Detecting Elder Fraud Without Surveilling Elders:

Legal and Criminological Conditions for Lawful, Privacy-Preserving AI Detection of Phone-Based Scams

Carl Lochstampfor

Department of Interdisciplinary Studies, Old Dominion University

IDS 300W: Interdisciplinary Theory and Concepts

August 21, 2026

Detecting Elder Fraud Without Surveilling Elders: Legal and Criminological Conditions for Lawful, Privacy-Preserving AI Detection of Phone-Based Scams

Phone-based scams against older adults are a complex problem that no single field can solve alone. Addressing this issue requires input from criminology, law, and computer science. Criminology helps explain why older adults are targeted and how scams unfold. Law sets the rules for when and how private conversations can be monitored. Computer science examines whether detection systems can work without violating privacy. These three fields are the best ones to choose from because the research question splits into three parts, each belonging to one of them: (1) why older adults are targeted and where to intervene is criminological; (2) whether monitoring is legally allowed is a question of law; and (3) whether a system can catch scams without exposing private data is a computer science problem. Leaving any one of those fields out of the equation would leave part of the question unanswered, which is exactly why an interdisciplinary approach is required (Repko & Szostak, 2026). When criminology, law, and computer science are combined with recent advances in artificial intelligence (AI), especially large language models (LLMs), it becomes possible to detect scams in real time by analyzing live conversations and spotting manipulative tactics. Still, just because something is technically possible does not mean it is legally or ethically acceptable. A system meant to protect could also be used for surveillance, depending on the legal and criminological context.

This paper asks: *under what legal and criminological conditions can AI-based detection of phone scams against older adults be used lawfully and privately in the United States?* The question requires an interdisciplinary approach, since looking at the problem from only one field can lead to incomplete or misleading answers. For instance, a technical solution might focus on accuracy but ignore legal limits on intercepting calls. A legal view might see consent as just a formality, while a criminological view might find the right time to intervene but overlook privacy or legality.

Using Repko and Szostak's (2026) interdisciplinary method, this paper brings together ideas from criminology, law, and computer science, highlights the main conflicts between them, and argues that system design, not just policy, is key to making AI scam detection both effective and lawful. Lastly, the focus of this paper is on scams targeting older adults in the U.S., especially during live phone conversations.

The Scale and Structure of Elder Fraud

Older adults are targeted by online and phone scams more than any other age group. According to the Federal Bureau of Investigation (2026), which compiles and analyzes fraud complaints reported to its Internet Crime Complaint Center, people aged 60 and older lost over $7.7 billion in 2025, with more than 201,000 victims and an average loss of over $38,000. Understanding this problem requires a criminological perspective. Lam et al. (2026) used Cyber Routine Activity Theory in a study with 23 focus groups and 142 adults aged 55 or older. They found that scams happen when a motivated offender, a suitable target, and no capable guardian come together. The most common scams involved pretending to be a relative, an authority figure, or using "guess who I am" tricks. These scams exploit older adults' regular smartphone use and their expectations in relationships, turning everyday actions into risks (Lam et al., 2026). Here, vulnerability comes not just from cognitive decline but from the way digital habits and trust overlap.

The way these crimes are structured shows why computer science solutions are needed. Leclerc and Morgenthaler (2023) reviewed different types of online fraud and found that most scams follow three common steps: the scammer contacts the victim, recruits other people to help carry out the fraud, and then uses money mules to move the stolen money. The first step, talking/communicating to the victim, is where AI detection can help most, since this is when scams

can be immediately spotted in the conversation. Seeing the problem this way means the AI detector acts as a technical guardian during the communication stage. Lam et al. (2026) also found that having a guardian makes people less likely to be targeted, while poor technical protection increases risk. However, relying on family to help "does not always translate into improved technical protection" (Lam et al., 2026). The lack of real-time technical guardianship, even with helpful family members, is the gap that AI detectors aim to fill. Filling that guardianship gap is the criminological reason for building such systems.

It is important to note that this gap does not mean older adults are helpless. Lam et al. (2026) found that older adults with better computer skills often helped their peers and used active strategies to cope, not just avoidance; these results challenge negative stereotypes about older adults. While the study took place in Hong Kong, it suggests that building community and peer support for technology skills can help close technology gaps (i.e., bridging the digital divide) and reduce the fear that keeps some older adults in the U.S. from using protective tools. This view matters for system design, as it sees older adults as active partners in their own protection, not just passive subjects of surveillance. This is a key idea for the opt-in, participatory systems discussed later.

Consent as the Master Variable

While criminology explains why a detector is needed and what it should do, legal rules decide if monitoring is allowed. Any system that listens to live phone calls must follow laws about wiretapping and interception, with consent being the main legal requirement. But some legal experts question whether consent, as it is usually given, is truly meaningful. Rothchild (2018) argues that the notice-and-choice approach—where people are considered to have agreed just by continuing after a disclosure—is "fundamentally flawed, cannot be fixed, and should be replaced"

with real limits on collecting and using personal information. Rothchild's argument applies to everyone, not just older adults: no one can realistically read and evaluate every privacy notice they come across, and because most privacy terms are nearly identical, there is rarely a more protective option to choose from. And for older adults, this problem impacts them even more. Thus, the problem lies in the system, not in the person (Rothchild, 2018).

This legal criticism matches what criminology finds and affects how systems should be built. Lam et al. (2026) found that some older adults hung up phone calls out of fear or had trouble managing passwords, showing the gap between formal consent and real control, as Rothchild (2018) describes. When it comes to AI scam detection, consent should not just be a box to check, since legal experts say this is not enough. Rothchild's main point for system design is that notice-and-choice fails when a company sends personal data to third parties for reasons it never disclosed. When someone clicks ‘agree’ on a privacy notice, they cannot consent to uses no one told them about, so that click does not actually cover what the company later does with their data. This violates fair information practices because a practice cannot be fair when the person never had a real way to know about it or to agree to it. This problem also sits at the heart of computer science, because where a system handles a conversation determines everything (i.e., on your own phone or on a company's server). This means that *how* the system is built, not just data policies, decides if it meets legal standards.

The Primary Conflict: Deterrence Versus Legitimacy

These fields not only focus on different things but also disagree about what an AI detection system should do. In criminology, the deterrence view says crime drops when people fear being caught and punished, so it supports more surveillance and prosecution. On the other hand, the procedural and legitimacy views argue that people follow rules more because they see the process

as fair, not just because they fear punishment. Tyler et al. (2015) found that perceived legitimacy affects behavior "as well as or better than" the risk of punishment, and that heavy surveillance and punishment have "mixed effects" and can be costly. This matters for how an AI detector is designed: deterrence-based systems use centralized detectors that keep conversation data for prosecution, while legitimacy-based systems prefer opt-in detectors that protect privacy and involve older adults as active participants.

Research supports the legitimacy approach, which has direct effects on how systems should be built. In the Queensland Speeding Engagement Trial, Bates et al. (2025) found that drivers aged 25 and older who got a fair and respectful letter along with a standard ticket were 11% less likely to get another speeding fine within a year, even though there were no extra threats of punishment. This effect was due to the sense of fairness, not stricter enforcement (Bates et al., 2025). While this study looked at traffic offenses, its lessons apply here to this paper's research question: fair, open, and voluntary interventions work better in the long run. So, an AI detector that is based on legitimacy and user opt-in is better than one that relies on secret surveillance for prosecution. In the end, this criminological debate becomes a computer science question about how to design the system. In this context, a legitimacy-based approach means the detector treats the older adult as someone to be respected and kept informed, not watched. The person willfully and consciously chooses to turn it on, understands what it does, and stays in control of it. The aim is to earn cooperation through fairness and transparency, which research links to lasting behavior change, rather than to impose monitoring the person never agreed to.

Common Ground: Architecture Resolves the Policy Debate

A key insight is that system design can help solve the policy debate about surveillance. Rothchild (2018) points out that privacy problems happen when personal data is sent to third parties for unknown reasons. This uses what Repko and Szostak (2026) call the Technique of Redefinition: the conflict between deterrence and legitimacy from the previous section is redefined from a policy argument into a design decision. Once the question shifts from 'which policy?' to 'how is the system built?', the disciplines stop conflicting and share common ground. Tyler et al. (2015) and Bates et al. (2025) say that protective goals are better met without collecting lots of data. Lam et al. (2026) note the lack of real-time technical guardianship for older adults. One design choice addresses all these issues: process conversations locally on the user's device instead of sending data elsewhere. Wang et al. (2025) show this is possible with MASK, a tool that removes sensitive information from phone transcripts before further analysis. This lets large language models detect scams while reducing privacy risks. Wang et al. show that computer science can solve legal problems, since cleaning data at the source avoids the third-party sharing that Rothchild says is the main flaw in the notice-and-choice model.

Hossain et al. (2026) use similar ideas for building and improving these systems. Their AI-in-the-loop setup detects scams in real time with a small local model and uses federated learning. This method updates a shared model across many users' devices without sending raw conversation data off the device. To protect privacy even more, they use differential privacy, which adds statistical noise so individual data cannot be reconstructed. In plain terms, federated learning works like a class of students who each study at home and share only their improved study tips with the group, never their private and personal notes, so the shared model gets smarter without anyone

handing over their actual data. Differential privacy then adds a little random ‘static’ to those shared tips, so no one can work backward to figure out any single person's information. The authors found that federated training worked well (about 0.80 engagement) and leaked very little personal information, showing that strong privacy can go ‘hand-in-hand’ with good performance (Hossain et al., 2026). Together, the work of Wang et al. (2025) and Hossain et al. (2026) proves that an on-device, consent-based system is not just an idea but can actually work. This means that what is possible in computer science meets the needs or requirements in law and criminology.

An Integrated Framework for Lawful, Private Detection

This section carries out the integration step of the interdisciplinary process, combining the insights from all three disciplines into a single framework. Bringing these ideas together leads to a set of rules for using AI to detect phone scams against older adults in the U.S. First, the system must be **opt-in**, started either by the older adult or a trusted person with clear authority. This is based on legal research showing that passive consent is not meaningful or helpful long term (Rothchild, 2018) and based on criminological evidence that protective actions work best when they are fair and involve the user (Tyler et al., 2015; Bates et al., 2025). Importantly, opt-in is about legitimacy and proper authorization, not just privacy. Privacy is protected by the technical safeguards described in the next points, following Rothchild’s (2018) call for real substantive limits instead of just procedural consent.

Second, all analysis must happen **on the user’s device**, so conversations are never sent to third parties. This directly solves the problem of sharing data with others, as Rothchild (2018) points out, and eases concerns about recording or monitoring private calls without clear consent, since no data leaves the device (Wang et al., 2025). Third, the system should clean any information it must process right at the source, a step called source-level redaction. Before any analysis occurs,

the device itself strips out personally identifiable information (PII) while keeping enough of the conversation's structure for detection to still work (Wang et al., 2025).

Fourth, the system should improve itself using privacy-preserving federated updates with differential privacy. In practice, each person's device learns from the scams it encounters and sends only those lessons, never the conversations themselves, to a shared model, with random noise added so no individual's data can be traced back. This lets the system keep up with changing scam tactics, as Leclerc and Morgenthaler (2023) note, without collecting all user data in one place (Hossain et al., 2026). Together, these rules create a detector that acts as a technical guardian during scam calls (Leclerc & Morgenthaler, 2023; Lam et al., 2026), and meets legal standards for real, design-based privacy protection instead of just relying on procedural consent, which Rothchild (2018) criticizes. This approach is truly interdisciplinary: it only works if all the relevant fields are included, and leaving any out weakens its legal, technical, or protective value.

Conclusion and Limitations

Whether AI systems for detecting phone scams against older adults are lawful and private depends more on how they are built than on laws alone. If detection is opt-in, runs locally on the device, uses source-level redaction, and updates through privacy-preserving federated learning, legal issues about interception and consent are solved by simple, architectural design. This also matches criminological findings that legitimacy-based approaches work better than coercive ones. While the debate between deterrence and legitimacy is important, it ultimately comes down to how the system is designed and built, which is a computer science issue.

There are some limits to these conclusions. The evidence for the legitimacy argument from Bates et al. (2025) comes from traffic offenses, not fraud, so its use for scam detection is based on analogy and needs direct proof. Lam et al. (2026) did their study in Hong Kong, and cultural

differences affect both victimization and guardianship, so applying the findings to the U.S. is only tentative. While the framework solves interception and consent issues through design, it does not cover important governance details like data retention, breach response, redress, or independent oversight, which still need to be developed. Also, while the framework sets out conditions for lawful use, it does not address complex questions about liability or responsibility if an opt-in detector fails. These issues need more research. The final step of the interdisciplinary process is to communicate and test the framework. Testing it would mean building a detector for these conditions and measuring, in real use, whether it protects older adults without compromising their privacy.

References

Bates, L., Bennett, S., Irvine, C., Antrobus, E., & Gilmour, J. (2025). A procedurally just flyer reduces subsequent speeding offences: Evidence from the Queensland Speeding Engagement Trial (QSET). *Journal of Experimental Criminology, 21*(1), 201–217. https://doi.org/10.1007/s11292-023-09582-w

Federal Bureau of Investigation. (2026). *2025 Internet crime report.* Internet Crime Complaint Center. https://www.ic3.gov/AnnualReport/Reports/2025_IC3Report.pdf

Hossain, I., Puppala, S., Alam, M. J., & Talukder, S. (2026). AI-in-the-loop: Privacy preserving real-time scam detection and conversational scambaiting by leveraging LLMs and federated learning. *Proceedings on Privacy Enhancing Technologies, 2026*(1), 87–114. https://doi.org/10.56553/popets-2026-0006

Lam, G., Lau, B. H.-P., Tsang, C.-K., Kwok, A. P.-K., & Shum, E. N.-Y. (2026). The use of IT-safety and coping measures against cybercrimes among older adults in Hong Kong: An application of cyber routine activity theory. *Frontiers in Sociology, 11*, 1739787. https://doi.org/10.3389/fsoc.2026.1739787

Leclerc, B., & Morgenthaler, E. (2023). *Examining emerging fraud facilitated by the internet through crime scripts* (Trends & Issues in Crime and Criminal Justice No. 680). Australian Institute of Criminology. https://doi.org/10.52922/ti77208

Repko, A. F., & Szostak, R. (2026). *Interdisciplinary research: Process and theory* (5th ed.). SAGE Publications.

Rothchild, J. A. (2018). Against notice and choice: The manifest failure of the proceduralist paradigm to protect privacy online (or anywhere else). *Cleveland State Law Review, 66*(3), 559–648. https://engagedscholarship.csuohio.edu/clevstlrev/vol66/iss3/7/

Tyler, T. R., Goff, P. A., & MacCoun, R. J. (2015). The impact of psychological science on policing in the United States: Procedural justice, legitimacy, and effective law enforcement. *Psychological Science in the Public Interest, 16*(3), 75–109. https://doi.org/10.1177/1529100615617791

Wang, K., Shen, Z., Zhang, Y., Cheung, M. M. K., Luo, X., Ngai, G., & Fu, E. Y. (2025). One size fits all? A modular adaptive sanitization kit (MASK) for customizable privacy-preserving phone scam detection. In *Proceedings of the 33rd ACM International Conference on Multimedia* (pp. 12381–12389). Association for Computing Machinery. https://doi.org/10.1145/3746027.3758164