

OECD啟動全球首創的《開發先進人工智慧系統組織的報告框架》



2025年2月7日，經濟合作暨發展組織（Organization for Economic Cooperation and Development，OECD）正式啟動《開發先進人工智慧系統組織的報告框架》（Reporting Framework for the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems，簡稱G7AI風險報告框架）。

該框架之目的是具體落實《廣島進程國際行為準則》（Hiroshima Process International Code of Conduct）的11項行動，促進開發先進人工智慧系統（Advanced AI Systems）的組織建立透明度和問責制。該框架為組織提供標準化方法，使其能夠證明自身符合《廣島進程國際行為準則》的行動，並首次讓組織可以提供有關其人工智慧風險管理實踐、風險評估、事件報告等資訊。對於從事先進人工智慧開發的企業與組織而言，該框架將成為未來風險管理、透明度揭露與國際合規的重要依據。

G7 AI風險報告框架設計，對應《廣島進程國際行為準則》的11項行動，提出七個核心關注面向，具體說明組織於AI系統開發、部署與治理過程中應採取之措施：

1. 組織如何進行AI風險識別與評估；
2. 組織如何進行AI風險管理與資訊安全；
3. 組織如何進行先進AI系統的透明度報告；
4. 組織如何將AI風險管理納入治理框架；
5. 組織如何進行內容驗證與來源追溯機制；
6. 組織如何投資、研究AI安全與如何降低AI社會風險；
7. 組織如何促進AI對人類與全球的利益。

為協助G7推動《廣島進程國際行為準則》，OECD建構G7「AI風險報告框架」網路平台，鼓勵開發先進人工智慧的組織與企業於2025年4月15日前提交首份人工智慧風險報告至該平台（<https://transparency.oecd.ai/>），目前已有包含OpenAI等超過15家國際企業提交報告。OECD亦呼籲企業與組織每年定期更新報告，以提升全球利益相關者之間的透明度與合作。

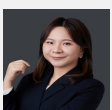
目前雖屬自願性報告，然考量到國際監管機關對生成式AI及高風險AI系統透明度、可問責性（Accountability）的日益關注，G7 AI風險報告框架內容可能成為未來立法與監管的參考作法之一。建議企業組織持續觀測國際AI治理政策變化，預做合規準備。

相關連結

[OECD AI Transparency Platform, OECD](#)

[OECD Launches Global Framework to Monitor Application of G7 Hiroshima AI Code of Conduct, OECD](#)

[Italian Ministry of Innovation, G7 Digital Tech Working Group Finalizes Reporting Framework for Advanced AI Systems, Innovazione.gov.it](#)



陳郁潔

副法律研究員 編譯整理

上稿時間：2025年05月

資料來源：

OECD AI Transparency Platform, OECD, <https://transparency.oecd.ai/> (last visited Feb. 21, 2025).
OECD Launches Global Framework to Monitor Application of G7 Hiroshima AI Code of Conduct, OECD, <https://www.oecd.org/en/about/news/press-releases/2025/02/oecd-launches-global-framework-to-monitor-application-of-g7-hiroshima-ai-code-of-conduct.html> (last visited Feb. 21, 2025).
Italian Ministry of Innovation, *G7 Digital Tech Working Group Finalizes Reporting Framework for Advanced AI Systems*, Innovazione.gov.it, <https://innovazione.gov.it/notizie/articoli/en/g7-digital-tech-working-group-finalizes-reporting-framework-for-advanced-ai-system/> (last visited Feb. 26, 2025).

文章標籤

人工智慧

人工智慧監管

人工智慧風險管理

推薦文章

