

Poster: Neeko：為青少年建立更安全的網路環境以擊潰網路誘拐

王瑋華, 國立中正大學資訊工程學系, 台灣

施巧柔, 國立中正大學資訊工程學系, 台灣

程芙茵, 國立中正大學資訊工程學系, 台灣

濫用網路進行色情和誘拐已成為國際關注的主要問題 [2]，目前已有許多學者投入網路誘拐的研究，然而，提前預防兒少受害的研究資源仍然相對缺乏，因此我們決定結合網路誘拐檢測模型，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，作為檢測及進一步干預網路誘拐的媒介，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。本研究將結合問卷調查和專家訪談之結果，歸納並確立 Neeko 的設計方向，並在 Neeko 的原型完成建置後，招募數名兒少受試者體驗及操作聊天機器人 Neeko，透過分析與討論受試者的回饋以評估聊天機器人的成效，並依此總結未來改善的方向。

Additional Key Words and Phrases: Data Mining, Natural Language Processing, Chatbot

1 研究動機與目的

在網路科技蓬勃發展的時代，隨著智慧型手機、個人電腦的普及，兒童及青少年交友的媒介也正在改變，2019 年臺灣兒童福利聯盟一項研究調查顯示 [10]，有超過八成的兒童曾透過網路結交陌生網友。然而，心智尚未發展成熟的兒少缺乏在網路上交友的正確觀念，網路所帶來的可匿名性 [9] 與便利性讓心懷不軌的成年人能夠輕易地偽裝成兒少的一員，以假身分伺機接近，與兒少建立關係，最終達到性剝削的目的，上述行動便稱為網路誘拐。若兒童成為網路誘拐的受害者，其可能會因為自責心態和社會對於性剝削受害者的汙名化等影響，不願意主動對外告知 [9]，進而導致錯誤與傷害一再發生。因此，在誘拐犯與兒少建立關係的過程中，若是沒有適當的干預措施，網路誘拐得逞的機率與其造成的危害可能會相對提高許多。

人類在與聊天機器人互動時，若經過適當的引導，能夠更好地表達自我和說出更多關於個人的隱私性資訊 [3, 4]，並且，聊天機器人願意重複同樣的內容而不會感到無聊或失去耐心 [1]。從上述研究的參與者反饋中可看出，聊天機器人的匿名性和隱私性提供人們相對舒適的管道提供資訊。因此，聊天機器人或許能夠成為一個令兒少相對容易接受的互動管道。綜合上述結果，本研究設計了聊天機器人 Neeko 作為干預兒少和誘拐犯培養信任的媒介，以促進兒少思考和增加其尋求幫助的機會。當兒少暴露在一定網路誘拐風險的聊天室中，Neeko 將向使用者提出反思性問題，使兒少將注意力轉移到談話內容的準確性上 [6, 7]，誘發兒少進行批判性思考，以抑制網路誘拐成功的可能性。

2 研究方法

本研究旨在探討如何設計 Neeko 的互動方式，能夠有效促進兒少網路交友時的思考。研究分為兩個階段，第一階段統整出 Neeko 互動方式和設計方向，在 Neeko 原型實作完成後進入第二階段，將招募數名兒少受試者進行體驗和蒐集回饋。

研究的第一階段為 Neeko 互動方式的設計，我們歸納與整理過去文獻，統整出 Neeko 的主要設計方向，為了獲得目標使用者的實際回饋，我們以網路問卷的形式向 6 至 17 歲的兒少及其家長（主要照顧者），進行聊天機器人互動方式和主要功能相關意見的蒐集和接受度調查，並與專家組成焦點小組，透過和專家的討論和設計調整，進行 Neeko 原型的實作，通過迭代設計確認 Neeko 的外型與互動方式。第二階段我們招募 5 至 10 名 13 至 17 歲的青少年受試者作為實驗對象，針對 Neeko 的使用情境進行體驗和操作，利用實驗觀察法，觀察受

Authors' addresses: 王瑋華, 國立中正大學資訊工程學系, 嘉義縣, 台灣, dahua089@alum.ccu.edu.tw; 施巧柔, 國立中正大學資訊工程學系, 嘉義縣, 台灣, sql101@alum.ccu.edu.tw; 程芙茵, 國立中正大學資訊工程學系, 嘉義縣, 台灣, fuyincherng@cs.ccu.edu.tw.

測者在不同情境下的使用狀況，並在體驗結束後進行半結構式訪談以了解受測者對 Neeko 的體驗感受，最後將所得的資訊進行分析與討論。

3 目前成果

3.1 聊天機器人設計

聊天機器人 Neeko 將結合網路誘拐檢測模型，偵測使用者與陌生人的對話是否存在網路誘拐的風險，當偵測到使用者受到網路誘拐的風險時，會傳送給使用者反思性問題並提醒使用者需要注意聊天室的交友關係，根據研究 [6, 7] 中指出，將人們的注意力轉移到準確性上可以提高他們對於判斷假新聞的準確性，因此，利用 Neeko 向兒少提出反思性問題，將兒少的注意力轉移到網路聊天對象所說內容的準確性上，藉此增加兒少對於該內容的思考，反思性問題例如：「這個人說的話是可信的嗎？」、「他說的資訊是不是該查證一下呀？」。Neeko 對聊天內容的檢測可以由使用者關閉，預設 Neeko 加入前已存在的對話與好友為「關閉檢測」，之後新增的對話及好友為「開啟檢測」。

協助孩子主動思考是 Neeko 的主要目標，而若孩子持續和陌生網友來往，並且聊天內容持續被檢測為高風險時，我們將會向緊急聯絡人傳送通知（使用者可自行設置緊急聯絡人），以讓大人介入協助了解情況。Neeko 將常駐於使用者的聊天室，並且在初次使用時顯示歡迎訊息並介紹機器人的用途，使用者可以隨時開啓機器人，開啓後聊天介面將會顯示簡短的問答句，當使用者想對話時可供選擇，向 Neeko 發送該問答句，使用者也可以直接輸入任何句子。

3.2 問卷調查與結果

採調查研究法，透過 SurveyCake 雲端問卷服務於網路發放問卷。為了蒐集兒少和其家長的意見，分別設計了家長（主要照顧者）和兒少調查問卷，招募 6 至 17 歲的兒少及家中有兒少之家長進行填答。最終兒少調查共計回收 23 份有效問卷，男性 7 位、女性 16 位，年齡眾數落在 15~17 歲。家長調查共計回收 31 份有效問卷，男性 13 位、女性 18 位，年齡眾數落在 40~49 歲。

問卷共分為四個部分，第一部分調查受試者之年齡、性別；第二部分針對聊天機器人和網路誘拐進行介紹，並說明目前本研究預期實作之聊天機器人的主要功能設計；第三部份為單選題，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，根據受試者對題目的理解進行填答；第四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，兒少問卷中針對 Neeko 的形象偏好設計了複選題，提出了一些相關選項如性別、是否為人類、形象風格（嚴肅、溫和親切、俏皮活潑），並提供讓受試者自行填寫回覆的「其他」選項。

根據所得數據，我們針對第二部分各題進行平均值和標準差之計算，如表1和表2。針對「使用意願」的調查，表1和表2題1之平均值皆超過3，顯示兒少及家長都有意願嘗試使用此聊天機器人；針對「向緊急聯絡人傳送通知」的功能，表2中題2、題4和表1題2結果平均值均超過3，顯示家長及兒少對於此功能具正面的看法；針對「聊天機器人對隱私（聊天內容）的侵犯」，表1題5平均值超過3.5而表2題3平均值低於3，顯示兒少對於自己和其他人對話隱私可能被侵犯的顧慮，而家長在此議題上和兒少有相對不同的看法。

第四部分的開放意見調查中，對於兒少問卷「偏好的聊天機器人形象」的複選題，得票結果如表3，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」的形象顯示，這樣的結果可能是因為部分使用者會對「機器人認為自己是人類」的想法感到不適，或是當使用者意識到對話對象為機器人時，人類的外型會讓他們覺得受到欺騙[5]。

針對開放性問答，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」、「在我無聊的時候陪我玩」等，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，但不會過於的去影響自己的想法」、「更直接一點，如果覺得是騙人的，就說是騙人的」等，其中，希望聊天機器人的互動方式能更直接的受試者皆為 6~11 歲，與過去文獻 [8] 的論點相符，年齡較小的兒童更傾向於聊天機器人給予直接的指示。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」、「可能可以與警察機關做配合，蒐集更多犯罪資料，連同詐騙回報一同綁定。」等。

表 1. 兒少問卷之平均值及標準差計算表

編號	題目	平均值	標準差
1	我會想要使用此聊天機器人。	3.22	0.74
2	我能接受在我身處網路誘拐風險時，此聊天機器人向緊急聯絡人傳送通知。	3.43	0.73
3	我不信任系統對於網路誘拐的判斷結果。	3.43	0.90
4	比起家長／我的主要照護者干涉我的交友情況，我更能接受聊天機器人和我互動。	3.30	0.82
5	我認為使用此聊天機器人可能會侵犯我和他人對話的隱私。	3.70	0.76

表 2. 家長問卷之平均值及標準差計算表

編號	題目	平均值	標準差
1	我會想讓我的孩子使用此聊天機器人。	3.52	1.06
2	我認為在檢測出高誘拐風險時向家長傳送通知能讓家長了解情況，進而有效阻止網路誘拐。	3.90	0.79
3	我認為使用此聊天機器人可能會侵犯孩子和他人對話的隱私。	2.87	0.99
4	我認為直接向緊急聯絡人傳送通知才是相較於聊天機器人在第一時間和孩子溝通更好的處理方式。	3.39	0.92
5	檢測系統的判斷未必準確，我認為收到並非百分之百準確的通知會造成不必要的麻煩，這降低我使用此聊天機器人的意願。	2.77	1.09

表 3. 兒少對於聊天機器人之形象偏好得票數統計表

形象	溫和親切的	俏皮活潑的	非人類	非二元性別 ／沒有性別	女性	人類
得票數	16	10	10	6	6	5

4 未來研究規劃

根據問卷調查結果可初步探究兒少及其家長（主要照顧者）對於 Neeko 之看法，後續本研究將結合與專家訪談的結果，完成 Neeko 原型之建置，並招募兒少受試者體驗及操作聊天機器人 Neeko，透過半結構式訪談以了解其使用上的實際感受及回饋，並統整出未來可調整方向。

參考文獻

- [1] Luke Fryer and Rollo Carpenter. 2006. Bots as language learning tools. *Language Learning & Technology* 10, 3 (2006), 8–14.
- [2] Sylvia Kierkegaard. 2008. Cybering, online grooming and ageplay. *Computer Law & Security Review* 24, 1 (2008), 41–55.
- [3] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.
- [4] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.
- [5] Andreea Muresan and Henning Pohl. 2019. Chats with Bots: Balancing Imitation and Engagement. In *CHI '19 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '19 (CHI EA '19)*. Association for Computing Machinery, USA. <https://doi.org/10.1145/3290607.3313084> null ; Conference date: 04-05-2019 Through 09-05-2019.
- [6] Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand. 2021. Shifting attention to accuracy can reduce misinformation online. *Nature* 592, 7855 (2021), 590–595.
- [7] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G Lu, and David G Rand. 2020. Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science* 31, 7 (2020), 770–780.
- [8] Lara Schibelsky Godoy Piccolo, Pinelopi Troullinou, and Harith Alani. 2021. Chatbots to Support Children in Coping with Online Threats: Socio-Technical Requirements. In *Designing Interactive Systems Conference 2021 (Virtual Event, USA) (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1504–1517. <https://doi.org/10.1145/3461778.3462114>
- [9] Janis Wolak, Kimberly J Mitchell, and David Finkelhor. 2006. Online Victimization of Youth: Five Years Later. (2006).
- [10] 兒童福利聯盟. 2019. 兒少使用社群軟體狀況調查報告。 https://www.children.org.tw/publication_research/research_report/726