

Neeko：為青少年建立更安全的網路環境以擊潰網路誘拐



組員 | 王瑋華、施巧柔 指導教授 | 程芙茵

摘要與研究問題

濫用網路進行色情和誘拐已成為國際關注的主要問題，目前已有許多學者投入網路誘拐的研究，在過去的文獻中我們觀察到，現今針對網路誘拐檢測的相關研究主要專注於其在司法調查上的應用，然而提前預防兒少受害也是我們需要關注的議題，並且目前關於網路誘拐檢測系統的研究多以英文為主要語言，因此本研究嘗試將中文應用於現有檢測系統的方法。除此之外，我們結合檢測模型，實作一款聊天機器人Neeko，以反思性問題作為誘發兒少思考的契機，在兒少與誘拐者建立穩定關係之前，幫助其意識到網路交友存在的誘拐風險，作為檢測及進一步干預網路誘拐的媒介。我們向6~17歲的兒少及其家長發放問卷，調查使用對象對於聊天機器人設計的看法，並依此總結未來改善的方向。

本研究主要分為以下三部分：

1. 將中文聊天內容經過翻譯及處理後，套用在現有針對網路誘拐的英語資料庫所建立之機器學習模型上，探討此方法對於辨識網路誘拐的成效。
2. 探討聊天機器人做為預防網路誘拐的方式在兒少與家長間的接受度。
3. 結合過去文獻與問卷調查結果設計實作一結合網路誘拐檢測模型的聊天機器人原型。

研究成果

檢測模型效果評估

本研究使用了 Vogt, Leser, 與 Akbik (2021) 提出的早期性誘拐者檢測模型。測試資料的方面，考量到聊天紀錄的隱私問題，蒐集網路誘拐中文資料在執行上十分困難，因此，本研究利用 Vogt 等 (2021) 所提出之模型的測試資料經過翻譯，作為中文測試資料。結果顯示 accuracy 為 95%，precision 為 99%，recall 為 67%，F1-score 為 72%。

問卷調查與結果

採調查研究法，分別設計了兒少和家長（主要照顧者）調查問卷，招募6至17歲的兒少及家中有兒少之家長填答。最終兒少調查共計回收23份有效問卷，年齡數落在15~17歲。家長調查共計回收31份有效問卷，年齡數落在40~49歲。

問卷第一部分以李克特五點量表的形式分為非常不同意（1）、不同意（2）、普通（3）、同意（4）以及非常同意（5）五個選項，此部分針對聊天機器人之主要功能設計進行調查；第二部分為複選題及開放式問題，針對聊天機器人的互動方式和建議進行開放意見調查，兒少問卷中對Neeko的形象偏好設計了複選題，提出了相關選項如性別、是否為人類、形象風格及「其他」選項。

第一部分調查結果中，對於「向緊急聯絡人傳送通知」的功能，兒少及家長對於此功能皆具正面的看法；針對「聊天機器人對隱私（聊天內容）的侵犯」，兒少存在對於自己和其他人對話隱私可能被侵犯的顧慮；針對「與聊天機器人互動分享的意願」，結果顯示相對於與家長分享交友狀況，兒少對於與聊天機器人互動有較高的意願。

第二部分的開放意見調查中，兒少問卷對於「偏好的聊天機器人形象」的複選題，其中有約70%的受試者選擇「溫和親切的」，且相較於「人類」的形象，有較多受試者偏好聊天機器人以「非人類」的形象。

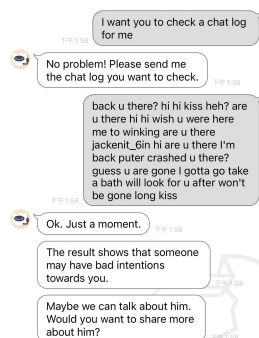
第二部分開放性問答，以下列舉數例問卷填答的結果，兒少問卷對於希望能夠新增的功能建議回饋中，與陪伴相關的回覆較為常見，如「談心」，對於希望的互動方式建議如「口氣很好的，很溫柔的方式」、「稍微的提醒，但不會過的去影響自己的想法」等。家長問卷對於聊天機器人之功能建議，部分受試者表示希望過程中能夠有真人介入，如「可以加入真人服務，有危險立即與真人連線」等。

聊天機器人設計

聊天機器人主要的目的為促進兒少對於網路交友的思考，透過適時的提醒幫助兒少建立網路交友的觀念、培養獨立思考的能力，以遏止他人對兒少進行網路誘拐。聊天機器人將常駐於使用者的聊天室，初次使用時會顯示歡迎訊息並介紹機器人的用途，開啟後使用者可以輸入任何句子與之交互。

本研究使用 Google Dialogflow cx 建構聊天機器人，對話流程分為兩部分，第一部分為以協助兒少思考為目標而設計的反思性問題，第二部分修改了 Dialogflow cx 的 prebuilt agent 以因應日常對話，達到兒少於問卷中希望擁有的陪伴功能。

第一部分中，使用者可以向聊天機器人提出偵測網路誘拐的需求，此時聊天機器人會請使用者傳送想要檢測的聊天紀錄，並告知模型檢測的結果（圖一），若該對話為網路誘拐，聊天機器人會對使用者詢問一系列反思性問題，反思性問題以開放性問題為主，包含如「Do you think he/she is a trustworthy person?」、「Do you think he/she ever lied to you?」等，讓兒少在回答問題的過程中重新思考與該網友交流的過程，以及「What would you do if he/she makes you feel uncomfortable?」、「Would you tell your parents that the risk of online grooming has been detected?」等問題讓兒少思索可能發生情況以及對該事件的應對方式，並於對話最後給予一些建議如「I think you can also find trusted friends or elders to chat with.」鼓勵兒少讓現實生活中的人介入事件（圖二）。



圖一、網路誘拐檢測功能展示



圖二、機器人向使用者詢問反思性問題

未來展望

1. 支援使用者以中文與聊天機器人交互及中文網路誘拐對話檢測。
2. 將聊天機器人設置於更多社群（Facebook、Instagram等）與遊戲平台，與該平台合作，使聊天機器人能夠自動取得並檢測平台上兒少與他人的對話內容，使用者可自行關閉檢測。
3. 當使用者和陌生網友聊天內容被檢測為具有風險時，聊天機器人將自行跳出並中斷兒少與該網友的對話，若兒少與他人的對話持續被檢測為高風險，系統將會向由使用者自行設置的緊急聯絡人傳送通知。

獲獎紀錄

- 通過111年度國科會大專學生研究計畫（編號：111-2813-C-194-020-E）
- 2022 台灣人機互動研討會（TAICHI）海報論文最佳人氣獎 第二名

參考文獻

Vogt, M., Leser, U., & Akbik, A. (2021, August). Early Detection of Sexual Predators in Chats. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 4985-4999).