**The Dark Side of Artificial Intelligence – Risks Arising in Dating Applications**

Rachel Fletcher, Calli Tzani & Maria Ioannou

Rachel Fletcher obtained a first-class Bachelor of Science degree in Psychology in 2022 (Forensic Psychology) at Nottingham Trent University, and a Distinction in Investigative Psychology MSc at the University of Huddersfield in 2023. She is a current PhD student researcher at the University of Huddersfield investigating sextortion, stalking and online criminal activity.

Dr Calli Tzani is a senior lecturer of Investigative and Forensic Psychology at the University of Huddersfield, Deputy Director of the Applied Criminology and Policing Centre, and the Forensic Consultant Editor of ADM. She is a Fellow of the Higher Education Academy, and Associate Fellow of the British Psychological Society.

Dr Maria Ioannou is a Professor of Investigative and Forensic Psychology at the University of Huddersfield, Course Director of the MSc Investigative Psychology and Course Director of the MSc Security Science. She is a Chartered Forensic Psychologist, Chartered Manager, Fellow of the Higher Education Academy, HCPC Registered Practitioner and Associate Fellow of the British Psychological Society.

**Key Digestive Message**

Hiding behind a smartphone screen, online dating applications provide a playground of opportunity for fraudsters and scammers. With ease of access to artificial intelligence, the technological capabilities of nefarious individuals are quickly growing. From sophisticated chatbots designed to engage in conversations and extract personal data, to deepfake technology used to create convincing false personas. This article summarises the current and upcoming risks which artificial intelligence poses to dating application and social media users. Deepfake

technology is a key risk; the world is experiencing greater use of attractive deepfake images to convince dating app users into involvement in a romance scam, face-swaps to target and blackmail social media users with their intimate images, and instant generation of child sexual abuse material. Other risks include stalkers tracking their victims with greater ease, and individuals downloading nefarious dating applications which utilise chatbots to gather information and get paid. Gaps in empirical research are identified and discussed.

## Introduction

Online dating applications offer a land of opportunity to fraudsters and scam artists, with platforms such as Tinder, Hinge and Match.com becoming conventional ways to establish relationships, short-term connections, and hook-ups. These platforms, including social media sites, allow users to date multiple users at once, 24/7 (Lamphere & Lucas, 2019). On top of this, disclosure about oneself on social media is pervasive (Luo & Hancock, 2020). The ever-increasing popularity of dating applications and social media sites mean that cyber-criminals have easy access to a pool of potential victims, making the potential for deceit, fraud, and technological-facilitated violence significant concerns for public protection (Zhou, 2023). This is further exacerbated by the introduction of artificial intelligence, which despite its benefits, may offer perpetrators new and advanced ways to exploit and manipulate others (Hayward & Maas, 2021; Blauth, Gstrein & Zwitter, 2022).

Artificial intelligence (AI) refers to the development of computer systems with the capabilities to perform tasks and make decisions which typically require human intelligence. From sophisticated chatbots designed to engage in conversations and extract personal data, to deepfake technology used to create convincing false personas, evidence suggests that criminologists must prime themselves for the dark side of this technology, and how artificial intelligence will influence online safety. The aim of this article is to summarise the current and

potential future risks arising in the media and empirical literature, surrounding AI-related safety concerns on dating applications.

## Methodology

This study gathered sources from general internet searches, using key search terms such as "artificial intelligence", "dating applications", "social media", "romance fraud", "sextortion", and a combination of these items together. Empirical literature was explored to examine the current state of the topic. The findings reflect concerns raised from across both source types.

## Findings

### Deepfake Technology

Deepfake technology utilises deep learning algorithms to manipulate audio-visual content, such as videos, images, or voice recordings, so that it appears that someone said or did something they never actually did. Key aspects include face-swapping and voice synthesis, manipulated to high realism so that it is difficult to identify deepfake material. While deepfake technology holds high potential in the entertainment industry, deepfake technology appears to be one of the most prevalent concerns affecting the authenticity of dating applications and potential for identity theft and impersonation.

***Detecting a catfish.*** The use of a fake profile, often using fake imagery, is a behaviour termed "catfishing" and depicted in the popular MTV series *Catfish* (MTV, 2020). Catfishing is a common precursor to several online crimes, namely romance fraud (Whitty, 2013; Cross, 2022) and sextortion (Carlton, 2019). Until recently anecdotal evidence has suggested that catfishing involves some form of identity theft; offenders take images of genuine persons and display them as their own (Cross & Layt, 2022). While this devastatingly means that catfishing involves two victims - both the deceived individual and the individual whose photographs were stolen - this has acted as a protective measure for identifying fraudulent profiles. Prevention messaging often suggests that internet-users conduct a reverse image search to verify the

identity of a person they are speaking to online, for instance, popular site WebMD states to "use an online reverse image search to find out if the person's photos are on anyone else's online profiles" (Sachdev, 2022). As recently suggested by Cross (2022), advancements in deepfake technology mean that perpetrators are now able to use entirely original content, unidentifiable by a reverse image search, raising serious consequences if the public are to rely on reverse image search results to identify a fraudulent profile.

***Sextortion blackmail.*** Sextortion is an online threat referring to perpetrators who blackmail social media and dating app users through threatening to disseminate their sexually explicit material. On the 5th of June 2023, the FBI produced an alert stating that they had observed an increase in the use of deepfakes in sextortion cases, with perpetrators using artificial intelligence to manipulate content posted by targets on social media into realistic, sexual images (FBI, 2023). This is a significant concern since it removes the need for victims to unknowingly share their sexually explicit content with a perpetrator before being blackmailed, as has typically been required for sextortion (Gamez-Gaudix et al, 2015). Due to artificial intelligence enabling perpetrators to morph the faces of dating app users from innocent photographs onto sexual images, any user who shares content online is more easily at risk of being blackmailed with the exposure of sexual imagery. Sextortion has already been linked with severe mental health consequences and suicide (Nilsson et al, 2019) and some victims have shared additional implications of deepfake sextortion such as body dysmorphia (Hendery, 2023).

***The proliferation of child sexual abuse material (CSAM).*** The opportunity for law enforcement to better detect and investigate child sexual abuse has been a positive influence of artificial intelligence, as discussed in several studies (Sanchez et al, 2019). However, ease of access to these technologies creates risk to the proliferation of child sexual abuse material in the online space. The Internet Watch Foundation describes this scenario as their "worst

nightmare" come true; in just one month, they detected 11,108 AI-generated images shared on a dark web child abuse forum, 2,978 of which depicted child sexual abuse, and some involve newly generated imagery of real victims of sexual abuse (IWF, 2023).

Ofcom (2022) reveals the critical risk this could pose to children on social media, showing that 62% of children (aged between 3-17 years old) have their own profile on at least one social media application, including 33% of five- to seven-year-olds and 60% of eight- to ten-year-olds. Although the images and videos shared by children of themselves on social media are typically innocent, artificial intelligence can morph these images into CSAM. Even without the addition of blackmail as seen in sextortion cases, the psychological and developmental impact that generation of these images poses on children is a worrying unknown.

**Automated Bots and Phishing**

Automated bots, driven by artificial intelligence, are designed to simulate human behaviour and engage users in conversations which appear genuine. Within dating applications they can initiate contact with real online users, convincing them to share personal information, click on malicious links, or provide financial details. This is termed phishing. Limited studies have highlighted malware authors utilising chatbots to feign an entire online dating application and receive payment, as discussed below.

***Mobile malware.*** Several years ago, Hu et al (2019) conducted the first systematic study on fraudulent dating apps. It was uncovered that malware authors were luring internet-users into installing fraudulent dating apps with subscription fees, yet instead of speaking to other genuine accounts, the existing accounts were fake identities managed by chatbots. For over 70% of apps identified of this nature, users were required to pay additional money to reply to the chatbots messages, which often involve seduction (Hu et al, 2019). To emphasise the potential impact of this malware, this study estimated the lower bound of potential victims to

be 30,932 based on user complaints, and the upper bound to be 2.5 million based on number of downloads, at risk of data breaches and/or financial loss. Since, research investigating fraudulent dating apps has been scarce, although it can be speculated that with the advancements of artificial intelligence these applications may now be more sophisticated. For instance, the chatbots were partly identified by Hu et al (2019) by their nonsensical responses to conversation, whereas chatbot technology now more successfully simulates human conversation. There is a need to understand the recent prevalence and nature of fraudulent dating applications, and to develop solutions to block the distribution of these.

**Cyberstalking and harassment.**

Using artificial intelligence, an article by Fox News quotes "a stranger in a coffee shop can watch you and learn virtually everything about you, where you've been, and even predict your movements" (Eberhart, 2023). Cyberstalking and harassment is another area where AI could have dangerous consequences, being used as a surveillance tool to track victims with greater ease.

***Facial recognition software.*** According to privacy campaign group the Big Brother Watch, facial "search engines" are available to upload a photo of a person's face and identify matches from social media, blog posts, and websites (Big Brother Watch Team, 2022). Working differently to a reverse image search which identifies duplicates of images on the internet, this advanced AI can find every instance of your face on the internet with just an online photograph or in person snapshot, creating concern that a stalker could find someone's workplace or indications of where they live uploaded onto private or previously unidentified social media sites. Some platforms are publicly available (PimEyes; Big Brother Watch Team, 2022) and others are reduced to law enforcement access only (Clearview AI; BBC, 2022). Regardless, this raises concern that those with access to these databases could exploit them.

## Conclusion and Future Research Directions

Despite anecdotal evidence sharing concern about the impact of artificial intelligence on persons safety, it is evident that much of the empirical literature investigates the positives of this technology. For instance, the immediate titles when searching for stalking, sextortion, and fraud along with artificial intelligence reflect the benefits of AI to police investigation and crime deterrence, with "how technology can mitigate and counteract cyber-stalking and online grooming" (al-Khateeb & Epiphaniou, 2016) serving as an example of these search results. Not disregarding these positive uses there are several research gaps pertaining to the dark side of artificial intelligence. Specifically, attention should be paid to the psychological consequences, such as possible body dysmorphia, and emotional support systems available to victims targeted by deepfake technology. Detection and prevention strategies to identify deepfake material and stop its distribution is also crucial, since it is a reality that these are increasing in circulation.

There's very limited research specifically focusing on how AI, such as facial recognition, contributes to harassment and stalking incidents, and whether this is a reality. Not unrelated to the speed by which technology develops, research is also outdated on fraudulent dating applications and how these are being used to target individuals. Empirical literature should seek to understand the genuine risks posed to dating application users, in order to get ahead of the problem and identify solutions. Aside from research, there is an immediate need for education about the possibilities of artificial intelligence so that the public are able to better identify the methods being used by criminals online and protect their privacy. For instance, as Cross (2022) suggested, messaging surrounding the use of reverse image search machines to catch out catfish perpetrators appears outdated given increased use of deepfakes.

## References

Big Brother Watch. (November 8, 2022). BBC – Pimeyes, a Facial "Search Engine" Could be Used by Stalkers. https://bigbrotherwatch.org.uk/2022/11/pimeyes-a-facial-search-

engine-could-be-used-by-stalkers/#:~:text=Privacy%20group%20Big%20Brother%20Watch,of%20individuals%20in%20the%20UK. [Accessed 5 December 2023).

Blauth, T. F., Gstrein, O. J., & Zwitter, A. (2022). Artificial intelligence crime: An overview of malicious use and abuse of AI. *IEEE Access*, *10*, 77110-77122.

Carlton, A. (2019). Sextortion: The Hybrid Cyber-Sex Crime. *NCJL & Tech.*, *21*, 177.

Cross, C. (2022). Using artificial intelligence (AI) and deepfakes to deceive victims: the need to rethink current romance fraud prevention messaging. *Crime Prevention and Community Safety*, *24*(1), 30-41.

Cross, C., & Layt, R. (2022). "I suspect that the pictures are stolen": Romance fraud, identity crime, and responding to suspicions of inauthentic identities. *Social Science Computer Review*, *40*(4), 955-973.

Eberhart, C. (May 30, 2023). Who is watching you? AI can stalk unsuspecting victims with 'ease and precision': experts. Fox News. https://www.foxnews.com/us/who-is-watching-you-ai-can-stalk-unsuspecting-victims-ease-precision-experts. [Accessed 5 December 2023]

Federal Bureau of Investigation (June 5, 2023). Alert Number I-060523-PSA. Malicious Actors Manipualting Photos and Videos to Create Explicit Content and Sextortion Schemes. https://www.ic3.gov/Media/Y2023/PSA230605. [Accessed 5 December 2023]

Gámez-Guadix, M., Almendros, C., Borrajo, E., & Calvete, E. (2015). Prevalence and association of sexting and online sexual victimization among Spanish adults. *Sexuality Research and Social Policy*, *12*, 145-154.

Hayward, K. J., & Maas, M. M. (2021). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, *17*(2), 209-233.

Hendery, S. (2023). Deepfakes of victims used in sextortion attacks spike, FBI warns. SCMedia. https://www.scmagazine.com/news/deepfakes-sextortion-spike-fbi. [Accessed 5 December 2023]

Hu, Y., Wang, H., Zhou, Y., Guo, Y., Li, L., Luo, B., & Xu, F. (2019). Dating with scambots: Understanding the ecosystem of fraudulent dating applications. *IEEE Transactions on Dependable and Secure Computing*, *18*(3), 1033-1050.

Internet Watch Foundation. (25 October 2023). 'Worst nightmares' come true as predators are able to make thousands of new AI images of real child victims. https://www.iwf.org.uk/news-media/news/worst-nightmares-come-true-as-predators-are-able-to-make-thousands-of-new-ai-images-of-real-child-victims/.

Lamphere, R. D., & Lucas, K. T. (2019). Online romance in the 21st century: Deceptive online dating, catfishing, romance scams, and "mail order" marriages. In *Handbook of Research on Deception, Fake News, and Misinformation Online* (pp. 475-488). IGI Global.

Luo, M., & Hancock, J. T. (2020). Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Current opinion in psychology*, *31*, 110-115.

Nilsson, M. G., Pepelasi, K. T., Ioannou, M., & Lester, D. (2019). Understanding the link between sextortion and suicide. *International journal of cyber criminology*, *13*(1), 55-69.

Ofcom. (2022). Children and parents: media use and attitudes report. Available at: Children and parents: media use and attitudes report 2022 [Accessed 5 December 2023].

Sachdev, P. (December 13, 2022). Signs of Catfishing. *WebMD. https://www.webmd.com/sex-relationships/signs-catfishing* [Accessed 5 December 2023).

Sanchez, L., Grajeda, C., Baggili, I., & Hall, C. (2019). A practitioner survey exploring the value of forensic tools, ai, filtering, & safer presentation for investigating child sexual abuse material (csam). *Digital Investigation*, *29*, S124-S142.

Vallance, C. (November 8, 2022). Stalking fears over PimEyes facial search engine. BBC. https://www.bbc.co.uk/news/technology-63544169. [Accessed 5 December 2023].

Whitty, M. T. (2013). The scammers persuasive techniques model: Development of a stage model to explain the online dating romance scam. *British Journal of Criminology*, *53*(4), 665-684.

Zhou, Y. (2023). The Benefits and Dangers of Online Dating Apps. *Canadian Journal of Family and Youth/Le Journal Canadien de Famille et de la Jeunesse*, *15*(2), 54-62.