INVESTIGATION · METHOD AND THE FUTURE

The Future of Investigation in a Digital World

Almost every investigation now involves digital evidence, and the tools, sources and threats are changing quickly. This article examines the main trends shaping the future of investigation, including data volume, artificial intelligence, open-source intelligence, synthetic media, encryption, cross-border data and genetic genealogy, and the skills and safeguards investigators will need.

ModQu research article · 1,955 words · 24 references (APA 7th edition)

Twenty years ago, digital evidence was a specialism within investigation. Today almost every investigation involves it. A workplace complaint turns on chat messages, a fraud on transaction records, a missing-person case on phone location data, and a war crimes inquiry on videos posted online. The volume and variety of digital material keep growing, and so do the tools for analysing it and the ways it can be faked or hidden. This article examines the main trends that are shaping the future of investigation, the opportunities and risks they create, and what investigators and organisations will need to do to keep their findings reliable and fair.

From scarcity to abundance

For most of the history of investigation, the problem was too little evidence. In the digital world it is often too much. Garfinkel (2010) warned that the growth of storage capacity, the diversity of devices and file formats, and the spread of encryption and cloud storage would end what he called the "golden age" of digital forensics, in which a single hard drive held most of the evidence and could be examined fully with standard tools. Quick and Choo (2014) reviewed the evidence and found that the volume of digital data seized in investigations had grown faster than forensic capacity, leading to large backlogs in many police forces.

Abundance changes the investigator's task. Instead of searching for evidence, investigators must decide which of millions of items to examine, and must be able to explain why they did not examine the rest. Casey (2011) emphasises that digital evidence must still be handled according to forensic principles of preservation, documentation and verification, but these principles become harder to apply as volumes rise. The future investigator will spend less time finding material and more time selecting, filtering and validating it.

Trends and their consequences

Trends shaping investigation

Each trend brings both new opportunities and new risks

	Opportunity	Risk
Data volume	More traces of what happened	Backlogs; important items missed
Artificial intelligence	Search, transcription and matching at scale	Bias, invented details, automation bias
Open-source intelligence	Public can document events worldwide	Misidentification and harm to bystanders
Synthetic media	Detection tools improving	Fakes, and real evidence dismissed as fake
Encryption and the cloud	Stronger protection for victims' data	Evidence out of reach or held abroad
Genetic genealogy	Cold cases solved	Privacy of relatives who never consented

ModQu analysis of sources cited in this article ModQu.com

Figure 1. Six trends that are changing investigation. In each case the same development that helps investigators also creates new ways for investigations to go wrong. Diagram by ModQu.

Artificial intelligence

Artificial intelligence is the most obvious response to abundance. Machine learning already ranks documents for review, transcribes and translates speech, recognises faces and objects, and links records that refer to the same person. Large language models can summarise material and answer questions about it. These tools will become standard in investigations of every size.

They also carry well-documented risks. Face recognition algorithms vary widely in accuracy, and many have shown higher false match rates for some demographic groups (Grother et al., 2019). Language models can produce fluent but false statements (Ji et al., 2023). People tend to over-trust automated outputs and to stop looking for contradicting evidence (Parasuraman & Manzey, 2010). Models trained on past policing data can reproduce past patterns of enforcement rather than patterns of crime (Lum & Isaac, 2016).

Regulation is beginning to respond. The European Union's Artificial Intelligence Act classifies many uses of AI in law enforcement, including some biometric identification, as high-risk and imposes requirements for data quality, documentation, transparency and human oversight; it prohibits some uses outright (Regulation (EU) 2024/1689, 2024). Investigators in many countries are likely to face similar rules. The future skill is not operating AI tools but knowing when their outputs can be trusted and how to check them.

Data-driven surveillance

Digital data allow investigators to see patterns across whole populations rather than only within a case. Brayne (2017), in a study of the Los Angeles Police Department, found that big data platforms allowed police to integrate previously separate databases and to track people who had no direct police contact. Ferguson (2017) argues that such tools offer efficiency but can extend surveillance to people never suspected of wrongdoing, and can embed bias in apparently neutral systems. Boyd and Crawford (2012) cautioned more generally that large datasets do not speak for themselves: they are shaped by how they were collected, and bigger data are not automatically more accurate or objective.

Courts have begun to set limits. In *Riley v. California* (2014), the United States Supreme Court held that police generally need a warrant to search a mobile phone seized during an arrest, recognising that phones hold "the privacies of life". In *Carpenter v. United States* (2018), it held that obtaining historical cell-site location records generally requires a warrant, because such records reveal a detailed picture of a person's movements over time. Kerr (2005) had earlier argued that rules designed for physical searches needed to be rethought for computers, and these decisions reflect that shift. The future will bring more such questions as new data sources, from cars to smart home devices, become available to investigators.

Open-source investigation

Some of the most significant investigations of the past decade were carried out using publicly available information: satellite images, social media posts, videos, flight and shipping data. Higgins (2021) describes how the open-source group Bellingcat used such material to identify those responsible for downing Malaysia Airlines Flight MH17 and for other incidents. International courts and human rights bodies now use open-source evidence, and the Berkeley Protocol sets out standards for collecting, preserving, verifying and analysing it so that it can be used in legal proceedings (Human Rights Center, University of California, Berkeley, School of Law & United Nations Office of the High Commissioner for Human Rights, 2022).

Open-source methods extend investigation beyond professionals to journalists, researchers and the public. That creates opportunity, but also risk: crowd-sourced investigations have misidentified innocent people, and publishing personal information can cause harm. The future of open-source investigation depends on applying the same standards of verification and ethics as any other form.

Synthetic media

Generative tools now produce realistic images, audio and video of events that never happened. Chesney and Citron (2019) describe two dangers for investigations and courts. First, fabricated material may be accepted as genuine. Second, genuine material may be dismissed as fake, a problem they call the "liar's dividend", because the mere possibility of fakes allows people to deny real evidence. Image forensics offers methods to detect manipulation by examining lighting, geometry, compression and other traces (Farid, 2016), but detection tools must keep pace with generation tools.

In future, investigators will need to establish the provenance of digital media more routinely: who created it, on what device, when, and how it has been handled since. Investigations that rely on a single image or recording without corroboration will become harder to defend.

Encryption, the cloud and borders

Encryption protects the data of billions of people, including victims of crime, but it can also put evidence out of investigators' reach. Governments have repeatedly proposed requiring companies to provide access to encrypted data. Abelson et al. (2015), a group of leading security experts, argued that such requirements would introduce vulnerabilities that criminals and hostile states could exploit, and that the risks outweigh the benefits. This debate is likely to continue.

Cloud storage means that evidence about events in one country is often held by a company in another. The Council of Europe's Convention on Cybercrime (2001) created a framework for international cooperation, and the United States' Clarifying Lawful Overseas Use of Data Act (2018) established a basis for agreements that allow foreign governments to request data directly from United States providers under certain conditions. Even so, obtaining cross-border data remains slow, and investigators will need to understand the legal routes available in each case.

Cryptocurrencies were once thought to make payments untraceable, but Meiklejohn et al. (2013) showed that analysis of the public Bitcoin ledger could cluster addresses and follow flows of funds to exchanges. Blockchain analysis is now a standard investigative technique, although privacy-focused currencies and mixing services make it harder.

Genetic genealogy

Investigative genetic genealogy compares DNA from a crime scene with profiles that people have uploaded to consumer genealogy databases, identifying distant relatives and then building family trees to narrow down a suspect. It led in 2018 to the identification of the "Golden State Killer" and has since resolved many cold cases. Erlich et al. (2018) estimated that a majority of Americans of European descent could already be identified through a third cousin or closer relative in such databases. The technique raises questions about the privacy of relatives who never consented to their DNA being searched, and many databases now require users to opt in to law enforcement matching.

Transparency of tools

As more investigative conclusions depend on software, the question of whether that software can be examined becomes more important. Wexler (2018) documents cases in which developers of forensic and risk assessment software have claimed trade secret protection to prevent defendants from examining how the software works. Courts and legislatures will need to decide how to balance commercial interests against the right to challenge evidence. Investigators can reduce the problem by preferring tools that have been independently validated and whose methods can be explained. They should also record the version and settings of every tool used, keep the raw data alongside the processed results, and be prepared to have a second examiner repeat key steps with a different tool. Where two independent tools give the same result, a court or tribunal can place more weight on it than on a single unexplained output.

What investigators will need

The trends point to several needs:

- **Data literacy.** Investigators will need to understand how data are created, what they can and cannot show, and how to test them, even when specialists do the technical work.
- **Validation habits.** Every tool output, from an AI summary to a location record, should be checked against an independent source before it supports a finding.
- **Provenance tracking.** Records of where each item came from and how it was handled will be essential as fakes become easier to produce.
- **Legal awareness.** Rules on privacy, surveillance, AI and cross-border data are changing quickly, and investigators will need to keep up with them.
- **Ethical judgement.** The ability to collect more data does not mean it should be collected. Proportionality, minimisation and respect for people not under suspicion will matter more, not less.
- **Collaboration.** Complex cases will require teams combining investigators, data analysts, forensic specialists and lawyers, often across organisations and countries.

A short example

The following hypothetical case shows how these trends combine. A company receives an anonymous complaint, with an audio clip attached, alleging that a senior manager offered a bribe to a supplier. The investigation team first examines the clip's provenance. Its metadata are missing, and an analysis of the audio finds signs that parts were generated synthetically, but the team does not dismiss the complaint on that basis. Instead it reviews the manager's email and messaging accounts, using AI tools to search tens of thousands of messages and transcribe call recordings, and checks every relevant result against the original data. It finds payments to a consultant linked to the supplier, traced through bank records and, in one case, a cryptocurrency exchange. Some of the manager's messages are held by a cloud provider in another country, and the team works with its lawyers to obtain them lawfully. The final report explains which evidence was relied on, how each item was verified, why the audio clip was given no weight, and what the AI tools did and did not do.

Conclusion

The future of investigation will be shaped by abundant data, artificial intelligence, open-source information, synthetic media, encryption, cross-border storage and new forms of forensic science such as genetic genealogy. Each development offers investigators new ways to establish what happened, and each creates new ways to get it wrong. The fundamentals will not change: investigators will still need to gather evidence lawfully, test it, consider alternative explanations and report honestly. What will change is the range of skills needed to apply those fundamentals, and the importance of safeguards that keep powerful tools subordinate to human judgement and the rights of the people affected.

References

Abelson, H., Anderson, R., Bellovin, S. M., Benaloh, J., Blaze, M., Diffie, W., Gilmore, J., Green, M., Landau, S., Neumann, P. G., Rivest, R. L., Schiller, J. I., Schneier, B., Specter, M. A., & Weitzner, D. J. (2015). Keys under doormats: Mandating insecurity by requiring government access to all data and communications. *Journal of Cybersecurity, 1*(1), 69–79. https://doi.org/10.1093/cybsec/tyv009

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review, 82*(5), 977–1008. https://doi.org/10.1177/0003122417725865

Carpenter v. United States, 585 U.S. 296 (2018).

Casey, E. (2011). *Digital evidence and computer crime: Forensic science, computers and the internet* (3rd ed.). Academic Press.

Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753–1820. https://doi.org/10.15779/Z38RV0D15J

Clarifying Lawful Overseas Use of Data Act, Pub. L. No. 115-141, div. V, 132 Stat. 1213 (2018).

Convention on Cybercrime, Nov. 23, 2001, E.T.S. No. 185 (Council of Europe).

Erlich, Y., Shor, T., Pe'er, I., & Carmi, S. (2018). Identity inference of genomic data using long-range familial searches. *Science, 362*(6415), 690–694. https://doi.org/10.1126/science.aau4832

Farid, H. (2016). *Photo forensics*. MIT Press.

Ferguson, A. G. (2017). *The rise of big data policing: Surveillance, race, and the future of law enforcement*. NYU Press.

Garfinkel, S. L. (2010). Digital forensics research: The next 10 years. *Digital Investigation, 7*(Suppl.), S64–S73. https://doi.org/10.1016/j.diin.2010.05.009

Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face recognition vendor test (FRVT) part 3: Demographic effects* (NISTIR 8280). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8280

Higgins, E. (2021). *We are Bellingcat: An intelligence agency for the people*. Bloomsbury.

Human Rights Center, University of California, Berkeley, School of Law, & United Nations Office of the High Commissioner for Human Rights. (2022). *Berkeley protocol on digital open source investigations: A practical guide on the effective use of digital open source information in investigating violations of international criminal, human rights and humanitarian law*. United Nations.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248. https://doi.org/10.1145/3571730

Kerr, O. S. (2005). Searches and seizures in a digital world. *Harvard Law Review, 119*(2), 531–585.

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance, 13*(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M., & Savage, S. (2013). A fistful of bitcoins: Characterizing payments among men with no names. In *Proceedings of the 2013 Internet Measurement Conference* (pp. 127–140). ACM. https://doi.org/10.1145/2504730.2504747

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Quick, D., & Choo, K.-K. R. (2014). Impacts of increasing volume of digital forensic data: A survey and future research challenges. *Digital Investigation, 11*(4), 273–294. https://doi.org/10.1016/j.diin.2014.09.002

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). (2024). *Official Journal of the European Union, L*.

Riley v. California, 573 U.S. 373 (2014).

Wexler, R. (2018). Life, liberty, and trade secrets: Intellectual property in the criminal justice system. *Stanford Law Review, 70*(5), 1343–1429.