

1、 快刀 AI 文章辨識寫作檢測功能是如何運作的

快刀 AI 文章辨識是一款利用快刀專利核心技術，結合困惑度分析文本中的語言特徵，如句子結構、語法、用詞等，以辨識由 ChatGPT、Gemini、BERT、Jasper、Deepseek、Copilot 等 AI 語言模型所生成的文章。此外，快刀 AI 文章辨識還能識別出文章是否使用翻譯軟體，為使用者提供全方位的文本分析與辨識服務。

2、 快刀 AI 文章辨識可以檢測過去已上傳文章中是否有 AI 寫作嗎？

快刀 AI 文章辨識的操作介面非常簡單易用，無需專業知識，只要將過去已上傳的文章再次複製貼上於系統介面上，再選取立即辨識，就可以檢測文章是否由 AI 生成的。

3、 快刀 AI 文章辨識文字檢測功能支援哪些語言？

快刀 AI 文章辨識文字檢測功能能支援繁中/簡中/英/日/韓/泰/越/西/法/葡/德/義/俄/波/荷/瑞典/羅馬尼亞/捷克/挪威/希臘/阿拉伯/孟加拉/保加利亞/克羅埃西亞/希伯來/印地/匈牙利/塞爾維亞/土耳其/烏克蘭等多國語言。

4、 系統會標示具體哪些段落是 AI 生成的嗎？

系統會以紅底及黃底兩種顏色，分別標示出極大機率為 AI 生成之文字，與有可能為 AI 生成之文字。

5、 如果上傳非英文的文章進行檢測，會發生什麼事？

快刀 AI 文章辨識所使用的自然語言技術與機器學習技術分析文章內的語言特徵，如句子結構、語法、用詞等，可準確判斷 AI 生成的文章；此核心技術不受限於僅使用單一語種，任何語種文字，尤其是中文(無論正繁體或簡體)，快刀 AI 文章辨識都能順暢的辨識與判斷；以下截圖顯示出快刀 AI 文章辨識卓越的中文文章檢測成果：

原創性辨識報告

報告摘要

報告編號: aisensei3534

作者姓名:

原創比例

純正原創

微少原創

作者原創內容 73%

疑似由AI生成內容 0%

高機率AI生成內容 26%

辨識時間: 2025-01-08 13:44

辨識字數: 5923字

原創程度: 73% 中度原創

原創程度說明

原創度	相似程度	人工檢查建議
100%	純正原創 系統偵測文本中沒有任何內容為【疑似 AI 生成內容】或【高機率 AI 生成內容】，意即文本內容皆為作者原創。	暫不需要
76% ~ 99%	高度原創 系統偵測文本中有 1% ~ 24% 為【疑似 AI 生成內容】，意即文本部分內容僅有輕度機率為 AI 生成內容，其餘內容 76% ~ 99% 皆為作者原創。	建議抽查 【疑似 AI 生成內容】 標記處
51% ~ 75%	中度原創 系統偵測文本中有 25% ~ 49% 為【疑似 AI 生成內容】或【高機率 AI 生成內容】，意即文本部分內容有中度機率為 AI 生成內容，僅部分內容 1% ~ 75% 為作者原創。	建議抽查 【疑似 AI 生成內容】 與【高機率 AI 生成內容】 標記處
26% ~ 50%	輕度原創 系統偵測文本中有 50% ~ 74% 為【疑似 AI 生成內容】或【高機率 AI 生成內容】，意即文本部分內容有高度機率為 AI 生成內容，僅少部分內容 26% ~ 50% 為作者原創。	建議逐句檢查 【疑似 AI 生成內容】 與【高機率 AI 生成內容】 標記處
0% ~ 25%	微少原創 系統偵測文本中 75% ~ 100% 為【疑似 AI 生成內容】或【高機率 AI 生成內容】，意即文本部分內容有高度機率為 AI 生成內容，極少部分內容 1% ~ 25% 為作者原創。	極度建議逐句檢查 【疑似 AI 生成內容】 與【高機率 AI 生成內容】 標記處

疑似 AI 生成內容 0%

高機率 AI 生成內容 26%

6、 為什麼將自己寫的文章上傳後，辨識結果卻顯示疑似由 AI 生成內容？

若文章撰寫中，曾有詞句為 AI 協作(如草稿、潤稿、改寫等)，也有極大可能會被系統判定為 AI 生成之文字。

7、快刀 AI 文章辨識優勢比較表

比較項目	快刀 AI 文章辨識	T 牌 AI 寫作檢測工具
可否試用	歡迎所有老師申請試用 (不限定是否是快刀用戶)	非 T 牌用戶，無法提供試用
上傳方式	介面操作非常便捷，直接複製貼上文字/文章，或是上傳 Word/PDF 檔就可以立即進行檢測	操作繁複，要先存成文字檔案，再選取上傳檔案才能進行系統檢測
能檢測的語言	中/英/法/德/義/西/葡/俄/波/荷/瑞典/羅馬尼亞/捷克/挪威等多國語言	僅能檢測英文、西文及日文
可檢測的 AI 寫作模型	能檢測出 GPT5、GPT4、GPT3、GPT2、BERT、Jasper 等多種 AI 語言模型所產生的內容	僅能檢測 GPT-3、GPT-3.5 以及其變種的模型

8、為什麼我明明是上傳 PDF 檔案，還是顯示上傳失敗或是辨識失敗？

因 PDF 檔檔案內建安全性問題，導致有時候上傳到系統上會被擋。遇到這種情況再麻煩改使用 word 或剪貼上傳進行比對

9、快刀 AI 文章辨識在自然語言處理上，使用了「困惑度」計算方法，請問什麼是「困惑度」計算法？

困惑度 (Perplexity) 是一個用來衡量語言模型表現的指標。簡單來說，就是用來評估模型預測下一個詞語的表現，困惑度越低表示模型越準確。

我們可以用一個猜詞語遊戲來說明這個概念；譬如說，有個人要猜下一個詞語是甚麼，他看到的提示是：「今天天氣真好，我們去 _ _ _ _ 」；在這個情境下，他可以猜「公園」、「游泳」、「購物」等詞語，因為這些選擇都很合理，困惑度就比較低。

但如果提示是「我喜歡吃 _ _ _ _ 」，這時有很多可能的選擇，比如「蘋果」、「披薩」、「冰淇淋」等等，困惑度就比較高，因為有很多選擇，他不知道應該猜哪一個詞語。

所以，語言模型的困惑度類似這個例子。當模型預測下一個詞語時能夠縮小可能性範圍，困惑度就會較低，也就表示模型對於預測下一個詞語更有信心。

快刀 AI 文章辨識利用一連串的困惑度方法來評估 ChatGPT 生成的文章，並透過機器學習推導出 AI 寫的詞語，很有效率的幫助我們得到分析結果。

10、快刀 AI 文章辨識在偵測 AI 文章時，會用到哪些分析方法？

如今 AI 技術可以生成相當自然而高水準的文章，因此檢測它們變得具有挑戰性；快刀 AI 文章辨識採用多種方法來檢測 AI 生成的文章，概略說明如下：

(1) 基於特徵的分析

快刀 AI 文章辨識首先通過分析文章中的語言特徵，如單詞使用、句子結構、語法和標點符號使用等，進行基於特徵的分析。通常 AI 生成的文章在這些方面可能存在較為明顯的規律性和模式，這些特徵可以被提取出來並使用機器學習演算法（如支持向量機、隨機森林等）進行分類，從而有效地檢測文章。

(2) 統計語言模型

快刀 AI 文章辨識使用統計語言模型評估文章在語言上的合理性；這些模型會計算文章中各個單詞和短語的出現概率，並與真實文本的概率分佈進行比較；如果文章中的概率分佈與真實文本存在顯著差異，則有可能是 AI 生成的。常用的統計語言模型包括 n-gram 模型、隱馬爾可夫模型等。這些模型可以幫助快刀 AI 文章辨識在文章中發現不自然的概率分布，從而判定文章的來源。

(3) 深度學習方法

快刀 AI 文章辨識在檢測 AI 生成文章方面，有許多不同方法。其中有運用到卷積神經網絡（CNN）和遞歸神經網絡（RNN）兩種的深度學習模型；這些模型通過對文章進行多層次的特徵提取，捕捉到更為微妙的語言模式；通過訓練這些模型以區分人類撰寫的文章和 AI 生成的文章，幫助快刀 AI 文章辨識在檢測過程中達到更高的準確率；此外，快刀 AI 文章辨識還運用遷移學習方式進行預訓練深度學習模型（如 BERT、GPT 等）來檢測 AI 生成的文章，從而提高檢測的準確性。

11、用於檢測文章是否為 AI 生成的模型是如何訓練的？

快刀 AI 文章辨識需設計和訓練相應的模型，這些模型的訓練過程大致可以分為以下幾個步驟：

(1) 數據收集

首先收集大量的文本數據作為訓練和測試樣本。這些數據通常包括人類撰寫的文章

和 AI 生成的文章。人類撰寫的文章可以從各種線上資源（如新聞網站、博客等）中獲得，而 AI 生成的文章則可以通過使用現有的 AI 文本生成器（如 GPT 系列、BERT 等）生成。

(2) 數據預處理

收集到數據後就需進行預處理，以便模型更好地學習其特徵；這包括將文本分割成單詞或詞組、去除停用詞、進行詞幹提取或詞形還原等。此外，還需要為每個樣本分配標籤，既標示該文章是人類撰寫的還是 AI 生成的。

(3) 特徵提取

接下來就要從預處理後的數據中提取相應的特徵；這些特徵可能包括語言特徵（如單詞使用、句子結構等）、統計特徵（如詞頻分布、n-gram 概率等）、深度學習特徵（如卷積神經網絡或遞歸神經網絡提取的特徵等）等；特徵提取的目的是將文本數據轉換為模型可理解的數值形式。

(4) 模型訓練

建構好特徵和標籤之後，就可以開始訓練模型了。根據前面提到的方法，可以選擇基於特徵的分析方法（如支持向量機、隨機森林等）、統計語言模型（如 n-gram 模型、隱馬爾可夫模型等）、深度學習方法（如卷積神經網絡、遞歸神經網絡等）等。在訓練過程中，模型會嘗試學習區分人類撰寫的文章和 AI 生成的文章之間的特徵差異。通常會使用交叉驗證或者將數據集分為訓練集和驗證集的方式來評估模型的性能；在訓練過程中，模型會不斷地調整其參數以最小化預測誤差，從而提高在未知數據上的泛化能力。

(5) 模型評估

模型訓練完成後，就進行使用獨立的測試集對模型的性能進行評估。測試集通常包含人類撰寫的文章和 AI 生成的文章，並且與訓練集沒有重疊。通過計算模型在測試集上的準確率、精確率、召回率等指標，快刀 AI 文章辨識可以衡量模型在檢測 AI 生成文章方面的效果。

(6) 模型優化與調整

接下來就要根據模型在測試集上的性能，進行優化與調整；這可能包括調整模型參數、使用更多或更少的特徵、尋找更好的訓練數據等。此外，還可以嘗試將不同的模型進行組合，使用集成學習等方法來提高檢測性能。

簡單來說，訓練 AI 生成文章的檢測模型需要進行數據收集、預處理、特徵提取、模型訓練、評估和優化等多個步驟。在這個過程中，需要持續探索和選擇最適合的特徵和模型，以提高檢測準確性和可靠性。此外，由於 AI 生成技術的迅速發展，還需要不斷關注新的檢測方法和技術，以應對可能出現的新型 AI 生成文章。

12、AI 寫作的準確度如何？

AI 生成文章檢測模型的準確度取決於多個因素，包括訓練數據的質量、特徵選擇、模型類型以及模型的訓練與優化等。在理想情況下，這些模型可以在一定程度上區分人類撰寫的文章和 AI 生成的文章。然而，隨著 AI 生成技術的進步，檢測準確度可能會受到影響。

(一) 訓練數據質量

檢測模型的準確度與訓練數據的質量密切相關。如果訓練數據包含了豐富的人類撰寫文章和 AI 生成文章樣本，並且這些樣本能夠代表真實世界中的各種情況，則模型的檢測準確度可能會較高。

(二) 特徵選擇

選擇合適的特徵對於提高檢測準確度至關重要。如果選擇的特徵能夠有效地區分人類撰寫的文章和 AI 生成的文章，則模型的檢測準確度可能會較高。然而，隨著 AI 生成技術的發展，這些特徵可能需要不斷地更新和調整。

(三) 模型類型

不同類型的模型在檢測 AI 生成文章方面可能具有不同的性能。例如，深度學習方法（如卷積神經網絡、遞歸神經網絡等）在某些情況下可能具有較高的檢測準確度。然而，這些模型通常需要大量的訓練數據和計算資源。

(四) 模型訓練與優化

模型的檢測準確度還取決於其訓練和優化過程；通過適當的參數調整、特徵選擇以及集成學習等方法，從而提高模型的檢測準確度。