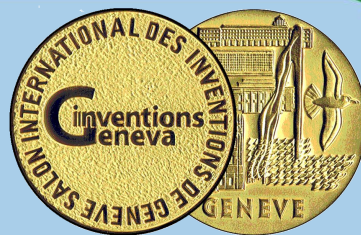


香港建造業議會  
建造業安全周 2024 研討會



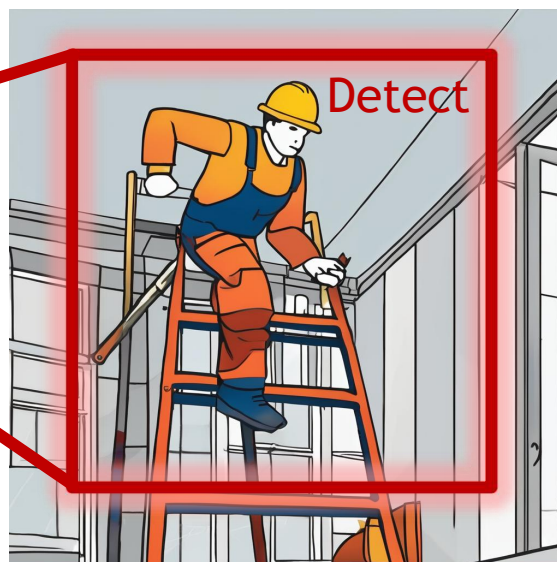
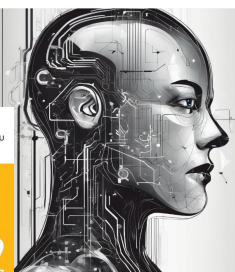
# 工地智能安全監測 — 視覺語言模型 的應用

黃子健  
渠務署高級工程師/機電工程5/研究及發展1  
2024年8月26日



# 視覺語言模型在工地安全監測中的應用

- 實時監察工地安全
- 將安全規則現場影像分析系統，協助安全人員監管工地安全
- 如果發現有安全違規行為，會通知安全人員



# 電腦視覺 (Computer Vision) vs 視覺語言模型



## 電腦視覺

擅長識別、分類及跟蹤物體

現有的人工智能  
工地安全監管系統

## 大型語言模型

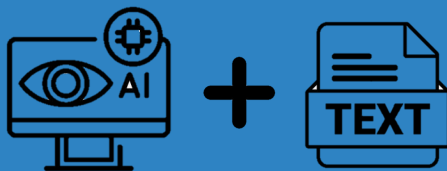
語言理解及語義推理  
(Semantic Reasoning)

Chat GPT-4



## 視覺語言模型

電腦視覺與大型語言模型  
的融合



虛擬工地  
安全助理

# 電腦視覺 VS 視覺語言模型



## 電腦視覺

擅長識別、分類及跟蹤物體

現有的人工智能  
工地安全監管系統

## 大型語言模型



語言理解及語義推理

## 視覺語言模型

電腦視覺與大型語言模型  
的融合



虛擬工地  
安全助理

# 人工智能工地安全監管系統

## 深度學習 (Deep Learning)

### 預訓練與微調:



標籤的圖像

深度神經  
網絡

物體檢測、分類等

應用示例: 請找出所有的貓。



深度神經  
網絡  
(已訓練)

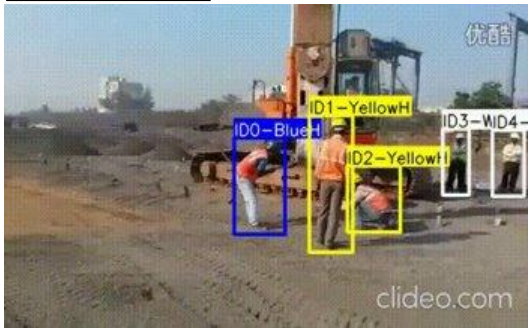


預測概率

$P_{[\text{Dog}]}$
$P_{[\text{Cat}]}$
$\vdots$
$P_{[\text{Cup}]}$

# 電腦視覺的局限性

模型1



個人防護裝備違規偵測

模型2



被撞預警

模型3



跌倒姿勢識別

模型4



火災/煙霧偵測

- 偵測物品的存在與缺失
- 偵測越界

- 難以擴展
- 特定場景及任務



安全警示: 沒有安全帶!!

缺乏推理能力

# 電腦視覺的局限性

電腦視覺的局限性：

- 只能輸出**基本**的結果(如物體、位置等)
- 無法檢測需要**高級語義推理 (Semantic Reasoning)**的違規行為



未偵測到違規！

上落時，**必須面向**  
輕便工作台 …

(參考資料：建造業議會  
「高空工作安全手冊」)



# 電腦視覺 vs 視覺語言模型



## 電腦視覺

擅長識別、分類及跟蹤物體

挑戰：

- ⊖ 無法理解圖像環境背景
- ⊖ 依賴訓練數據
- ⊖ 缺乏常識推理能力



## 大型語言模型

語言理解及語義推理



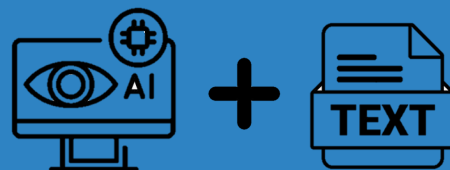
語義推理能力

任務語言理解能力



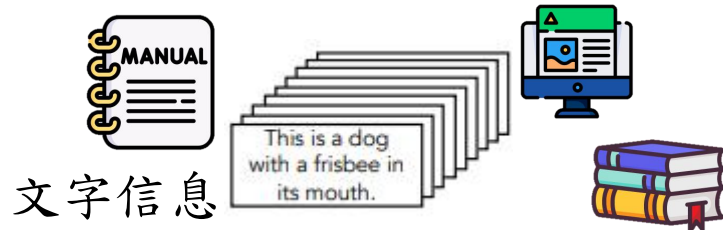
## 視覺語言模型

電腦視覺與大型語言模型  
的融合

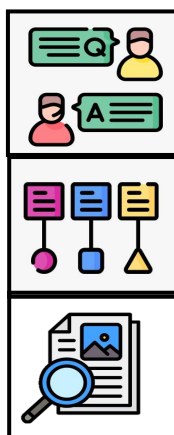


虛擬工地  
安全助理

# 大型語言模型



## 大型語言模型



問題回答

文本分類

信息檢索



GPT-4

分享

Please list 5 safety requirements of work at height activities in construction sites in short.

GPT-4 Poe

Here are five concise safety requirements for work at height activities on construction sites:

### 1. Risk Assessment and Planning:

- Conduct thorough risk assessments and establish a detailed plan including emergency rescue procedures.

### 2. Use of Personal Protective Equipment (PPE):

- Equip workers with appropriate fall protection gear such as safety harnesses and ensure regular inspection and maintenance of the gear.

### 3. Training and Competence:

- Provide comprehensive training and certify workers on safe practices and emergency procedures for working at heights.

### 4. Safe Access and Egress:

- Ensure stable and secure access points and platforms for all work at height activities.

### 5. Regular Inspection and Maintenance:

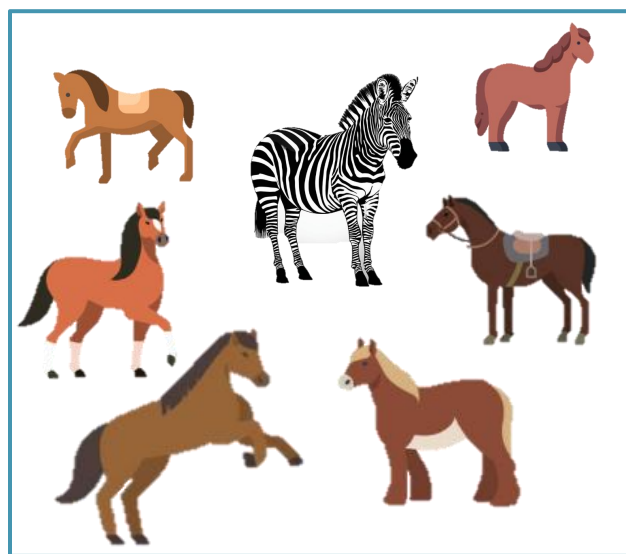
- 自然語言交流
- 類人的「虛擬助手」
- 語義推理能力

# 視覺語言模型

視覺語言模型是帶有視覺功能的大型語言模型

提示:

「有黑白斑紋的馬就是斑馬」。請找出所有的斑馬。

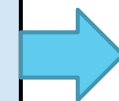


## 視覺語言模型

馬



數據庫中沒有斑馬的圖像，只有馬的圖像。



輸出:



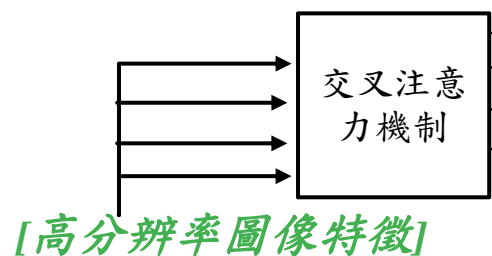
根據提示，雖然數據庫中沒有斑馬的圖像，但視覺語言模型能理解「有黑白斑紋的馬就是斑馬」這個語義推理，並找到斑馬圖像。

身教勝於言教？  
(Example is better than precept.?)

# 解決方案 - 虛擬工地安全助理

## 模型架構

### 高分辨率交叉模組



高分辨率圖像

### 原視覺語言模型



### [圖像特徵序列]

編碼器(圖像+視覺自注意力模型)



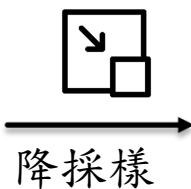
低分辨率圖像

### [文本特徵序列]

詞向量

上落時，必須面向輕便工作台，並保持三點接觸。  
When ascending or descending, worker should face to the working platform and keep 3 points contact.

文本



當安裝、改動或拆除樓邊/孔洞/窗邊等防護措施時，應提供適當的防護系統予工友。  
Whilst installing, altering or dismantling fall protection facilities at the floor edge, opening and windows, suitable fall arresting system should be provided to workers.

工作平台應在穩固、平坦和水平的地面上。裝置工作平台的周圍應無廢物和雜物。  
Working platforms should be erected on firm, even and level ground. The surrounding of working platforms should be kept free from waste and miscellaneous materials.

在搭建、更改、拆卸棚架、或設置安全工作平台或提供安全進入並不切實可行時，須配戴全身式安全吊帶，並繫於穩固的繫穩點或獨立救生繩上。  
When erecting, altering, dismantling of scaffolds or it is impracticable to erect a working platform or provide safe access and egress, the use of full body safety harness attached to a secure anchorage point or an independent lifeline is required.

Safety Manual (2018)

(參考資料:建造業議會「高空工作安全手冊」及渠務署「安全手冊」)

# 從「通才」到「專才」

模型微調：

從「通才」到「專才」  
建築知識和**安全意識**

視覺語言模型優化的  
**三階段**方案

## 第一階段：描述性微調



針對**描述性**任務進行訓練：

從預訓練的模型到具有  
**工地知識**的模型

## 第二階段：以安全為導向的指令微調

Is this site safe?

Not completely, there is one worker not wearing harness, which is regulatory when working at height

Is this site safe?

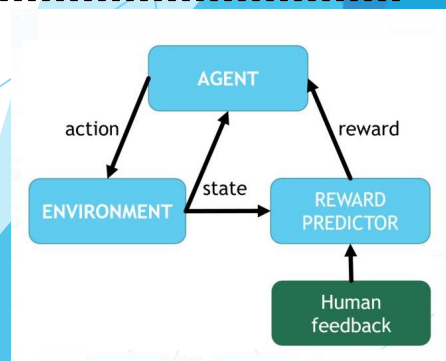
The site shown in the picture does not appear to be completely safe due to the absence of safety harnesses or ladders, which are essential for workers' safety when working at heights.

針對**高階推理能力**進行訓練：

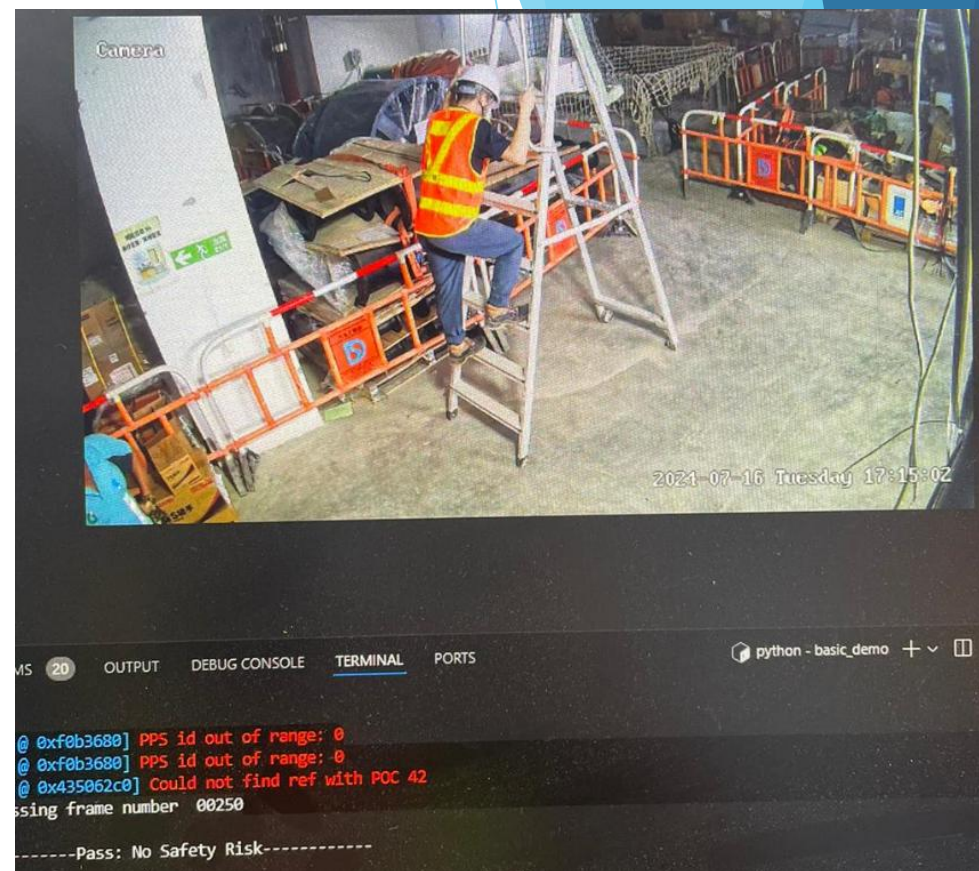
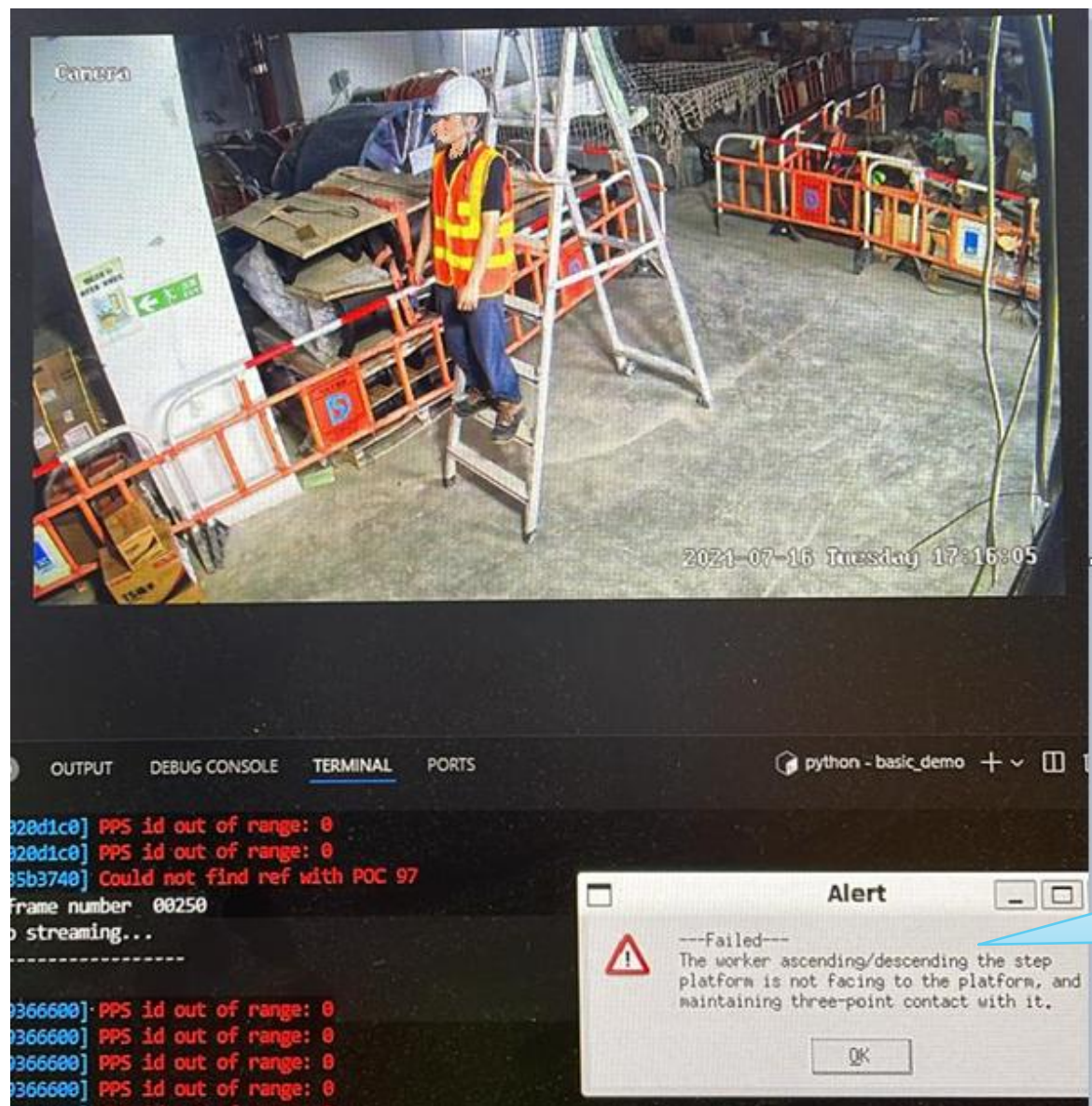
具有**工地知識**  
及**安全意識**的模型

## 第三階段：強化學習與人類反饋 (RLHF)

由實際**從業者**進行持續優化



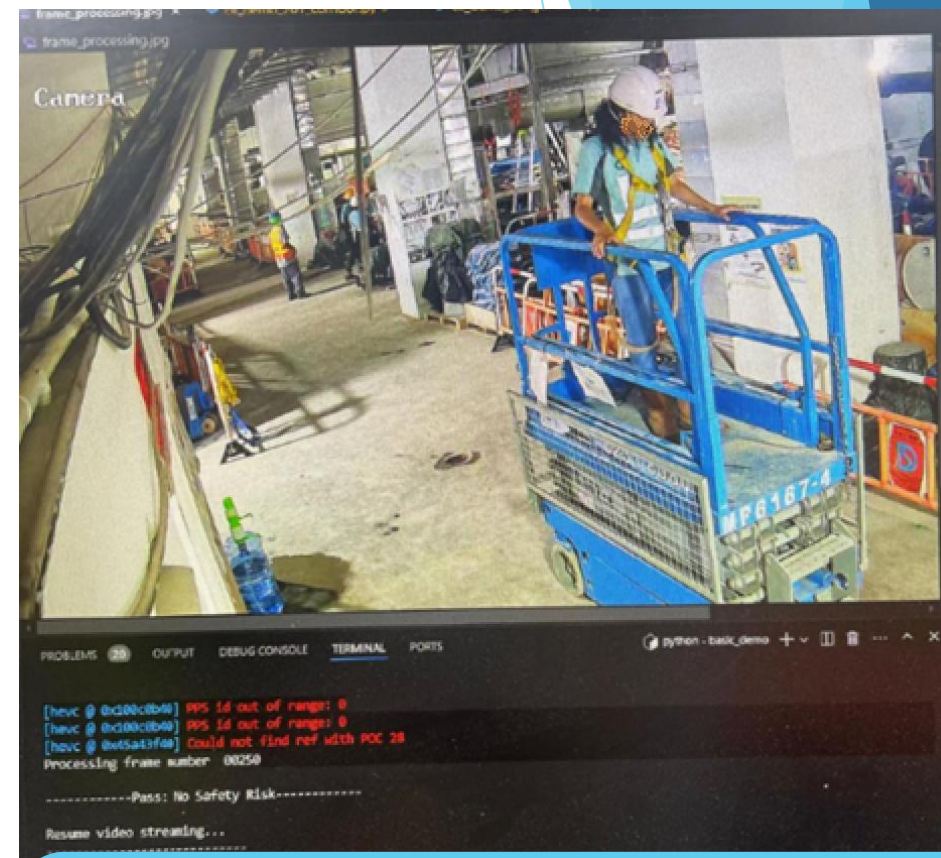
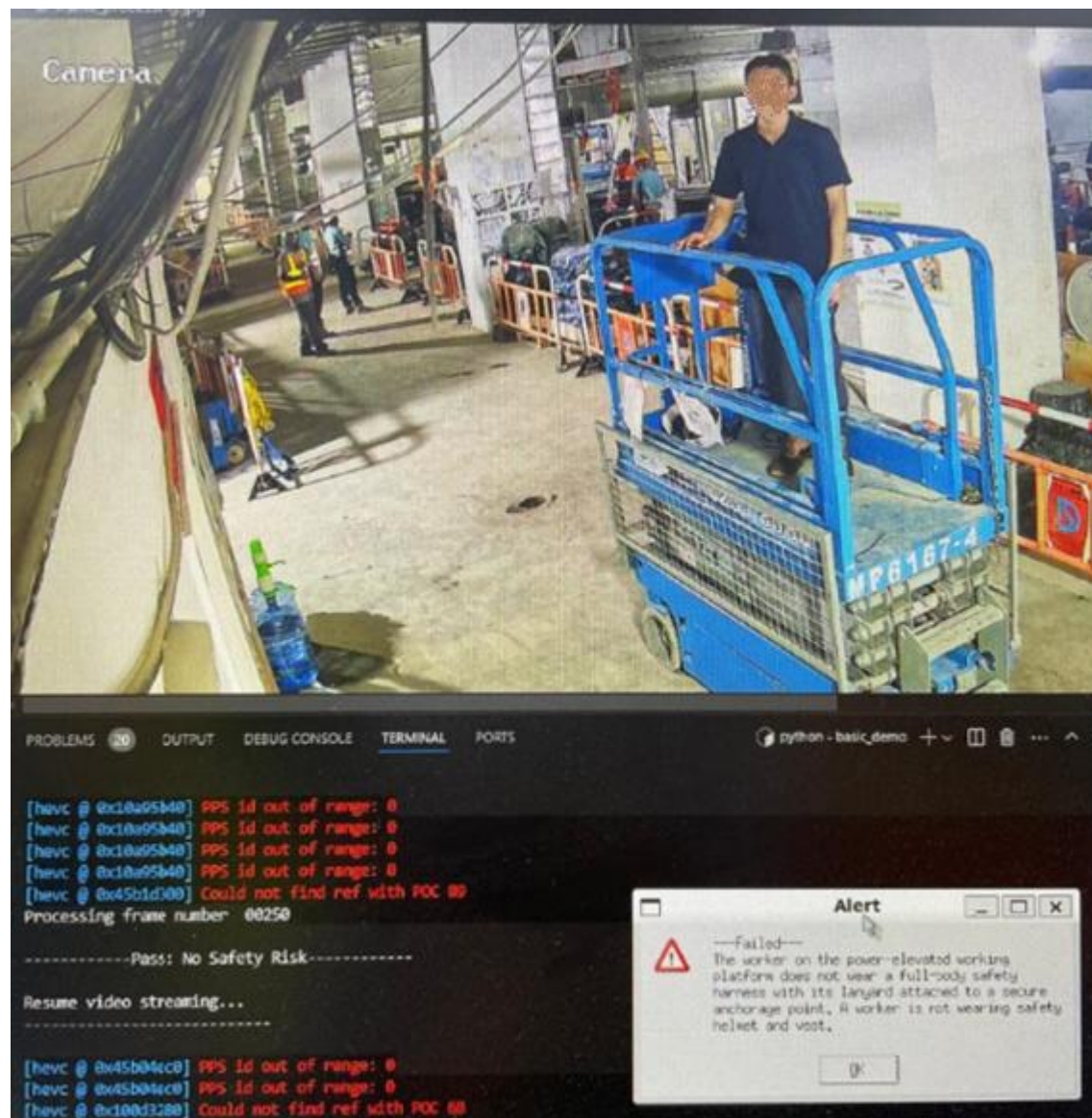
# 示範：安全使用梯子



## 警報

正在上落階梯式工作台的工人沒有面向工作台，也沒有保持三點支撐。

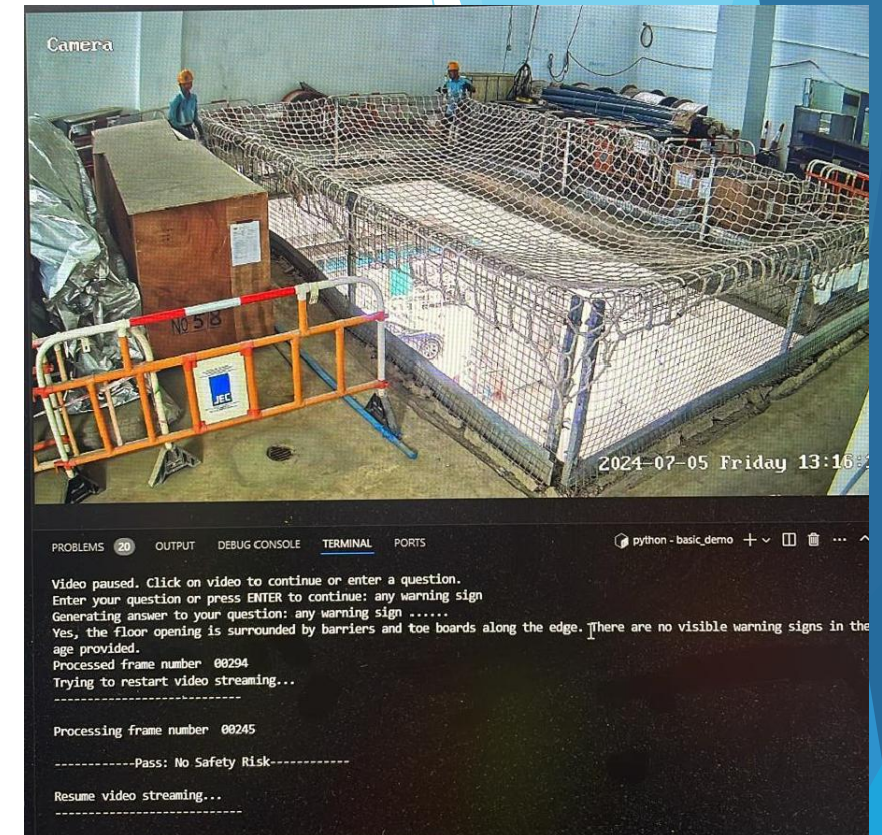
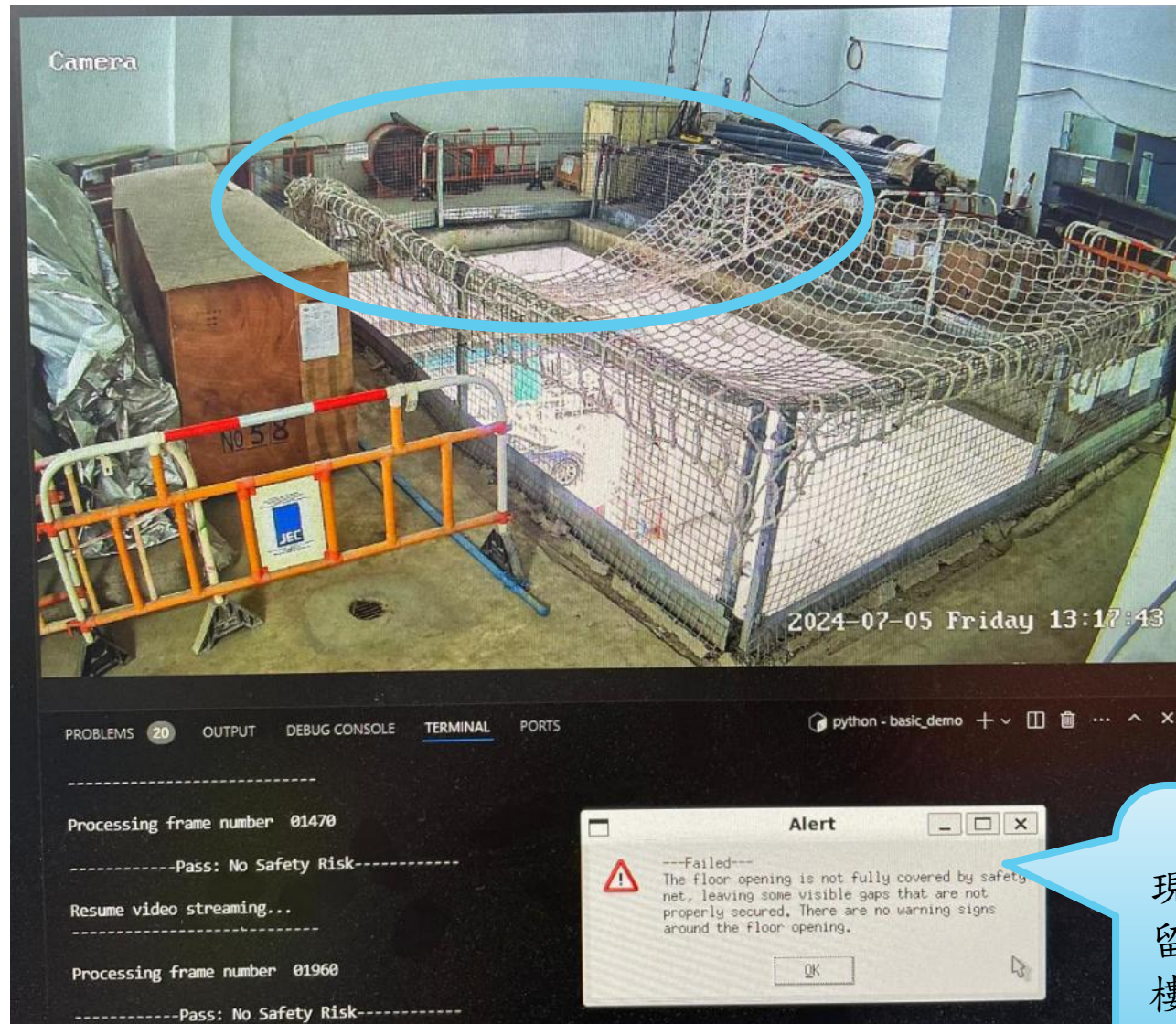
# 示範：高空作業



## 警報

在高架工作台上的工人沒有配戴全身安全帶，並將安全帶固定到可靠的錨固點。工人也沒有配戴安全帽和穿著反光衣。

# 示範:樓面洞口



## 警報

現場樓面洞口未完全蓋上安全網,  
留有一些可見的間隙未妥善固定。  
樓面洞口周圍也沒有設置警示標誌。



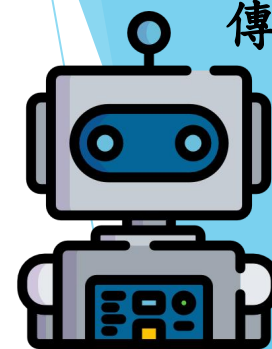
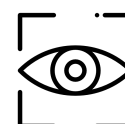
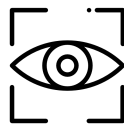
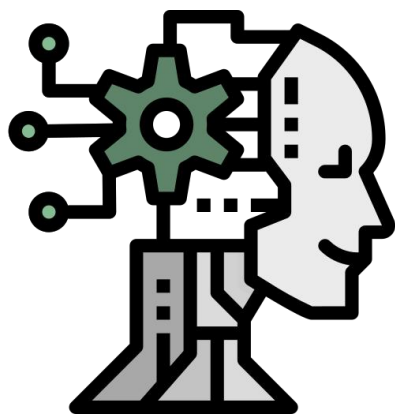
# 虛擬工地安全助理 - 視覺語言模型

「理解」有關建築的安全法規和指引的內容



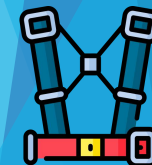
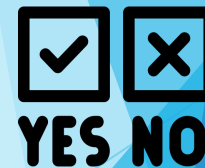
文字理解 + 影像分析

## 識別危害及不安全行為



傳統人工智能

只分析工作安全有關物件是否存在

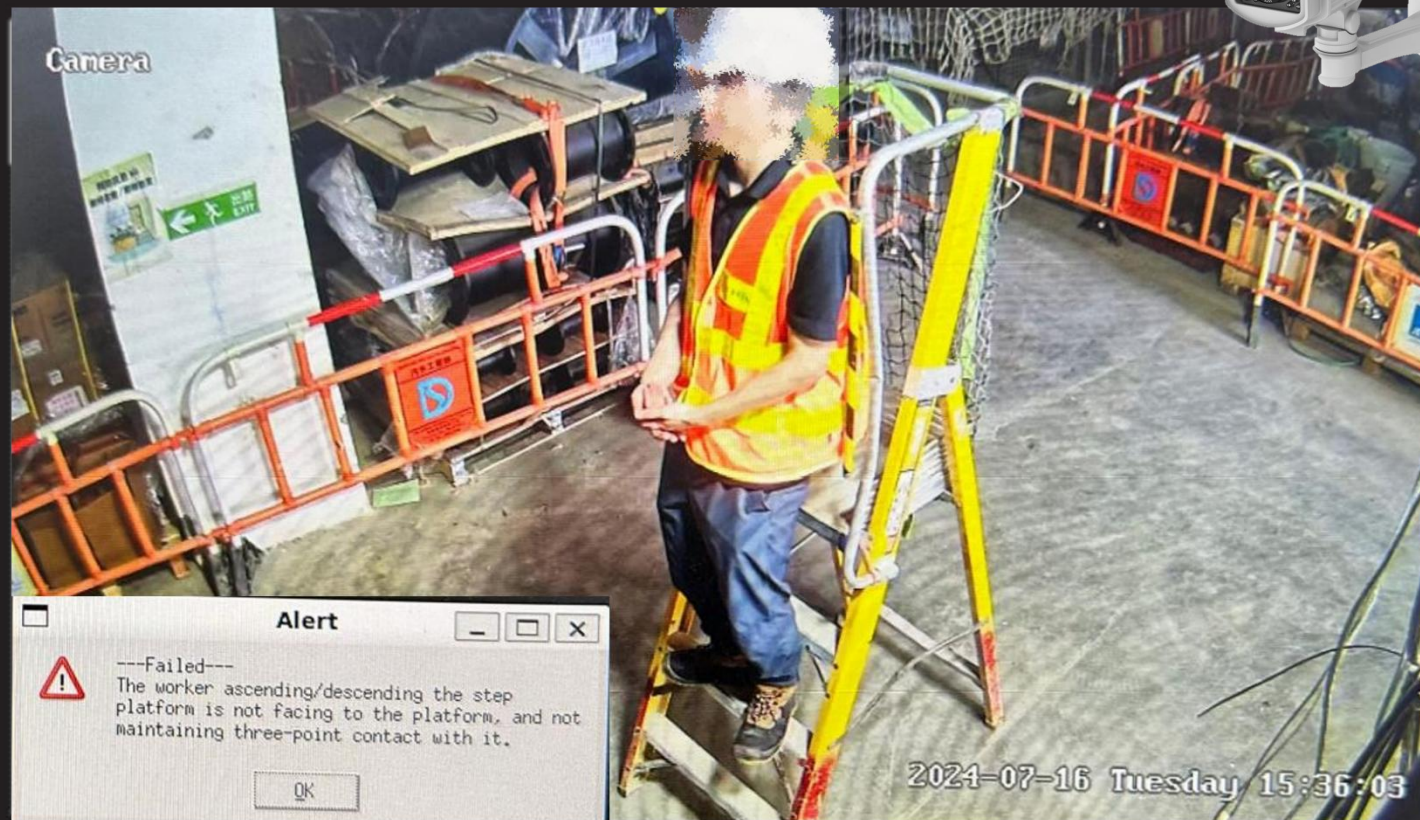




將圖像與相關文本  
進行配對



準確地分析工地  
實況



### 警報

工人沒有面向工作平台，亦沒有  
保持三點接觸。



及時對工地實況  
發出警示



提出更具針對性的  
理解



多謝