IBM提出「人工智慧日常倫理」手冊作為研發人員指引

隨著人工智慧快速發，各界開始意識到人工智慧系統應用、發展過程所涉及的倫理議題，應該建構出相應的規範。IBM於2018年9月02日提出了「人工智慧日常倫理」（Everyday Ethics for Artificial Intelligence）手冊，其以明確、具體的指引做為系統設計師以及開發人員間之共同範本。作為可明確操作的規範，該手冊提供了問責制度、價值協同、可理解性等關注點，以促進社會對人工智慧的信任。

一、問責制度（Accountability）

由於人工智慧的決策將作為人們判斷的重要依據，在看似客觀的演算系統中，編寫演算法、定義失敗或成功的程式設計人員，將影響到人工智慧的演算結果。因此，系統的設計和開發團隊，應詳細記錄系統之設計與決策流程，確保設計、開發階段的責任歸屬，以及程序的可檢驗性。

二、價值協同（Value Alignment）

人工智慧在協助人們做出判斷時，應充分考量到事件的背景因素，其中包括經驗、記憶、文化規範等廣泛知識的借鑑。因此系統設計和開發人員，應協同應用領域之價值體系與經驗，並確保演算時對於跨領域的文化規範與價值觀之敏感性。同時，設計師和開發人員應使人工智慧系統得以「了解並認知」用戶的價值觀，使演算系統與使用者之行為準則相符。

三、可理解性（Explainability）

人工智慧系統的設計，應盡可能地讓人們理解，甚至檢測、審視它決策的過程。隨著人工智慧應用範圍的擴大，其演算決策的過程必須以人們得以理解的方式解釋。此係讓用戶與人工智慧系統交互了解，並針對人工智慧結論或建議，進而有所反饋的重要關鍵；並使用戶面對高度敏感決策時，得以據之檢視系統之背景數據、演算邏輯、推理及建議等。

該手冊提醒，倫理考量應在人工智慧設計之初嵌入，以最小化演算的歧視，並使決策過程透明，使用戶始終能意識到他們正在與人工智慧進行互動。而作為人工智慧系統設計人員和開發團隊，應視為影響數百萬人甚至社會生態的核心角色，應負有義務設計以人為本，並與社會價值觀和道德觀一致的智慧系統。

相關連結

- Bias in AI: How we Build Fair AI Systems and Less-Biased Humans
- Requirements for Moral Judgments

相關附件

- Everyday Ethics for Artificial Intelligence [pdf]
- The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems [pdf]

你可能會想參加

→ 【2023科技法制變革論壇】AI生成時代所帶動的ChatGPT法制與產業新趨勢
→ 「跨域數位協作與管理」講座活動
→ 新創採購-政府新創應用分享會
→ 【線上場】113年「新創採購機制及鼓勵照護機構參與推動」說明會
→ 【北部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【南部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 113年新創採購-照護機構獎勵說明會
→ 【南部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【北部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會

→ 【中部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【臺北場】113年度新創採購-招標作業廠商說明會
→ 【臺中場】113年度新創採購-招標作業廠商說明會
→ 【高雄場】113年度新創採購-招標作業廠商說明會

陳明
法律研究員 編譯整理

上稿時間：2018年11月

資料來源：

1. Every day Ethics for Artificial Intelligence, https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf. (last visited Oct. 10, 2018).
2. Institute of Electrical and Electronics Engineers, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_general_principles.pdf. (Dec. 13, 2017).

延伸閱讀：

1. 陳譽文，人工智慧規範性議題綜觀，科技法律透析，第29卷第4期，頁48（2017）。
2. Bias in AI: How we Build Fair AI Systems and Less-Biased Humans, https://www.ibm.com/blogs/policy/bias-in-ai/ (last visited Oct. 20, 2018).
3. The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_general_principles.pdf. (last visited Oct. 03, 2018). 4. Requirements for Moral Judgments, Mt. San Antonio College, https://faculty.mtsac.edu/cmcgruder/moraljudgements.html (last visited Oct. 03, 2018).

文章標籤

人工智慧

推薦文章