Final Project

1. **AI 的未來能力**

   假如存在未知且能預見對生物身體有危險的環境，比如說星球間的探索建設，與其讓人類穿上太空服在無氧的環境中作業，不如讓 AI 機器人工作來得好。自古以來，人類就在擴張生存範圍，未來往太空擴張也是能預期的，這也是我認為這個能力對人類社會有重大意義的原因。為了達到這個目的，AI 需要自行應對未知新環境的能力，建設新星球時遇到異常天候、落石、輻射強化等等問題，無法預先建立模型，AI 需要自行診斷問題和調整方案。

2. **所需的成分與資源**

   我認為這個 AI 需要從多種即時感測資料(需要各種對應的感測硬體)中學習，例如溫度、壓力、地形等等，這些資料不可能事先標籤，需要自監督式學習或非監督式學習建立環境的模型。需要的工具可能有神經網路、PDE 模擬溫度、壓力等等，和規劃策略的演算法、強化學習的回饋訊號。

   使 AI 能透過自監督式學習建立環境模型，加上強化學習更新行動的策略，在星球上自主工作。

3. **涉及的機器學習類型**

   我認為達成這個能力需要自監督式或非監督式學習和強化學習的組合。

   使用自監督或非監督式學習，輸入來自源於新環境中沒有標籤的資料，模型需要自行找出資料中的規律，讓 AI 理解新環境；結合強化學習，透過與環境互動得到的回饋（目標訊號），更新行動策略，以達到於新情境中自行學習並解決問題的能力。

4. **可實作模型**

1. 問題設計：

   設定一個存在 4 顆燈泡(編號 0,1,2,3)與 4 個開關的環境，機器人只能觀察到現在燈泡的狀態(一個四維向量)。每次機器人會選擇按下一個開關，然後回到新的燈泡狀態並得到正一(0 號燈泡亮起)或是負一(其它)的獎勵。

   目標期望機器人學會開關對應哪個燈泡亮起的規則，使它每一步做出得到預估最高獎勵的動作。

   機器人會輸入目前的觀察與動作，對每一個預選動作輸出下一步的預測燈泡狀態，並選擇預估獎勵最高的動作。

   透過上述的簡化模型，機器人有「從經驗中學到未知的規則以指導行動」能力的雛形，是未來希望達到 AGI(人工通用智能)的其中一項能力。

2. 模型與方法：

   首先建立一個環境，有四顆燈泡與四個開關，第一、二、三和四的開關分別打開燈泡 0,1、燈泡 2、燈泡 1,3、燈泡 0,2,3。(燈泡對應簡化的世界，開關對應機器人能做的動作)

再使用一個簡單的神經網路模型：

輸入層：8 個神經元對應燈泡狀態的四維向量與按下不同四個開關的四個動作。

Hidden layer：32 個神經元配上 Relu 作為激活函數。

輸出層：4 個神經元配上 sigmoid 激活函數對應 4 個燈泡明暗的機率。

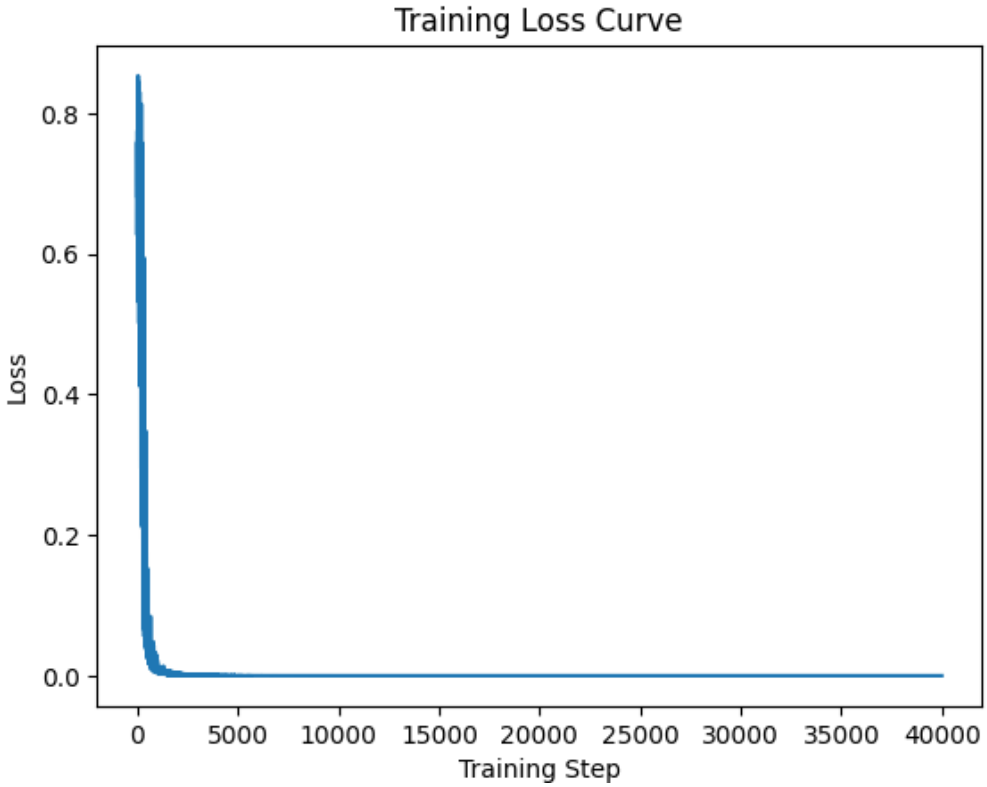
配合輸出是機率，所以 Loss function 使用 binary cross entropy。

期望這個預測燈泡狀態的機率的神經網路，能學會燈泡狀態的規則。

再來設定好機器人的策略(在最終目標需要 AI 自己形成策略)：

給定目前燈泡狀態，使用上述神經網路模型對每一個預選動作做模擬，若是下一步燈 0 的機率大於 0.5，則估計獎勵加一。若是估計獎勵更好，更新最佳動作。

3. 實作與結果：



上面兩張圖是 loss curve 和 reward 的比較。

學習過後的機器人相比於隨機亂按的機器人，一個的 reward 是 50，相當於每一步都得到獎勵，而另一個則是只有 2，接近於亂按的期望值 0。

這能明顯地看出此模型成功學到了開關對應哪些燈泡的規則。

4. 討論：

設計簡化模型，我一開始直覺是要使用 Unsupervised learning 去做，因為輸入資料不能經過標籤。後來我學到怎麼用強化學習，讓機器人與環境互動來對資料產生標籤，再用學過的監督式學習的方法，使機器人的神經網路學到如何預測下一步燈泡的狀態。

設計這個簡化模型時，我讓機器人已經知道了哪顆燈泡亮起可以得到獎勵。若是想要達到 AGI 的程度，在一個未知的環境中，應該把哪個現象的發生視作獎勵可能十分困難。還有這個模型的學習方法是透過多次試錯嘗試，來訓練神經網路，在現實世界中的環境規則可能很複雜，需要大量的嘗試才能學好，這個過程的成本會是個大問題。