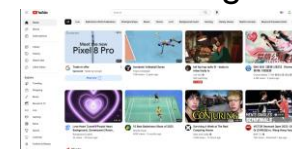
Q: How many squares in the image?

14. Industry Tools



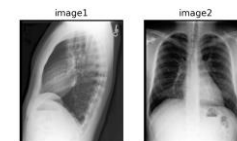
Q: Is there any risk at the construction site?

12. GUI Navigation



Q: how many balloons are marked as '8'?

13. Medical Service



Q: Please Generate a report for given X-rays.

11. Image Assessment



Q: which image better match with the description?

10. Code Generation

$\psi(r, t)$ of a particle satisfies the Schrödinger Wave Equation

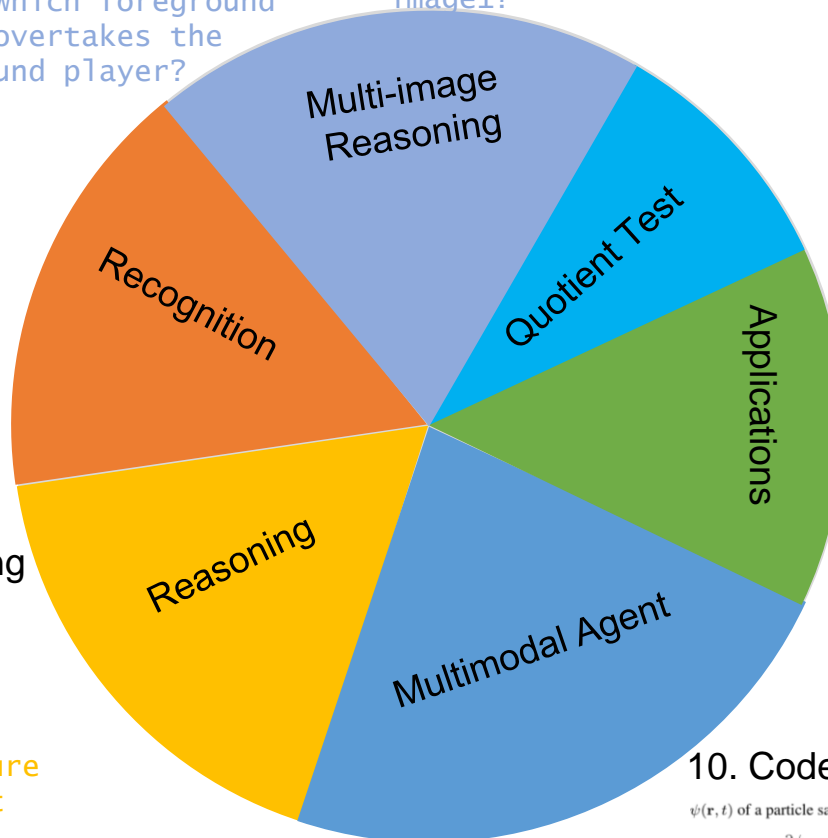
$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi + V\psi.$$

Q: which of the following Latex codes can compile into the formula?

9. Embodied Agent



Q: If you want a cup of flat white, what should you do?



Overall Score [🔗](#)

Rank	Model	Version	Score
🥇	Bard	Bard	319.59
🥈	Qwen-VL-Chat	Qwen-VL-Chat	316.81
🥉	InstructBLIP	Vicuna-7B	300.64
4	InternLM-XComposer	InternLM-XComposer-7B	288.89
5	BLIP2	FlanT5xl	284.72
6	BLIVA	Vicuna-7B	284.17
7	Lynx	Vicuna-7B	279.24
8	Cheetah	Vicuna-7B	258.91
9	LLaMA-Adapter-v2	LLaMA-7B	229.16
10	VPGTrans	Vicuna-7B	218.91
11	Otter-Image	Otter-9B-LA-InContext	216.43
12	VisualGLM-6B	VisualGLM-6B	211.98
13	mPLUG-Owl	LLaMA-7B	209.40
14	LLaVA	MPT-7B	197.00
15	MiniGPT-4	Vicuna-7B	192.62
16	Otter	Otter-9B	180.87
17	OFv2_4BI	RedPajama-INCITE-Instruct-3B-v1	176.37
18	PandaGPT	Vicuna-7B	174.25
19	LaVIN	LLaMA-7B	97.51
20	MIC	FlanT5xl	94.09

Object Hallucination [🔗](#)

Rank	Model	Version	Score
🥇	Qwen-VL-Chat	Qwen-VL-Chat	90.00
🥈	Lynx	Vicuna-7B	86.33
🥉	InstructBLIP	Vicuna-7B	85.00
4	BLIP2	FlanT5xl	82.67
5	Cheetah	Vicuna-7B	77.00
6	BLIVA	Vicuna-7B	76.67
7	Otter-Image	Otter-9B-LA-InContext	74.00
8	Bard	Bard	70.67
9	InternLM-XComposer	InternLM-XComposer-7B	67.67
10	VPGTrans	Vicuna-7B	62.33
11	mPLUG-Owl	LLaMA-7B	61.00
12	LLaMA-Adapter-v2	LLaMA-7B	60.67
13	VisualGLM-6B	VisualGLM-6B	54.00
14	Otter	Otter-9B	53.33
15	PandaGPT	Vicuna-7B	53.00
16	MiniGPT-4	Vicuna-7B	50.67
17	MIC	FlanT5xl	50.33
18	OFv2_4BI	RedPajama-INCITE-Instruct-3B-v1	49.00
19	LLaVA	MPT-7B	49.00
20	LaVIN	LLaMA-7B	20.00

Model A

The image shows debris scattered across the floor, including wood planks, bricks, and a white safety helmet.

There is also a black trash bag and a broom on the ground. The scene appears to be in an alleyway with a concrete wall on the right side.

Potential safety hazards in this scene include:

- **tripping** hazards from the debris on the floor, the possibility of hidden obstacles or tools in the cluttered space,
- the risk of **falling** objects from the unfinished structure.

Workers at the site must exercise caution while navigating the area and ensure proper safety measures are in place to minimize the risk of accidents.



Model B

The scene depicts a cluttered and potentially hazardous urban environment. Here are the elements observed and the associated potential safety hazards:

Garbage and Debris: There are scattered **trash bags**, **wooden** materials, and **assorted waste**. This can lead to: Tripping hazards, Possible **spread of diseases** if there's organic waste or pests.

Pallet: A **wooden pallet** is seen in the middle, which may cause: Tripping hazards, Potential for splinters or cuts if the wood is damaged.

Liquid on the Ground: There appears to be a **wet or oily surface**. Slippery surface can lead to falls and injuries. The nature of the liquid isn't clear; it could be harmful if touched or ingested.

Exposed Metal Frames: **Rusty metal frames** are seen lying around. Risk of tetanus or other infections from cuts. Tripping hazards.

Overtured Objects: An **overtured plastic container** can be seen. Potential tripping hazard. If the container had contents, there's a risk of exposure to whatever it held.



Model B

Wires and Poles: Some **poles and potential wiring** are seen upright. Risk of electrocution if they are live wires. Falling hazards if not secured properly.

Poor Visibility: The area seems to be in a **dimly lit alley** or **narrow passage**. Reduced visibility can amplify all the other risks, as one might not see hazards until it's too late.

General State of Disrepair: The **walls and ground seem to be in a state of neglect**. Structural integrity might be compromised, leading to collapses or falling debris.

Recommendations for addressing these hazards:

Clear all debris and waste from the area.

Secure any loose wires or poles.

Dry up or clean any spilled liquids and check for the source to prevent future spills.

Check the structural integrity of the surroundings.

Ensure proper lighting for the area.

In conclusion, the scene indicates a need for immediate attention to safety standards and cleanup to prevent potential accidents or health issues.



Model A



HKU
MMLAB



The image shows a **large hole in the ground** with a **shovel** placed in it. The hole appears to be **filled with water** and there is **a large amount of debris** scattered around it.

From a safety perspective, the most significant hazard in this scene is the large hole in the ground.

If someone were to **accidentally step in or fall into the hole**, they could suffer serious injuries.

The presence of the **shovel** in the hole also suggests that **someone may have been working in the area**, which increases the risk of someone tripping or falling over the shovel or the debris.

Additionally, there are several **bottles and other pieces of litter** scattered around the hole and the surrounding area. These items can create additional hazards, as they may be trip hazards or cause accidents if someone were to step on them.

Overall, it is important to exercise caution in this area and ensure that the hole is properly fenced off and marked as a hazard to prevent accidents.



Model B



The scene shows a **broken** or **dislodged manhole cover** on an asphalted surface. Here are the elements observed and the associated potential safety hazards:

Broken/Dislodged Manhole Cover: The primary concern in the image.

Tripping Hazard: Someone could easily trip over the uneven surface or the edges of the manhole.

Falling Hazard: There's a risk of someone **stepping onto the dislodged cover and falling into the manhole**, potentially leading to severe injuries.

Objects Falling: **Objects could fall into the manhole, potentially causing damage to underground utilities.**

Water or Liquid Inside the Manhole: There appears to be **stagnant water or some liquid inside.**

Slipping Hazard: If this liquid overflows or splashes outside, it could make the surrounding area slippery.

Health Hazard: Stagnant water can be a breeding ground for insects like mosquitoes, leading to the potential spread of diseases.

Debris around the Manhole: There are **small pieces of garbage and debris** scattered around.

Tripping Hazard: These small items can be overlooked and cause someone to trip.

Environmental Concern: Debris entering the sewer system can lead to blockages and other issues.



Model B

Cracked Surface: The asphalt around the manhole appears to be cracked and in poor condition.

Tripping Hazard: The uneven surface can cause people to trip.

Infrastructure Concern: The state of the surface suggests that there might be deeper infrastructure or maintenance issues that need addressing.

Recommendations for addressing these hazards:

- Immediately secure the area around the broken manhole cover to prevent people and vehicles from approaching.
- Replace or correctly position the manhole cover as soon as possible.
- Clean up the debris and garbage surrounding the area.
- Assess the state of the asphalt and consider repairs or resurfacing to prevent further degradation.
- Regularly inspect such utility access points to ensure their safety and proper maintenance.





Model A

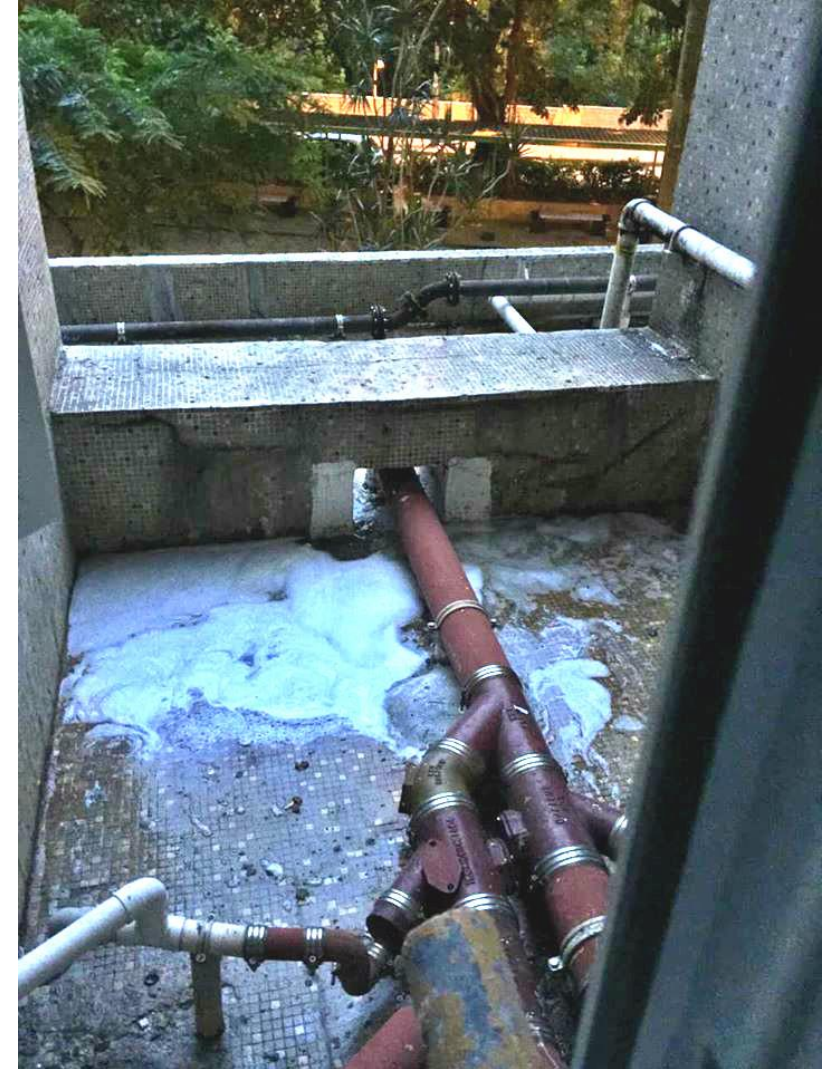
The image shows a concrete walkway with a few pipes running through it.

One of the pipes is red and has several metal clamps on it.

There is a large amount of foam on the ground near the pipes, which could be a safety hazard as it could cause individuals to slip and fall.

Additionally, there is a pile of tools on the ground, including a hammer and a screwdriver, which could pose a risk of injury if someone were to step on or come into contact with them.

Overall, the scene highlights the importance of keeping construction sites organized and clear of clutter to minimize the risk of accidents.



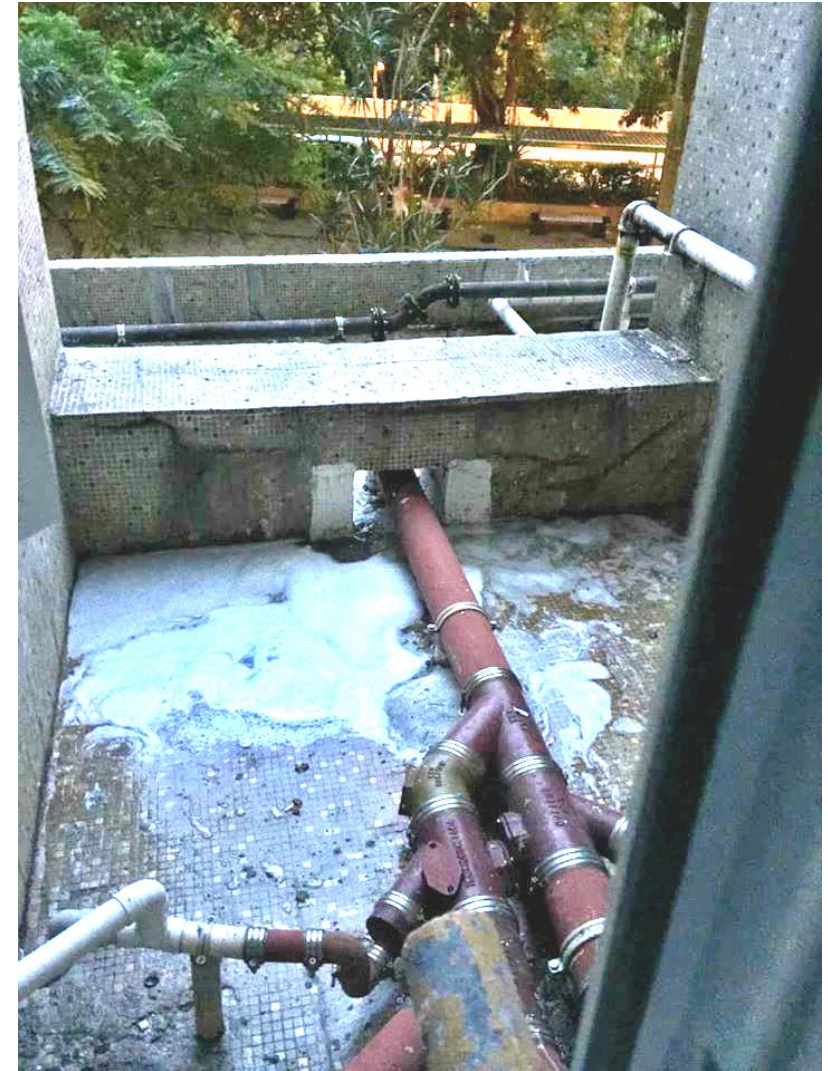
Model B

The scene appears to depict an **outdoor or semi-outdoor area**, possibly a **service or utility section of a building** or structure, with visible **plumbing and piping systems**. Here are the elements observed and the associated potential safety hazards:

Piping System: There's an extensive piping system, both **vertical and horizontal**, with varying **sizes and materials**. **Leakage Risk:** If any of the pipes or joints are not properly secured, there's a risk of leakage which can damage nearby structures and pose safety hazards. **Impact Hazard:** Protruding pipes could pose a hazard if someone were to accidentally walk into or collide with them.

Foamy Substance on the Ground: A white, foamy substance is visible on the ground. **Slipping Hazard:** The substance might make the floor slippery, posing a risk for slips and falls. **Chemical Exposure:** Depending on the nature of the substance, there could be risks associated with **skin or respiratory exposure**.

Debris: There seems to be some **debris** or loose particles on the ground. **Tripping Hazard:** Small objects and debris can pose a tripping risk.



Model B

Uneven Surface: The concrete or tiled surface appears to be uneven and possibly damaged in places. **Tripping Hazard:** Uneven surfaces can lead to trips and falls.

Exposed Brickwork: The walls or structures show **exposed brickwork or masonry**. Structural Concerns: Exposed brickwork may suggest potential structural issues or unfinished construction.

Vegetation: There are **trees and greenery visible** in the background. **Falling Debris:** Trees can drop branches, leaves, or other debris which might add to the clutter and pose tripping hazards.

Recommendations for addressing these hazards:

- Regularly inspect the piping system for any signs of wear, tear, or potential leaks.
- Clean up the foamy substance and investigate its source to prevent future occurrences.
- Regularly clear debris from the area to maintain a safe walking surface.
- Consider repairs or renovations for uneven or damaged flooring.
- If the exposed brickwork is a sign of unfinished construction, ensure it's completed or properly secured.
- Regular pruning or maintenance of the vegetation can reduce the risk of falling debris.



Model A



HKU
MMLAB



The image shows two construction workers wearing safety gear, one of them is wearing a yellow vest, standing on a large scaffold.

The workers are positioned at the top of a staircase, which appears to be the main access point to the scaffold.

From a safety perspective, the workers are taking the necessary precautions by wearing safety gear and using proper equipment.

However, there are potential safety hazards that should be considered.

One potential hazard is the risk of falling from the scaffold, which can be mitigated by using safety harnesses and adequate guardrails.

Another potential hazard is the risk of electrical shock from the nearby power lines, which should be addressed by proper grounding and isolation measures.

Additionally, proper training and supervision of the workers should be provided to ensure they are aware of the potential hazards and can work safely on the site.



Model B

The scene showcases a **multi-level scaffolding structure** at a construction site. **Two workers** are visible on the scaffolding:

Scaffolding Structure: A **metal** scaffolding system, which is commonly used in construction to provide temporary elevated platforms for workers and materials.

Stability Concerns: If not properly anchored or if overloaded, scaffolding can become unstable, risking a collapse.

Workers:

Worker 1 is seen wearing a hard hat, **high visibility vest**, and possibly safety goggles.

Worker 2 is wearing a hard hat, **sunglasses**, a high visibility shirt, and **appears to be secured with a harness**.

Fall Risk: Given the height, there's always a risk of falling from the scaffolding, which can result in severe injuries or fatalities.

Dangling Equipment: Worker 2 has some equipment attached which, if not properly secured, could drop and pose a risk to workers below or damage the property.

Equipment and Materials: There's a piece of machinery or **equipment** to **the right with the label "specialized concrete pumping."**

Struck-by Hazard: If the equipment malfunctions or isn't operated correctly, it can pose a hazard to workers.

Overloading: Adding too much weight on the scaffolding, either by workers or equipment, can cause it to become unstable.



Model B



HKU
MMLAB



Safety Harness: It appears that Worker 2 is using a **safety harness**, which is an essential safety measure when working at heights.

Improper Usage: If the harness isn't attached correctly or is not inspected regularly for wear and tear, it might not provide adequate protection in case of a fall.

Recommendations for addressing these hazards:

- Ensure the scaffolding is erected, modified, and dismantled by competent individuals and according to manufacturer's guidelines.
- Regularly inspect the scaffolding for structural integrity and stability.
- Ensure all workers are trained in the safe use of scaffolding and are aware of potential hazards.
- Ensure that workers wear appropriate personal protective equipment (PPE) and that it's in good condition.
- Regularly inspect harnesses and other fall protection equipment for signs of wear and tear.
- Establish and enforce a tool tethering policy to prevent tools and equipment from falling.
- Limit the number of workers and the amount of equipment/material on the scaffolding to avoid overloading.



Model A

The image shows a construction site with multiple vehicles and construction equipment, including several yellow vehicles and a large crane.

In the center of the image, a yellow truck is in a state of collapse, with its front and side portions falling down, and its roof is also damaged.

Several people are gathered around the site, some near the vehicles and some further away.

Potential safety hazards in this scene include the unstable and damaged truck that could pose a risk of falling or collapsing, especially if it is not secured properly.

Additionally, the presence of multiple vehicles and construction equipment in a relatively confined space could also increase the risk of accidents if proper safety measures are not taken.

It is essential for the workers to be cautious and follow safety protocols while working in such environments to minimize the risk of accidents and injuries.



Model B

The scene presents a construction site with several significant elements:

Tipped Crane: The most prominent safety concern is the yellow crane that appears to have tipped over, with its back wheels off the ground. This suggests a major accident or a near-miss situation.

Equipment Damage: The tipped crane could damage other equipment or structures.

Risk to Workers: Falling or unstable equipment poses a direct threat to workers in the vicinity.

Second Crane: Another crane (white) is situated near the yellow crane. It's extended towards a building, and its positioning suggests it might be operational.

Collision Risk: Close proximity of two large pieces of equipment increases the risk of collision.

Overloading: Using multiple heavy pieces of equipment in the same area can stress the ground or structures below.

Workers: Several workers are visible in the scene, some in close proximity to the tipped crane.

Direct Threat: Workers near the unstable crane face an immediate risk.

Lack of PPE: Some workers seem to be without hard hats or other visible safety equipment, making them more vulnerable to falling objects or other hazards.

Construction Debris: There are scattered construction materials, tools, and debris around the site.

Trip Hazard: Unorganized sites can lead to tripping hazards.

Struck-by Hazard: Loose tools or materials can become projectiles if knocked or if there's an equipment malfunction.



Model B

Pedestrians: There appear to be non-construction personnel or pedestrians on the road adjacent to the construction site. External Risk: They might be at risk from any unforeseen accidents or debris coming from the construction site.

Proximity to Buildings: The construction site is close to a large building. Structural Damage: Any mishap, especially with the cranes, could cause damage to the nearby building, posing a risk to its occupants.

Recommendations:

- Investigate the cause of the tipped crane and take corrective measures to prevent future incidents.
- Regularly inspect all equipment for functionality and safety.
- Clear and organize the construction site regularly to reduce tripping hazards and improve overall safety.
- Ensure all workers have and wear appropriate personal protective equipment (PPE) at all times.
- Consider installing barriers or clear demarcation for construction zones to prevent unauthorized entry and protect pedestrians.
- Regularly train workers on safety protocols, especially when working with or around heavy machinery.
- Monitor the weight and positioning of equipment to ensure the ground or structures are not overloaded.



Vision Generative AI: A New Frontier for Sewage Treatment Industry

Thank You!