Romance Scam Research Center

SOCIAL TECHNOLOGY AND SAFETY FOUNDATION

LIVING EVIDENCE SYNTHESIS

Technology and AI

Evidence Synthesis: Technology and AI in Romance, Catfishing, and Pig-Butchering Scams

Across the reviewed, human-curated evidence, social media and general-purpose platforms (not only dating apps) are repeatedly identified as common initiation channels for romance and relationship-based fraud; offenders frequently move conversations from those public-facing platforms into less-monitored messaging services to continue grooming and evade detection. Generative AI and deepfake tools are accessible and have been documented or plausibly attributed in some cases to support fabricated identities (images, voices, or videos), reducing the reliability of simple provenance checks such as reverse-image search. Empirical and qualitative studies document semi-industrialized offender practices (assembled digital infrastructure, role allocation, scripted materials) and partial automation (email-harvesting, semi-automated scripts, fake-profile farms), but the precise share of operations run end-to-end by large language models (LLMs) or fully automated systems is unknown. Romance-initiated investment fraud (“pig-butchering”) commonly uses fabricated trading dashboards, links/QR distribution, crypto transfers, mixers, and on-chain recirculation patterns that complicate full forensic attribution. Prototype detection and OSINT workflows (logistic classifiers, Maltego, reverse-image guidance) report promising in-sample results but are limited by dataset selection, validation gaps, privacy trade-offs for dynamic

message analysis, and uneven external evaluation. Methodologically, the corpus is heterogeneous and often relies on purposive, self-selected, or secondary-sourced datasets; authors consistently identify these constraints and call for interdisciplinary, privacy-sensitive real-world evaluations, living forensic syntheses, and targeted empirical measurement of AI’s operational role in scams.

Published Sep 6, 2026 Updated Sep 6, 2026 Synthesis version 67

PUBLICATIONS	EVIDENCE STATEMENTS	CITED SOURCES	SYNTHESIS VERSION
51	**187**	**51**	**67**

Across the reviewed, human-curated evidence, social media and general-purpose platforms (not only dating apps) are repeatedly identified as common initiation channels for romance and relationship-based fraud; offenders frequently move conversations from those public-facing platforms into less-monitored messaging services to continue grooming and evade detection. Generative AI and deepfake tools are accessible and have been documented or plausibly attributed in some cases to support fabricated identities (images, voices, or videos), reducing the reliability of simple provenance checks such as reverse-image search. Empirical and qualitative studies document semi-industrialized offender practices (assembled digital infrastructure, role allocation, scripted materials) and partial automation (email-harvesting, semi-automated scripts, fake-profile farms), but the precise share of operations run end-to-end by large language models (LLMs) or fully automated systems is unknown. Romance-initiated investment fraud (“pig-butchering”) commonly uses fabricated trading dashboards, links/QR distribution, crypto transfers, mixers, and on-chain recirculation patterns that complicate full forensic attribution. Prototype detection and OSINT workflows (logistic classifiers, Maltego, reverse-image guidance) report promising in-sample results but are limited by dataset selection, validation gaps, privacy trade-offs for dynamic message analysis, and uneven external evaluation. Methodologically, the corpus is heterogeneous and often relies on purposive, self-selected, or secondary-sourced datasets; authors consistently identify these constraints and call for interdisciplinary, privacy-sensitive real-world evaluations, living forensic

syntheses, and targeted empirical measurement of AI's operational role in scams. [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16]

Platform ecology and channel shifting

Multiple empirical sources identify social media and general-purpose platforms (not only purpose-built dating apps) as frequent initiation channels for romance and relationship-based fraud. Large complaint and survey datasets and qualitative reports repeatedly list Facebook, Instagram and similar social-network sites among the most commonly reported first-contact origins. [1] [17] [18]

Authors document a consistent pattern of platform mixing: initial contact often begins on a visible public platform and then shifts to less-monitored messaging applications (for example WhatsApp, Telegram, Google Chat) where grooming and payment requests continue. Some studies report that operations commonly employ two to three distinct applications during a single scam. [19] [20] [2]

Platform affordances and public profile content are described as enabling targeted, tailored narratives: publicly visible friend lists, hobby posts, and profile details provide material offenders use to craft plausible 'love-story' pretexts. Researchers therefore highlight social-media profile structure and algorithmic friend-suggestion features as factors that can facilitate target selection. [21] [22] [23]

AI, deepfakes, and synthetic media

Multiple sources document that accessible deepfake and generative-AI tools can produce realistic synthetic images, voices, and videos and that these tools require lower technical thresholds than earlier synthetic-media methods. Authors explicitly link these capabilities to potential increases in deceptive plausibility for romance fraud and catfishing. [3] [5]

Researchers and commentators emphasize practical consequences for verification: reverse-image and provenance checks sometimes fail (for example, searches returning no matches), and AI-generated unique images reduce linkability to a real source, undermining existing consumer-facing verification strategies. [8] [4] [24] [14]

The reviewed evidence does not support a validated estimate of how many operations are fully automated by LLMs or end-to-end synthetic systems. Several authors note this proportion is unknown and call for empirical measurement of (a) the operational role of generative AI in live

scams and (b) the effectiveness of deepfake detection tools in real-world settings. [25] [15] [26] [27]

Automation, industrialization, and offender infrastructure

Qualitative and courtroom-sourced studies document assembled digital infrastructure used by offenders: false social-media and dating profiles, dedicated email accounts, foreign phone numbers, SIM changes, VPNs, and centralized IT resources that support multi-step operations and laundering. This evidence supports the characterization of some operations as semi-industrialized rather than ad hoc individual efforts. [28] [29] [30] [31]

Empirical work also documents partial automation in acquisition and early stages: email-harvesting, automated initial-contact processes, and honeypot experiments indicate low but measurable automated engagement rates; yet platform-specific studies show significant manual interaction remains necessary for sustained romance scams, which limits claims of full automation. [32] [33] [34] [35]

Authors caution that technical countermeasures (improved verification, fake-profile detection) raise an adaptive risk: offenders change impersonated identities and tactics when scripts become well known, so detection improvements can shift offender behavior or migration across platforms rather than fully eliminating threats. [36] [37] [38] [39]

Pig-butchering, cryptocurrency interfaces, and forensic tracing

Multiple mixed-method and forensic studies characterize ‘pig-butchering’ as a hybrid romance-to-investment fraud where long-term trust-building funnels victims into fabricated trading dashboards or apps. Empirical case series and reviews document distribution of fake apps/websites via links or QR codes and use of simulated dashboards to sustain investment-style manipulation. [7] [6] [40] [10]

Single-case and small-series forensic investigations combine OSINT, Maltego link analysis, and on-chain/off-chain linkage to produce partial attribution (wallet → social account → platform), but authors repeatedly report limitations such as missing email headers, unverifiable wallet–identity links, and incomplete subpoenaed data that prevent conclusive attribution in many cases. [8] [11]

[41]

Because of these attribution limits, the literature recommends living forensic syntheses and real-time collaboration between researchers and forensic analysts to keep pace with laundering techniques and to validate methods on continuously updated seeds and ground truth when available rather than relying on static, retrospective datasets alone. [10] [12]

Detection prototypes, OSINT workflows, and privacy trade-offs

Other technical detection approaches include automated reverse-image searches, stylometric and LIWC analyses, multimodal systems combining behavioral, IP, photographic and text signals, and proposed voice-biometric tools to identify impersonation. Authors emphasize contextual and language limits of lexical tools and recommend combining modalities to reduce false positives. [27] [42] [43] [44]

Prototypes in the batch combine open-source OSINT data collection (including Maltego transforms on romance-scam websites), TF-IDF feature extraction on cleaned emails and links, and a logistic-regression classifier that assigns a suspiciousness score to each data point. One reported evaluation, using a 75/25 train/test split on a dataset compiled from online resources and Maltego-derived information, produced 85% accuracy, 88% precision, 80% recall, and an F1 score of 0.84. Future work noted by the authors includes expanding the dataset and exploring more complex machine-learning algorithms. [12]

Several authors explicitly identify privacy and governance trade-offs: dynamic analysis of private messages could improve early warning or trigger advisories but would intrude on user privacy and raise acceptability questions. Work in this area calls for research on user acceptance, transparency, bias, and governance before operational deployment of message-monitoring interventions. [13]

Authors recommend partnership research with platforms to access higher-quality labeled data and to test detection systems under realistic, threat-adaptive conditions rather than relying solely on in-sample performance claims from curated OSINT datasets and laboratory splits alone. Future work proposals include integrating more complex ML, expanding scenario coverage, and conducting platform-partnered evaluations. [12] [45] [46]

Methodological patterns, limitations, evidence gaps, and future directions

The reviewed corpus is methodologically heterogeneous: qualitative interviews, complaint and administrative reports, corpus linguistics, blockchain forensics, case files, offender interviews, and prototype evaluations are all present. Many studies rely on purposive, self-selected, or secondary-sourced datasets (public complaints, forums, news reports), which authors consistently identify as limiting generalizability and prevalence estimation. [1] [4] [15] [3] [12]

Commonly reported methodological limitations include uncertain offender attribution (VPNs, caller-ID masking), incomplete forensic seed data, unverifiable victim-reported addresses, small purposive samples, and dataset contamination (controls containing unlabelled scam profiles). Researchers explicitly link these constraints to cautious interpretation of prevalence or causality claims. [2] [8] [11] [46] [47] [48]

Evidence gaps identified within the literature (and explicitly recommended by authors) include: (a) empirical measurement of how generative AI and LLMs are operationally used in live scams, (b) real-world validation of automated detection systems with platform data, (c) living forensic syntheses that integrate on-chain and off-chain evidence, and (d) robust evaluation of privacy-sensitive early-warning or advisory systems. [25] [27] [10] [12] [13] [49]

Across the corpus, authors repeatedly call for interdisciplinary, platform-partnered research that balances technical detection, legal and policy remedies, and ethical acceptability, and they caution that any deployed interventions must be evaluated for false-positive risk, bias, privacy intrusion, and likely offender adaptation. [50] [13] [51] [31]

References

1. Grace Carvalho Fernandes, S.; BASTOS DO NASCIMENTO, V.; Castelo Branco Do Nascimento, V. (2023). Romance Scam: Victim Manipulation and Human Trafficking DOI
2. Faber, P. (2024). The Frames of Romance Scamming DOI
3. Cross, C. (2022). Using artificial intelligence (AI) and deepfakes to deceive victims: the need to rethink current romance fraud prevention messaging DOI

4. Cross, C. (2023). “I knew it was a scam”: Understanding the triggers for recognizing romance fraud DOI
5. Gauci, Christine; Vella, Mary Grace (2026). Love and Technology DOI
6. Cross, C. (2023). Romance baiting, cryptorom and ‘pig butchering’: an evolutionary step in romance fraud DOI
7. Wang, F.; Zhou, X. (2022). Persuasive Schemes for Financial Exploitation in Online Romance Scam: An Anatomy on Sha Zhu Pan (□□□) in China DOI
8. Botha, Johannes George; Singh, Kreaan; Leenen, Louise (2025). Analysis of a Cryptocurrency Investment Scam: Pig Butchering DOI
9. Ditasya Anisa Riani; Ruslan Abdul Gani; Maryani, Maryani (2026). Catfishing on Social Media: A Criminal Law And Islamic Criminal Law Analysis In Jambi DOI
10. Gujarathi, Palak; Verma, Shankey; Nair, Vipin Vijay (2026). Pig Butchering Scams as Cyber-Enabled Financial Crime: A Scoping Review of Dimensions, Modus Operandi, and Victim-Offender Dynamics DOI
11. Griffin, John M.; Mei, Kevin (2024). How Do Crypto Flows Finance Slavery? The Economics of Pig Butchering DOI
12. Vedhanayagam, Priya; Singh, Moulik; Dadhania, Arora Preksha; Ikram, Sumaiya Thaseen (2026). Forensic footprints in digital love: Unveiling romance scams with Maltego and machine learning DOI
13. Dickerson, S.; Apeh, E.; Ollis, G. (2020). Contextualised Cyber Security Awareness Approach for Online Romance Fraud DOI
14. Kassem, R.; Carter, E. (2023). Mapping romance fraud research – a systematic review DOI
15. Cross, C.; Holt, TJ. (2023). More than Money: Examining the Potential Exposure of Romance Fraud Victims to Identity Crime DOI
16. Shaari, AH.; Kamaluddin, MR.; Paizi@Fauzi, WF.; Mohd, M. (2019). Online-Dating Romance Scam in Malaysia: An Analysis of Online Conversations between Scammers and Victims DOI
17. Dr. Rhem Rick N. Corpuz; Mary Joy G. Galang; Kim Hope S. Gueco; Kate D. Pamintuan (2025). ROMANCE AND RUIN: THE INTERPLAY OF FINANCIAL LOSS AND PSYCHOLOGICAL WELL-BEING AMONG LOVE SCAM VICTIMS IN THE PHILIPPINES DOI
18. Alavi, Khadijah; Mahbob, Maizatul Haizan; Sooed, Mohammad Syahrul Azha (2020). Strategi Komunikasi Penjenayah Cinta Siber Terhadap Wanita Profesional DOI

19. Dickinson, T.; Wang, F.; Maimon, D. (2023). What Money Can Do: Examining the Effects of Rewards on Online Romance Fraudsters' Deceptive Strategies DOI
20. Ordekian, Marilyne; Papasavva, Antonis; Mariconti, Enrico; Vasek, Marie (2024). A Sinister Fattening: Dissecting the Tales of Pig Butchering and Other Cryptocurrency Scams DOI
21. Christian Kopp; Robert Layton; Jim Sillitoe; Iqbal Gondal (2016). The Role of Love stories in Romance Scams: A Qualitative Analysis of Fraudulent Profiles DOI
22. Unknown (2023). The Crime of "Pig-butchering Scams" in the Securities Market and Legal Regulation DOI
23. Lamphere, RD.; Lucas, KT. (2019). Online Romance in the 21st Century DOI
24. Cross, C.; Layt, R. (2021). "I Suspect That the Pictures Are Stolen": Romance Fraud, Identity Crime, and Responding to Suspicions of Inauthentic Identities DOI
25. Dominguez Castillo, Lorena (2026). Industrialized heartbreak: how generative AI enables romance fraud at scale DOI
26. Herrera, LD.; Hastings, J. (2024). The Trajectory of Romance Scams in the U.S DOI
27. Shapiro, LR. (2022). Online Romance Scammers DOI
28. Abubakari, Yushawu; Lazarus, Suleman; Oseh-Ovarah, Valeen (2026). A crime script perspective on mapping the entry, continuation, and exit pathways of online romance fraud DOI
29. Yetunde O. Ogunleye; Ojedokun, Usman A.; Adeyinka A. Aderinto (2020). Pathways and Motivations for Cyber Fraud Involvement among Female Undergraduates of Selected Universities in South-West Nigeria DOI
30. Luong, Hai Thanh; Ngo, Hieu Minh (2024). Understanding the Nature of the Transnational Scam-Related Fraud: Challenges and Solutions from Vietnam's Perspective DOI
31. Hasibuan, Juneidi; Syam, Syafrudin (2023). A Legal Analysis on Online Fraud Using Fake Identity DOI
32. Atta-Asamoah, Andrews (2009). Understanding the West African cyber crime process DOI
33. Jakobsson, Markus (2016). Understanding Social Engineering Based Scams DOI
34. Huang, JingMin; Stringhini, Gianluca; Yong, Peng (2015). Quit Playing Games with My Heart: Understanding Online Dating Scams DOI
35. Robinson, Jemima; Edwards, Matthew (2024). Fraudsters target the elderly: Behavioural evidence from randomised controlled scam-baiting experiments DOI

36. Kim, Hyo-shin; Seo, Jun-bae (2019). A Study on Romance Scam : The Current Situation and Effective Countermeasures DOI
37. Abubakari, Yushawu; Oseh-Ovarah, Valeen (2025). The Gamification of Online Romance Fraud through Offenders' Cards DOI
38. Diallo, Ousmane; Hassan, Hasnani; Najeeb Zaidan, Muslim (2025). Leveraging Naive Bayesian Machine Learning for Detecting Pig Butchering Scams: A Cybersecurity Social Engineering Perspective in Africa DOI
39. Valentin Le Normand (2026). Romance Scam Statistics 2024-2026: verified, source-attributed figures from FBI IC3, FTC, City of London Police, Cybermalveillance.gouv.fr, GASA and Chainalysis (Baromètre 2026 des arnaques sentimentales) DOI
40. Franceschini, Ivan; Li, Ling; Bo, Mark (2023). Compound Capitalism: A Political Economy of Southeast Asia's Online Scam Operations DOI
41. Reiter, Jonathan; Team, Bitrace (2024). Connecting Chinese and American Scam Victims DOI
42. Ma, Katelyn Wan Fei; McKinnon, Tammy (2021). COVID-19 and cyber fraud: emerging threats during the pandemic DOI
43. Hamsi, Ahmad Safwan; Bahry, Farrah Diana Saiful; Tobi, Siti Noraini Mohd; Masrom, Maslin (2015). Cybercrime over Internet Love Scams in Malaysia: A Discussion on the Theoretical Perspectives, Connecting Factors and Keys to the Problem DOI
44. Toma, Catalina L.; Hancock, Jeffrey T. (2012). What Lies Beneath: The Linguistic Traces of Deception in Online Dating Profiles DOI
45. Rabby, F.; Chowdory, MMU. (2024). Romance Scamming: Uncovering the Transnational Crime and Legal Challenges DOI
46. Lee, KF.; Chan, MY.; Mohamad Ali, A. (2022). Self and desired partner descriptions in the online romance scam: a linguistic analysis of scammer and general user profiles on online dating portals DOI
47. Lee, KF; Chan, MY; Ali, Afida Mohamad (2024). Drawing on social approval as a linguistic strategy: A discourse semantic analysis of judgement evaluation in suspected online romance scammer dating profiles DOI
48. Khukhunaishvili, Dina (2024). Romance Scam as one of the main Challenges of cybercrime DOI
49. Sorell, T.; Whitty, M. (2019). Online romance scams and victimhood DOI

50. Boland, Michael James (2025). Developments in the Law Governing Online Activity: The Criminalisation of Catfishing and Civil Relief in Cases of Image-Based Sexual Abuse DOI

51. Krause, David (2025). The Deceptive Allure: Understanding and Combating Cryptocurrency Pig Butchering Scams DOI

This AI-assisted synthesis is based on reviewed evidence records. It may contain errors or omissions. Follow the publication and DOI links, consult the original works, and make your own judgment about the evidence.

About this document

This living evidence synthesis was produced through AI-assisted analysis of approved RSRC publications. Claims are linked to supporting evidence and checked for source consistency. It may change as the corpus is updated. Readers should consult the cited publications when making research, policy, or practice decisions.

Current web version: https://romancescamresearch.org/research-insights/technology-and-ai-in-relationship-based-fraud-synthesized-evidence-on-platforms-automation-synthetic-media-and-countermeasures