

AI 是怎麼學會的？

兩個問題、一台機器、一段歷史

什麼是模型？

怎麼訓練模型？



今天只講兩件事

01

什麼是模型

一台有很多旋鈕的機器

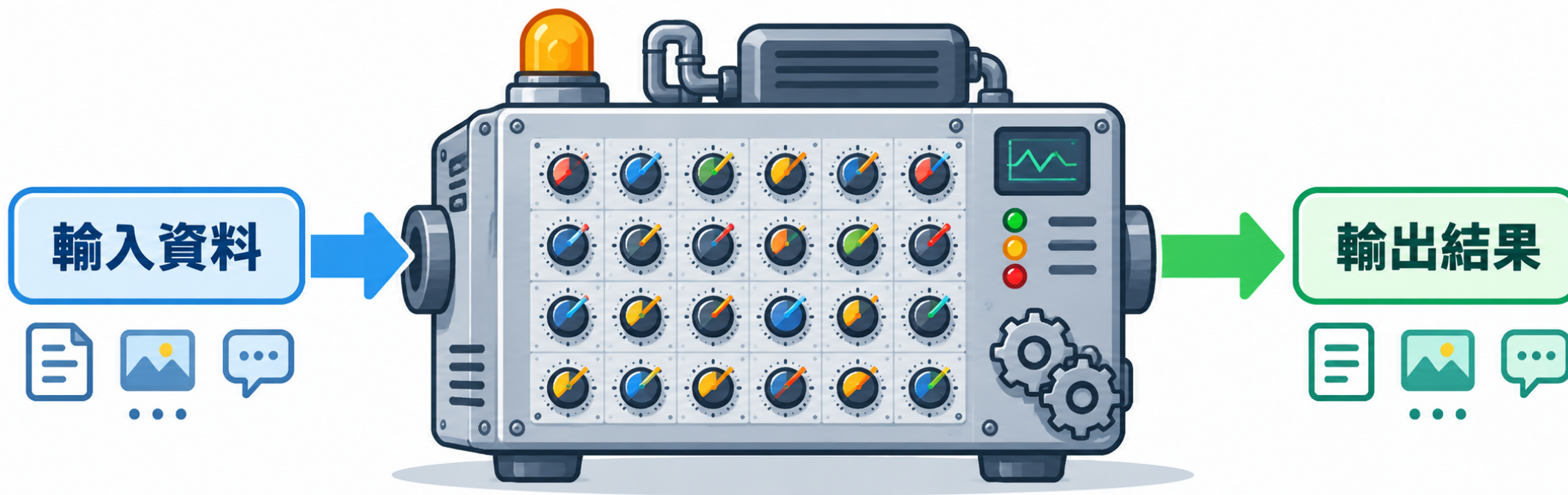
02

怎麼訓練模型

把旋鈕轉到對的位置——用 Enigma 的故事來講

模型，就是一台機器

資料從左邊進去，結果從右邊出來。



輸入

一句話、一張照片、一段聲音

機器

裡面是一排又一排的旋鈕

輸出

接下去的字、辨認的結果、生成的圖

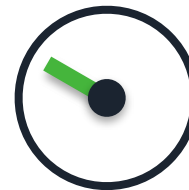
同一個輸入， 旋鈕位置不同， 出來的東西好壞就不同。

面板上的每一個旋鈕，專業上叫「參數」。

旋鈕決定機器做得好不好，不決定它做什麼——做什麼由輸入決定。

ChatGPT 這台機器，旋鈕有幾千億個。

輸入：寫一首關於月亮的詩



一首通順、有畫面的詩
接近我們想要的



押韻但不知所云
有點像，還不夠好



一堆亂碼
完全不行

問題來了：旋鈕該轉到哪裡？

幾千億個旋鈕，沒有人知道正確位置——連製造它的人也不知道。
所以需要一個方法，讓機器自己找。

這件事，八十年前就有人做過了。

1930-40 年代 · 德軍密碼機

Enigma

按一個字母，燈板亮出另一個字母。

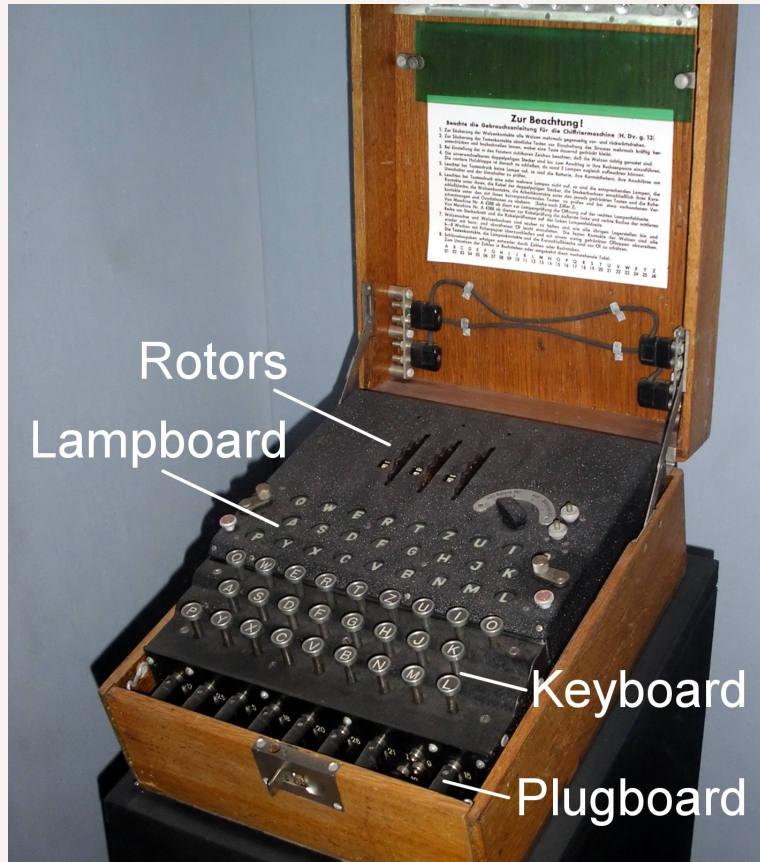
同一個字母再按一次，亮出來的又不一樣。

收訊的人有同一台機器，反著按回去就是原文。



秘密不在機器，在每天的設定

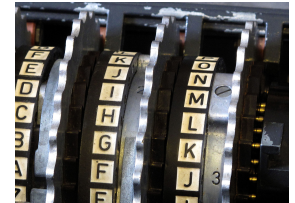
機器人人一樣，連盟軍都有。保密的是機器上三個可以動的地方，每天怎麼擺。



1

轉子 (Rotors)

5 個轉子，挑 3 個裝進去，左中右順序有差。



2

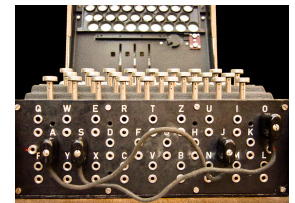
轉子的起始位置

每個轉子刻著 A-Z，裝好後還要各轉到指定的字母。

3

接線板 (Plugboard)

前面 26 個字母插孔，用 10 條線把 20 個字母兩兩對調。



機器人人一樣，連盟軍都有



算一算：一共有幾種設定？

1

轉子挑 3 個、排順序

第一槽 5 種選法，第二槽剩 4 種，第三槽剩 3 種

$$5 \times 4 \times 3 = \mathbf{60}$$

×

2

三個轉子的起始位置

每個轉子 26 個字母，三個各自獨立

$$26 \times 26 \times 26 = \mathbf{17,576}$$

×

3

接線板 10 條線

26 個字母挑 20 個、兩兩配對，這項最兇

$$\mathbf{150,738,274,937,250}$$

三項相乘

可能的設定共有

$$\mathbf{158,962,555,217,826,360,000}$$

種（約一億五千萬兆），德軍每天午夜換一組。

困難

每秒試

1,000,000 組

也要試 5,000,000 年

而盟軍每天只有 24 小時——午夜一到，設定就換了。
硬試是不可能的。

突破口：德軍太規律了

有些電文，內容是猜得到的。

每天早上六點的天氣報告

開頭一定是 WETTER（天氣）這個字

很多電文的結尾

一定是 HEIL HITLER

某個北非哨站每天回報

「無特殊狀況」——連續好幾個月

攔到的密文

Q F Z W R W I V T Y R E S X B F O G K U H Q B

猜到的原文

W E T T E R V O R H E R S A G E B I S K A Y A

這種「已知答案的電文」，破譯員叫它 **crib**。有了答案，就能檢驗一組設定對不對。

圖靈的機器：拿設定去對答案

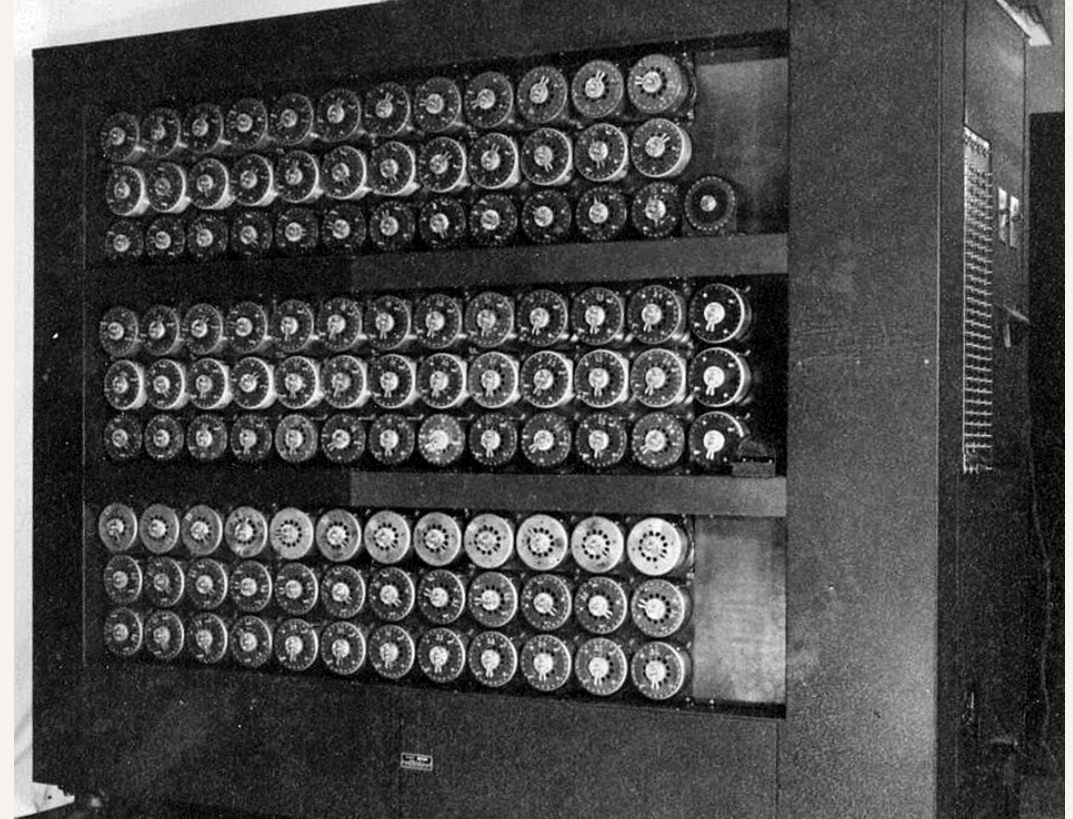


Alan Turing, 1936

設計了 **Bombe**，把「試」變成「淘汰」。

- 1 先假設一組設定
- 2 用它解 crib，解出來對不上就淘汰
- 3 一組設定錯，跟它同一族的幾千組一起出局
- 4 剩下的那一組，就是今天的答案

1940 年起，常常幾個小時就找出當天設定。



模仿遊戲



破解敵軍每天每則訊息

這，就是訓練模型

破解 Enigma

Enigma 的設定

已知答案的電文 (crib)

解出來對不上的字母數

找到今天的設定

訓練 AI 模型

模型的旋鈕 (參數)

訓練資料：一堆「輸入 + 正確答案」

誤差：機器猜的和正確答案差多少

訓練完成：旋鈕轉到誤差最小

訓練 = 用已知答案的資料，把旋鈕轉到誤差最小。

比喻有三個地方不一樣

① 誰在轉旋鈕

破解 Enigma 的是敵人；訓練 AI 的是機器的主人。

可以想成：你撿到一台 Enigma，還有幾封已經翻譯好的信。

② 旋鈕能不能慢慢轉

Enigma 的轉子動一格，整段輸出全變——只能淘汰，不能「越轉越接近」。

AI 刻意把機器做成「轉一點點、輸出只變一點點」，於是可以一路往誤差變小的方向轉。這是 AI 能學會的關鍵。

③ 有沒有唯一正解

Enigma 只有一組正確設定。

AI 沒有，誤差夠小就收工，所以同一個問題可以訓練出很多個都能用的模型。

今天的 AI：幾千億個旋鈕，同一件事

1

收集

一大堆「輸入 + 正確答案」的資料——網路上的文章、標好的照片

2

猜、對答案、量誤差

讓機器先隨便猜，和正確答案比，看差多少

3

轉旋鈕

往誤差變小的方向轉一點點，重複幾百萬次

模型＝一台有旋鈕的機器；訓練＝用已知答案把旋鈕轉到對的位置。

圖靈八十年前用淘汰做到了，今天的 AI 用「往好的方向轉」做到，規模大了幾兆倍。