經濟部

DOIT 經濟部產業技術司

「ISO/PAS 8800 道路車輛 - 安全與人工智慧」簡介

陳嘉珮 副法律研究員
報告單位：資策會科法所
報告日期：2025年7月31日

1.1 概述 - 適用範圍

ISO 26262

聚焦於軟硬體之功能安全，規範汽車功能安全開發流程，以確保汽車零組件準確且及時發揮功能

難以應對AI系統之不確定性與難以解釋性

ISO 21448

預期功能安全(Safety Of The Intended Functionality, SOTIF)，針對自駕車之感知、決策與控制系統，進行功能分析及補強

未完善規範AI系統故障之措施

ISO 8800

2024年12月國際標準組織（International Organization for Standardization, ISO）發布針對車用AI制定之安全驗證框架

適用系統：將AI應用於安全功能之道路行駛量產車輛電子電氣系統

適用階段：AI系統設計、模型訓練、模型驗證、布建與運行過程

適用對象：車廠、AI系統開發商、整車設計單位

1.2 概述 - 風險與挑戰

AI系統具備學習能力與自我調整特性，使其難以進行安全性分析與驗證

功能性不足

因輸入條件、資料特徵與模型特性，導致特定條件下預測失常

不可預測性

開發階段無法完全預測，於罕見情況發生時AI系統可能失效

不可解釋性

AI模型結構高度複雜且缺乏可解釋性，使開發者難以掌握模型內部決策邏輯

高度依賴資料

AI系統之學習行為與表現取決於資料之代表性、完整性與品質

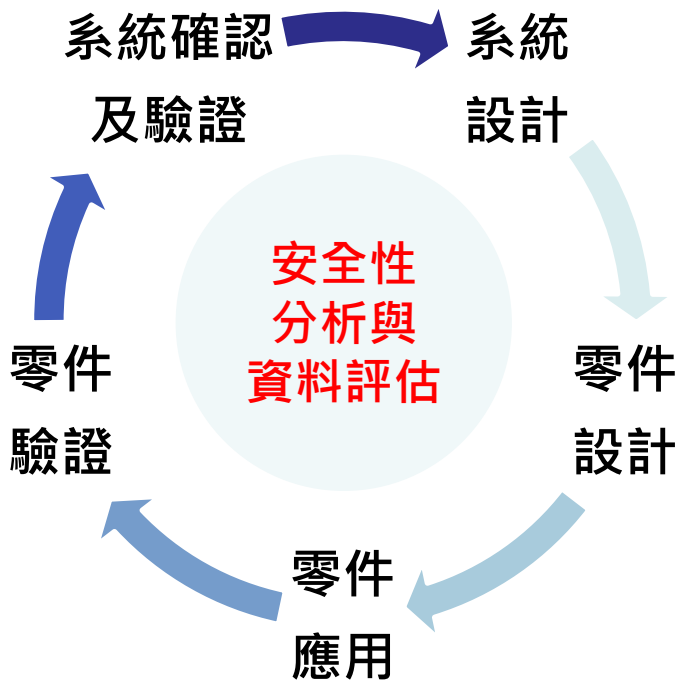
環境變異性

運行環境充滿動態變數，可能改變輸入資料之特性，使原有保證論證失效

2.1 驗證原則與方法 - AI安全生命週期

標準建議依照個案制定生命週期，並說明重點關注事項

AI（設計）安全生命週期範例



定義AI系統開發所需過程，確保運作過程之安全性

需符合ISO 26262、IS021448之規範

開發活動將持續更新，直至證明具備足夠安全性

用以識別、分析潛在功能缺陷及其原因，並建立緩解措施

蒐集安全需求所需資料

2.2 驗證原則與方法 - 安全需求

安全需求為ISO 8800之核心，需確保AI系統持續符合之並滾動式更新

目的

確保AI系統安全性且具備完整性與一致性

輸入空間

需定義AI系統之輸入空間，藉由減少操作之不確定性確保設計和運行安全

需求改良

需以開發、驗證和確認過程之資料改善安全需求

運行安全

需定義運行後之安全需求

- 偵測AI系統失效狀態
- 蒐集資料以改良AI系統

偵測與報告

偵測到以下情況需報告：

- AI系統無法符合安全需求
- 僅部分輸入空間符合安全需求

評估指標

設計指標時需提供正當理由，例如過往產品經驗、業界共識、系統分析結果等

2.3 驗證原則與方法 - 保證論證

標準提出證明AI系統滿足安全需求之考量與認證方法，可適用整體生命週期或單一階段

AI系統特有考量

AI系統係以平均績效指標設計，不足以涵蓋罕見但危險之情況

實際運作環境可能產生變化，產生未知風險

AI系統多依賴間接驗證，無法進行直接確認

機器學習之資料選擇與訓練，須確保資料妥適性

論證方法

產品導向

參考已實際使用之AI系統功能與特徵

流程導向

著重開發流程與驗證程序

風險導向

影響AI系統剩餘風險之因素

循環論證

應加入前次安全生命週期之論證成果為基礎

2.3.1 保證論證 - 資料管理

AI系統依賴資料訓練與應用，故須持續確保資料品質以預防或緩解風險

1.安全性檢驗

識別安全相關資料之缺陷，確定其因果後制定預防或緩解措施，並可作為措施之績效指標

2.需求辨識

識別資料缺陷如何影響AI系統之安全性

3.設計

設計資料之演繹和歸納分析方法，並得引入開發階段未使用之風險緩解措施

4.應用

識別資料準備和分類流程、方法或工具之潛在問題，並使之符合安全需求

5.驗證

確保資料之取得與使用符合安全需求

6.維護

因動態變數繁多，需使資料管理方法持續符合安全需求

2.3.2 保證論證 - 系統安全分析

AI系統設計期間

識別可能導致安全性相關故障或錯誤之原因

- 使用適合辨識AI模型安全性錯誤之分析技術
- 識別違反安全需求之錯誤並分析原因（例如潛在性功能不足）

制定預防或緩解措施

- 改善AI系統設計
- 修改安全需求
- 蒐集相關資料並運用

2.3.3 保證論證 - 系統驗證與確認

完成設計後須進行系統之獨立性分析及零件層級檢測

定義安全需求階段

確保安全需求之正確性與完整性

系統開發階段

透過模擬或分析評估成果，確保符合安全要求

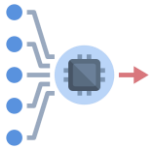
測試階段

評估運行是否符合安全需求

挑戰

目前缺乏廣泛被接受之安全性評估標準

安全需求缺乏精確描述



資料來源繁多，格式存在不一致

訓練樣本不足導致預測結果不一致

不可預測之錯誤限制預測能力

2.3.4 保證論證 - 持續監督

持續檢視與修正AI系統及緩解措施，確保實際運行後之安全性

建立維持運行期間符合安全需求及保證論證之措施及流程

設置配套軟硬體控制機制，維護AI系統運行

評估運行期間之安全性事件，風險超過可接受範圍時須採取緩解措施

檢視已執行之緩解措施，如仍有不可接受之剩餘風險須修改之

定期維護AI系統之安全，因應可能變因