
前言

「要當綠色真不容易。」

— Kermit the Frog 譯註

2022 年的政府間氣候變遷專門委員會（IPCC）報告，已經證實氣候變遷真實存在。世界各國開始採取行動，企業也需要跟上能源轉型的步伐，否則可能被淘汰。不幸的是，正如睿智的文化偶像所說，為了實現可持續發展，所需的變革絕不會輕鬆。

好消息是，大多數公有雲服務供應商已經承諾實現雲端淨零營運，當然還需要持續監督。我們可以向他們以及其他行業中的可持續領導者學習，並模仿他們的做法；並且有些工具已經開源或以商業化模式而可供選用。

這對我們來說是件幸運的事，因為很快地，客戶、基礎設施提供商、高昂的帳單和即將推出的法規等，可能就會逼著我們設立並達到自己的碳排放目標。所以，軟體開發和營運需要怎樣的改變，才能拯救我們的地球和公司呢？

這本書旨在幫助回答這個問題。《綠色軟體開發》涵蓋從國家電網應對可再生能源發展的方式，到改變營運、能源轉型對開發者日常生活的影響等。你可能會注意到，書中引用許多曾在大型科技公司工作的人所說的話，這並不是說他們改變立場，或是這些公司並不在意這議題；他們純以個人身分發言，不用受某些大型企業的公關規則約束。能聽到未經過濾的觀點總是件好事，因為從新手開發者到老練的首席技術長，每個人都有義務一同塑造未來。

譯註 Kermit the Frog 為 Walt Disney 旗下的布偶角色。

綠色軟體簡介

我生氣時沒有人會喜歡我。

— 布魯斯·班納（Bruce Banner）博士，《綠巨人浩克》

環保及永續倡議人士會生氣的原因不難想像，因為很少有行業能快速支援能源或綠色轉型，科技業也不例外。

但當我們關注綠色軟體議題時，一切就已經開始改變了。

在 IT 領域實現綠色環保的意義

根據綠色軟體基金會（GSF，<https://greensoftware.foundation>），綠色軟體（Green Software）或可持續軟體（sustainable software）的定義，是運行時造成碳排放最少的軟體。換句話說：

- 綠色軟體運行時，不需要太多能源功耗和硬體需求，就能完成單位工作，這叫做碳效率（Carbon efficiency）。因為產生電力和製造硬體通常會導致碳排放。
- 綠色軟體具備將運作狀態，依風力、太陽能、地熱能、水力或核能等低碳電力的時間和地點靈活調整的能力。或者可在電網電力碳密集時減少操作。例如，在只能使用煤電時降低資訊服務品質，等切換至低碳電力時再恢復，這可稱為碳意識（Carbon awareness）。

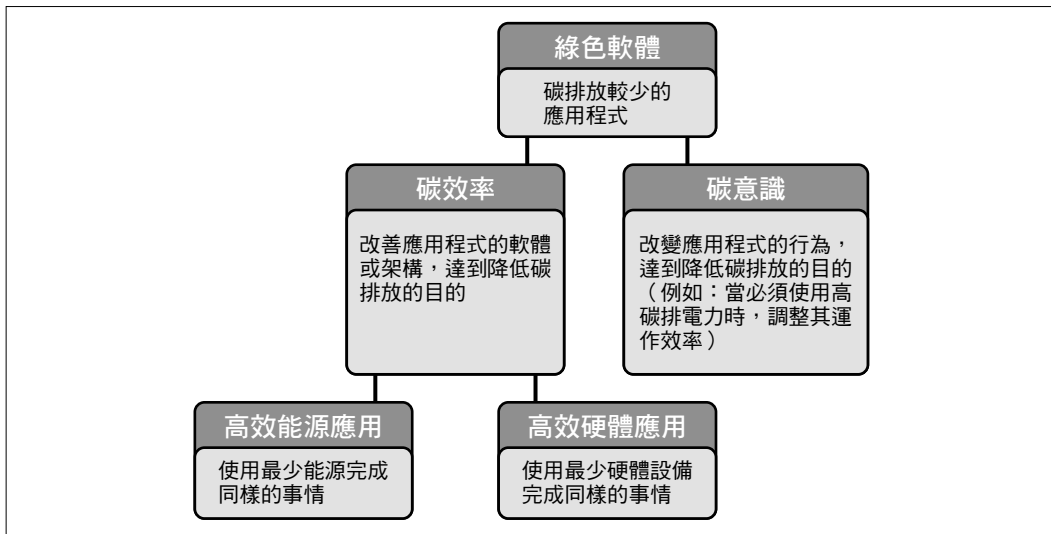


圖 1-1 綠色軟體基金會對綠色軟體的定義

在知道綠色軟體的定義後，又該如何開發與建構呢？

我們的想法

本書由 13 個技術章節組成：

1. 綠色軟體簡介
2. 設計模塊
3. 程式碼效率
4. 營運效率
5. 碳意識
6. 硬體效率
7. 網路
8. 更環保的機器學習、人工智慧和 LLMs
9. 測量
10. 監測
11. 共同效益
12. 綠色軟體成熟度矩陣
13. 我們該何去何從？

接下來將介紹每個章節，並提供關鍵重點。

營運效率

抵抗是徒勞的！

— 博格人（The Borg），電影《星際爭霸戰》（*Star Trek*）

實際上，抵抗並非徒勞無功。情況比這更糟。

與機器的戰鬥：運行效率與綠色技術的挑戰

總有一天，超導伺服器（<https://oreil.ly/gQHoi>）將在極低溫（<https://oreil.ly/BRDsl>）的資料中心（<https://oreil.ly/Rik8s>）中運行，徹底消除電阻，讓資料中心只需耗用現今的一小部分電力。或許未來的通用人工智慧（AGI）霸主已經在著手這件事。但可惜的是，渺小的人類等不了那麼久。

現階段，資料中心為機器供電時會產生熱能；這些能量往往以高昂的氣候代價取得，但最終卻會消散。當人們努力提升資料中心的電力使用效率（PUE，見第 2 章）時，他們其實正是在與電阻奮鬥。這同時也是「營運效率」概念背後的驅動力，而這正是本章要探討的主題。



那些超導資料中心可能會設在太空（<https://oreil.ly/k-oG8>），因為那裡可以解決低溫難題；超導體必須非常、非常冷——如太空般的那種冷，不是加件小背心就行的冷。然而，把超導資料中心搬到地球外還得等上一整個世紀，對解決眼前的問題來說緩不濟急。

在過去的三十年間，我們透過 CPU 設計的創新與提升其他硬體效率，來對抗資料中心中的電力浪費。這些進展讓開發人員得以用更少的機器、消耗更少的電力，卻仍能達到相同的功能輸出。然而，靠摩爾定律帶來的升級如今已不足為靠，因為摩爾定律正在放緩（<https://oreil.ly/aiuh7>），甚至有可能停滯（雖說大概不至於真的停止）。

好消息是，硬體效率並不是我們對抗「電阻」的唯一武器。所謂的「營運效率」正是像均維這類 Green DevOps 或 SRE 人員，在其資料中心用來減少電力浪費的法寶。所以，營運效率究竟是什麼呢？

供暖機制的演變

先暫且把超導體放一邊，使用電力時最大的「副產品」便是因電阻產生的熱。只要不是用在暖氣上，這股熱能就是壞事一件，而在現代資料中心裡，你幾乎看不到暖氣設備。

對加熱元件來說，電阻熱並非壞事——它正是電暖器運作的機制，能把電能 100% 轉成熱能。放在從前，100% 效率已經夠嚇人；如今，我們要的是每一瓦電都物超所值。熱泵便靠著冷媒的巧妙循環，效率竟可突破 400%（！），直接把門檻拉到新高度。可惜，熱泵的設計、製造與維運，比簡單的電暖器複雜得多，隱含碳排也高上好幾階。

隨著熱泵的普及，我們勢必會越來越擅長生產並操作它們，但這不會一夕完成，更不會毫無陣痛。事實上，熱泵清楚示範了任何轉型到更高效率技術時，前期投入與取捨絕非小事。

我們已砸下巨額資金在熱泵上，未來仍需加碼。這正說明：面對氣候變遷，真正的解方幾乎無法「唾手可得」，絕大多數都得花許多力氣變成可大量複製、可商業化的產品。然而解方確實存在，我們需要加速這個「商品化」過程——換句話說，要把「研發—安裝—營運」整條鏈的效率提升百倍、甚至千倍。而那些無法商品化的技術，恐怕撐不到能源轉型完成的那一天。

電力的碳強度：清潔與骯髒的區別

平常要為任何設備供電，往往只是把插頭往牆上一插，故事就結束了。之所以會「理所當然」地覺得所有電力都是乾淨、低碳的，大概是因為插座不會冒煙、也聞不到味道。（或許學校應該安排學生去燃煤電廠實地參訪？）

身為最終用電者，我們要怎麼判斷手上的電「乾淨」還是「髒」呢？在揭曉方法之前，先複習一下兩者的差異，以及區分方法。

不只是軟體產業，幾乎所有行業現在都會用到一個指標：碳強度（carbon intensity），它用來衡量每消耗 1 度電（kWh）會排放多少二氧化碳當量（CO₂e）。

還記得第 2 章提到的嗎？「碳」是對所有溫室氣體（GHG）的簡稱；而碳強度的標準單位是「每度電排放多少克 CO₂e」（gCO₂e/kWh）。有了這個統一量尺，才能比較不同電力來源的環境衝擊。

舉例來說，假設你能把筆電直接插在風力發電機上，理論上碳強度可視為 0——先暫不計入風機材料的生命週期排放。不過現實世界裡，大家不可能真的直接用到「單一」風電。

對大多數人來說，從電網拿到的是一鍋大雜燴：當下可用的所有電源全混在一起。（可能是「燃煤主餐配風電花生米」之類的組合。）所以日常用電的碳強度，其實是該國 / 區域電網中各種高、低碳電源的平均值。

談到「電網的運作方式」、「各國制度差異對區域與全球電力市場的影響」時，故事就變得複雜又精采。若想深究，推薦參考國際能源署（IEA）的公開資源（<https://oreil.ly/j7UK3>）。

不過對普羅大眾而言，掌握「碳強度」這個數字就夠判別電力是否乾淨。以 Google Cloud 為例，只要某區域電網的碳強度低於 200 gCO₂e/kWh，就會標成「低碳」（https://oreil.ly/qb6w_）區域。

理解碳強度，不只是為了寫出更有效率的程式；更重要的是建立一個新的思維模型，驅使我們去追求更聰明、有效地利用「低碳電力」！

碳強度的變異性

想必你已經熟悉，電力的「碳強度」會隨時間與地理位置而變化。

如圖 5-1 所示，像台灣或英國這些地區，因為全年風勢強勁，在地理條件上特別適合發展風力發電。事實上，有報導指出（<https://oreil.ly/uOqdX>），在 2023 年第一季，英國風電發電量首次超越燃氣發電廠。

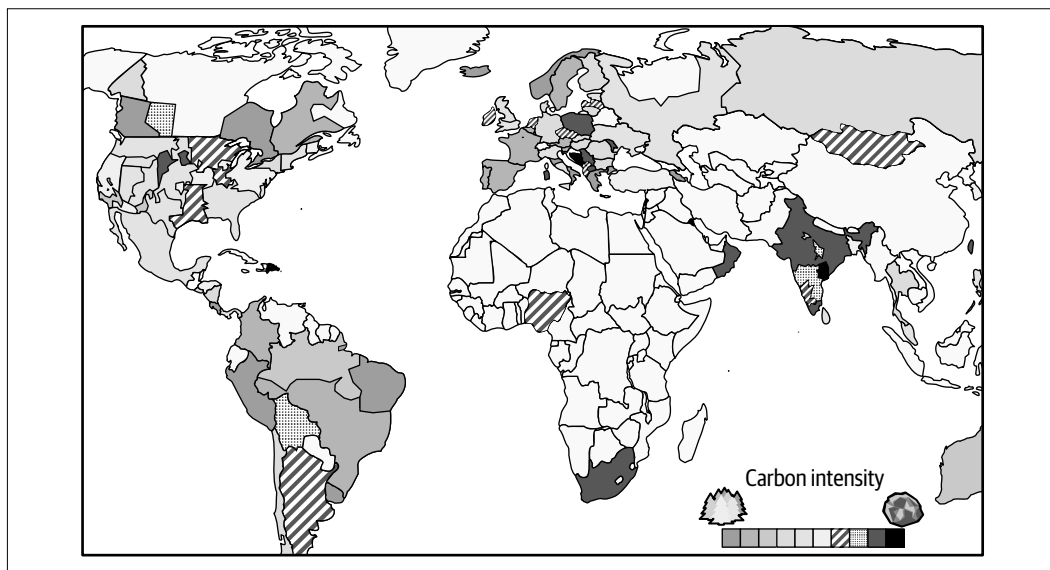


圖 5-1 顯示電力碳強度地理差異的地圖，樹木表示較乾淨的能源，煤炭表示較骯髒的能源

反觀多數北歐國家，因擁有終年豐沛，主要來自河川的水資源，水力發電便成了它們的主要再生能源。

然而，即使你身處擁有最新再生能源技術的地區，屋頂上閃閃發亮的太陽能板鋪滿整面屋頂，一旦太陽決定長時間隱身、不再灑下陽光，或是風停了，你仍然需要尋找替代的電力來源。

如圖 5-2 所示，由於各地天氣狀況變化無常，電力的碳強度會隨時間波動。



24/7 碳中性能源（CFE）代表能夠不斷地將每一度電的消耗與碳中性發電對齊，全天候、每天、在每個地點都能如此。根據聯合國的說法，達成這項目標是「完全脫碳電力系統」的終極願景（<https://oreil.ly/7cw53>）。

電力地圖提供「全球超過 50 個國家暨 160 多個區域的每小時碳電力消耗和生產資料。」這些資料可分為三種不同的時間範疇：歷史資料、即時資料以及未來 24 小時的預測資料。

請善用 SDK！

由於我們這些軟體開發者經常被指責懶惰，因此有人已經做了繁重的工作，基於另一個碳強度 API（WattTime，<https://www.watttime.org>）創建了一個開源軟體開發工具包（SDK），以促進碳意識技術的快速上市。

由綠色軟體基金會（Green Software Foundation，GSF）推出的碳意識 SDK（Carbon Aware SDK，<https://oreil.ly/QsPmh>），在去年的首屆 GSF 黑客松中成為焦點，參與者能夠使用該工具包作為網路應用程式介面（WebApi）和命令列介面（CLI），將碳強度資訊整合進他們的解決方案中。

參賽者共完成超過 50 個黑客松專案，從 Kubernetes 擴展到 UI 設計調整。請前往後續的「實際案例」部分，看看哪些專案引起了作者的關注。

碳強度是計算一個軟體或系統總碳足跡的關鍵因素之一。第 9 章會更詳細分析相關工具，並考慮其優勢與缺點。

需求轉移

既然已經了解並認識到並非所有電力的生產方式皆相同，接下來就要探討百萬美元的問題：該如何應對這些新獲得的資訊（也就是那個指標）？

再次強調，碳意識計算的核心在於對碳強度波動的應對。其基本策略是根據關鍵指標的變化，將應用程式重新配置到不同的時間或地點。

負載平衡（Load Balancing）和內容傳遞網路（Content Delivery Networks，CDN）是科技產業常用的技術，能夠根據當前情況將流量引導至最具處理能力的區域，以滿足需求。

如果你的應用情境允許在運行時間和地點上有一定的彈性，你可以考慮相應地調整工作負載。

例如，如果你的任務對時間的要求並不迫切，即不需要 99.99% 的時間保持運作，而是每月只有一次或兩次，比如訓練機器學習模型，就可以考慮在電力最清潔的時候訓練大型語言模型（LLM）。

如果當地法規允許，也可以考慮在碳強度更低的地點訓練模型。不過第 8 章會更深入探討這一主題，並思考調整時間和地點，以應用於綠色 AI 實踐中的方法，同時檢視機器學習生命週期每個階段需要考慮的事項。

需求轉移可以分為時間和空間兩個面向，一般認為前者比後者更易於執行。這不僅是因為實施和部署的修改可能是必須的，還因為涉及資料的當地法律可能成為一個難以克服的障礙（稍後會詳細說明這些問題）。

時間轉移

碳強度在一天的時間中會有所變化。如圖 5-3 所示，它在午夜至早上 8 點之間呈現出一個倒鐘型的曲線，因此，如果能將示範工作負載延遲一小時，可能就能節省接近 100 克的二氧化碳當量（CO₂e/kWh）。

這種時間調整的技術稱為時間轉移（*time shifting*），我們認為，這是減少軟體碳足跡的低垂果實之一，每個從事這項業務的人都應該牢牢把握住。

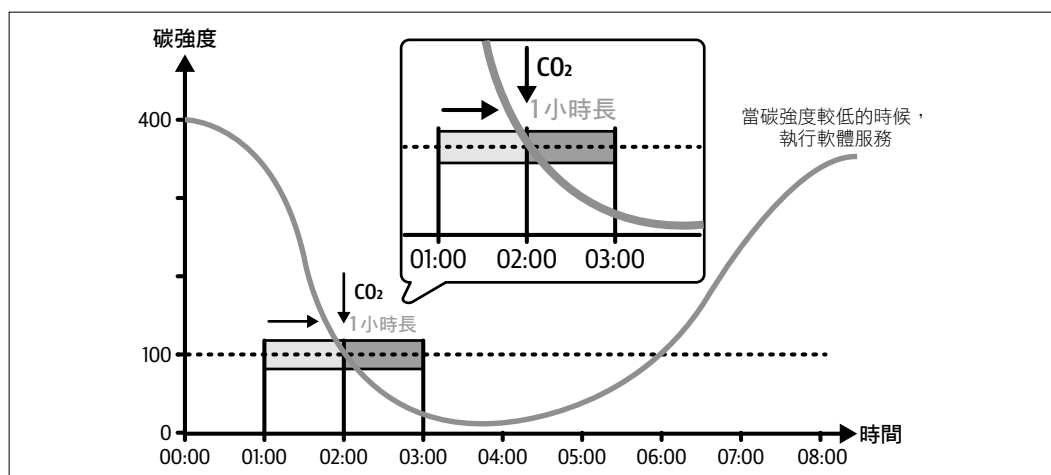


圖 5-3 倒鐘形曲線描述出碳強度的變化，同時展示其可能帶來的低碳時間轉移機會

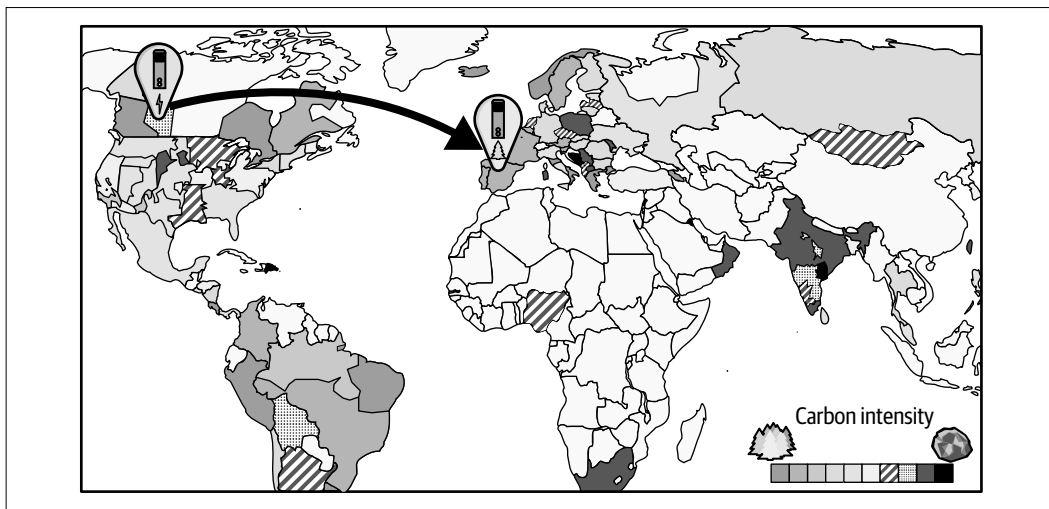


圖 5-4 顯示地區碳強度的全球地圖，以及將應用從高碳強度地區轉移到低碳強度地區的建議

需求塑造

前面提到透過調整應用程式的時間或位置來進行需求移轉的策略中，仍假設電力供應穩定。然而，也正如之前的探討，可再生能源供應並非總是穩定或無限的。因此，以下要很自豪地介紹另一個新策略來應對這一情況：需求塑造（*demand shaping*）。

需求塑造在資訊領域並不陌生，是已多次實施的概念，以滿足各種需求。例如，視訊通話應用程式如 Zoom 和 Google Meet，就使用類似概念來應對不穩定和波動的網路頻寬。

在通話過程中，如果你的網路頻寬受限，應用程式會自動降低影音品質。這樣做的目的是為了優先保護聲音的傳輸與品質，讓你仍然能夠對話，這是視訊會議的主要目標。

另一個值得提到的例子，是漸進式網頁應用程式（PWA）透過各種方法以有效解決網路頻寬問題。例如，漸進式加載技術，也稱為惰性載入（*lazy loading*），允許應用程式逐步分階段載入，這種方法有效地提升應用程式在網路較慢時的效能，避免同時載入所有資源。



第 7 章會詳細介紹需求塑照的歷史。

更環保的機器學習、 人工智慧和 LLMs

如果電腦能夠欺騙人類，讓人類相信它是人類，
就有資格稱它為智慧電腦。

— 艾倫·圖靈（Alan Turing）

2023 年 3 月底，Bill Gates 在其部落格「GatesNotes」中宣告：「人工智慧時代已經來臨。」他認為，人工智慧將如同行動電話或網際網路一般，成為生活中不可或缺的一部分。對於年紀不足以經歷沒有手機和網路時代的兩位作者而言，這番話格外具有分量。無庸置疑，我們正身處人工智慧（AI）和機器學習（ML）的時代，AI 對我們生活的影響之深，前所未見，從藝術創作、醫療診斷、軍事應用、教育評量到氣候議題，AI 在各個領域都成為焦點。

AI 並非新興科技。1950 年，Alan Turing 在其論文「Computing Machinery and Intelligence」中首次提出「思考機器」的概念¹。在這篇開創性文章中，他提出現今廣為人知的「模仿遊戲」理念，也就是我們現在所稱的圖靈（Turing）測試。這個測試旨在評估電腦是否具備人類或類人的認知能力。七十多年後的今天，這個概念依然引人入勝且發人深省，不只對資訊產業的從業人員如此，對好萊塢也是一樣，《西方極樂園》（*Westworld*）、《機械姬》（*Ex Machina*）和《模仿遊戲》（*The Imitation Game*）等多部改編作品就是明證。

1 A. M. Turing, "Computing Machinery and Intelligence," *Mind* 49 (1950): 433–60, <https://oreil.ly/kcDuI>.

自 1950-60 年代第一個 AI 模型問世以來，AI 的發展一直遵循摩爾定律的速度，直到 2012 年出現轉折²。此後，模型規模呈現爆炸性成長。2023 年初，隨著大型語言模型（LLM）透過 ChatGPT、Microsoft365 Copilot 和 Google 的 PaLM 2 等模型與服務向大眾開放，世界經歷新的典範轉移。LLM 是一種運用深度學習技術和海量資料集來理解、摘要、生成和預測新內容的 AI 演算法，這些模型在訓練和推理時所使用的資料量，遠超過先前的語言模型。

本章將聚焦於建構更環保的 ML 和 AI 系統辦法，而非討論 AI 是否會改變我們的生活或主宰世界。不過我們認為有必要在此指出，氣候變遷對人類的威脅程度仍高於 AI 的崛起。

為了解釋 AI 值得獨立成章的原因，我們將帶你了解 AI 應用與模型的快速發展；先劇透一下：其速度遠遠超越摩爾定律³。本章主要介紹 ML 生命週期中的各種減碳方法。雖然生命週期可以有多種定義方式，但本章將採用簡化模型（見圖 8-1）。綠色 AI 本身足以成書，而本章更像是一道開胃菜。

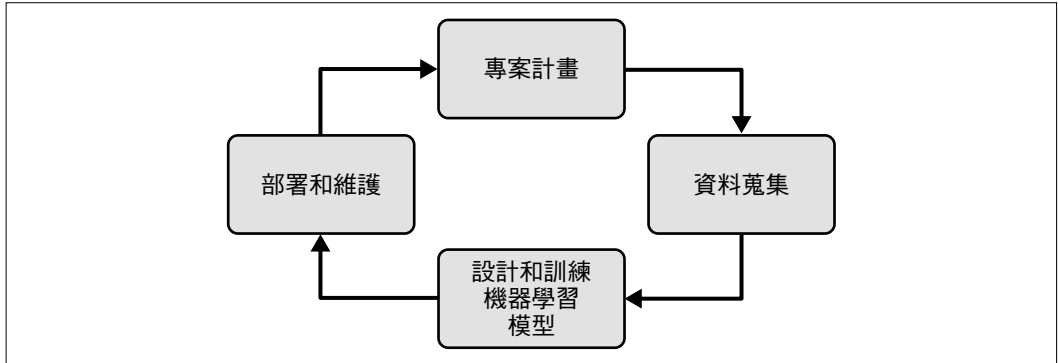


圖 8-1 描述簡化版 ML 生命週期的流程圖

規模與應用的爆發性成長

如前所述，AI 模型最初依循摩爾定律的速度發展，硬體一直是 AI 發展的推進器。早期摩爾定律的進展速度已足夠支撐發展，而後來幾年，平行運算、GPU 和其他專用硬體如 AI 加速器的廣泛應用，進一步加速這個領域的發展。

2 “AI and Compute,” OpenAI, accessed May 16, 2018, <https://oreil.ly/B063Q>

3 Gordon E. Moore, “Cramming More Components onto Integrated Circuits,” Electronics 38, no. 8 (April 19, 1965), <https://oreil.ly/r73IX>.

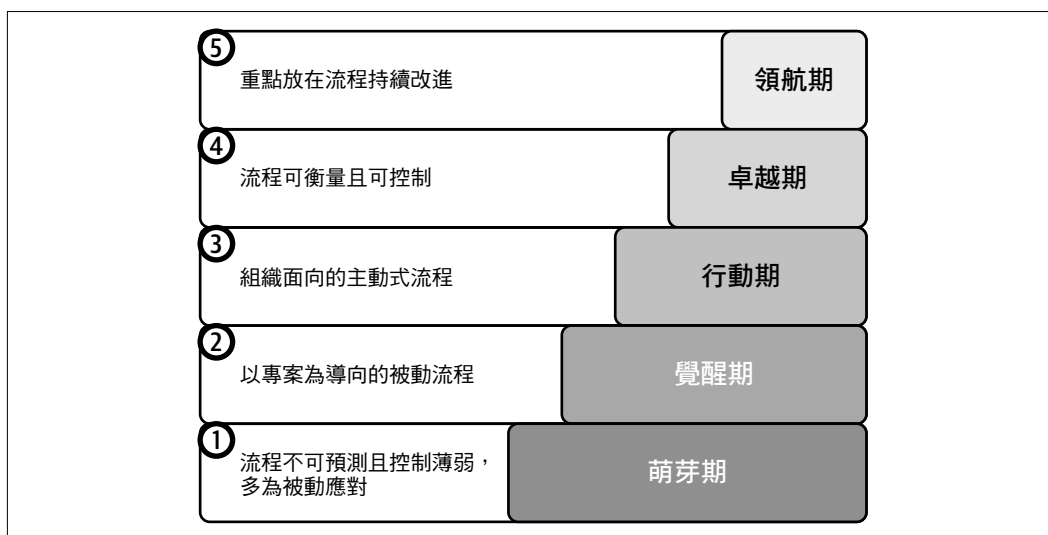


圖 12-2 成熟度矩陣的階梯遞增

基礎層級：

- 第一級（萌芽期）：流程不可預測且控制薄弱，多為被動應對。
- 第二級（覺醒期）：各專案獨立運作，仍以被動反應為主。
- 第三級（行動期）：流程標準化、文件完備，易於理解和稽核。
- 第四級（卓越期）：流程可量測且可控制。
- 第五級（領航期）：根據量化回饋持續改進。

就綠色軟體而言，這些層級可理解為：

- 第 1 級萌芽期：缺乏組織策略，停滯不前。僅有個人層面的環保意識。
- 第 2 級覺醒期：初步掌握要領但欠缺一致性，各團隊程度不一，具備基本排放資料。
- 第 3 級行動期：組織具備全面認知和完整歷史資料，流程標準化並全面實施，已完成無需大量資料支持的綠色改善。
- 第 4 級卓越期：可動態衡量進展，掌握即時資料。
- 第 5 級領航期：依據即時資料監控並改進排放效能，在業界扮演標竿角色，技術卓越傲視群倫。



此成熟度矩陣主要支援兩類核心使用者：決策層與內部永續推動者。它不僅為高層主管提供清晰的發展藍圖，更為內部倡議者提供具體依據，讓永續轉型更有說服力。

對顧問團隊而言，成熟度矩陣是理想的溝通工具，能有效傳達服務內容與品質標準，協助客戶掌握專案進展，並以客觀指標展現專案的實質價值遠超過表面的環保行動。這套客觀的評估框架不僅能提升客戶投入資源的意願，更有助於擴大綠色軟體顧問服務的市場規模。期待看見更多專業人士投身綠色軟體顧問領域，為產業永續發展注入新動能。

綠色軟體成熟度矩陣的實踐之道

值得慶幸的是，組織無需等待完整的資料就能啟動永續轉型，從目前大多數組織所處的第一級「萌芽期」，邁向第二級「覺醒期」。這特別重要，因為如第 9 章所述，溫室氣體排放的衡量仍具挑戰性，相關工具正持續演進中。

在成熟度矩陣的前兩個層級，組織可以立即採取有效行動。雖然管理大師彼得·杜拉克常說「無法衡量就無法改進」，但在永續領域，這句名言未必適用。

有許多公認能確保減碳效益的行動可以立即展開，例如：

- 關閉未充分利用的「殭屍」機器，進行資源最佳化
- 選擇更環保的託管服務
- 轉移工作負載至雲端服務，或善用多租戶功能
- 提升營運能力，如導入 GitOps 或 SRE 原則

一個簡單的起點是：只要能降低託管成本，就能減少碳排放。這可以作為初期的參考指標，讓你立即展開行動。

級別

接著來更詳細地了解一下綠色軟體成熟度矩陣 v1.0 的各個級別。



需要複習其中一些測量術語的話，請快速瀏覽第 9 章。

第一級「萌芽期」

- 組織內有些人關心科技的綠色目標，但整個組織還沒有統一的興趣、流程或承諾。
- 某些開發團隊可能會測量他們的碳排放，但無法得知公司整體的科技碳足跡是多少。
- 還沒有零碳目標或綠色軟體策略，或者只有一些空泛的零碳目標，比如 2050 年達到淨零排放（*Net Zero by 2050*，https://oreil.ly/TEro_），但尚未有具體計畫。

說實在的，大多數人現在都處在這個階段，別難過，這表示你用比較輕鬆的方式就做出很多改進。

第二級「覺醒期」

- 知道你的科技組織的範疇 1 和範疇 2 排放量，可能是每週、每月或每年統計一次。
- 科技組織內有幾個推動減碳的專案，可能會用到你們所蒐集的資料，但也不一定。
- 你的工程師接受過一些綠色系統和營運的培訓。
- 正在處理第 4 章提到的那些不用動腦就知道該做的事，比如手動關閉已經沒在用的殭屍服務、調整適當規模，或是把可以轉換的功能改用單次運算實例。

五年前，你們大概只會抵銷範疇 1 和範疇 2 的排放，就宣稱資訊系統達到碳中和，然後發個自吹自擂的新聞稿就完事了。但現在，因為大家的期待提高了，這樣只能算是五個階段中的第二階段：碳抵銷的價值有限，而碳中和雖然不錯，但還遠遠不夠。

不過呢，如果你們已經做到這個程度，其實表現還不錯啦！畢竟大多數人連這樣都還沒做到。



「碳中和」曾經是一個令人嚮往的目標，只要透過購買碳權抵銷自身排放即可達成。這些抵銷方案包括植樹造林或支持新建再生能源電廠。然而，隨著認知的提升，這類抵銷的效益受到質疑：再生能源產業已蓬勃發展，不再仰賴此類支持；而新栽種的樹木未必能長期固碳，實質效益有限。不過，這種做法仍有其歷史意義：它展現了組織衡量直接排放的決心，且在再生能源發展初期，這些早期企業用戶確實發揮了重要推動作用。