FEATURE FOCUS

Defending against Invisible Risk: The Implicative Data Lens for Security Governance Professionals

By: Emmi Bane and Carl Mathis

Abstract
As organizations increasingly rely on advanced analytics and AI-driven systems, security programs face growing risk from data that fall outside traditional classifications of "personal" or "sensitive." Data that do not directly identify individuals can nonetheless generate meaningful inferences, reveal patterns, or influence outcomes when analyzed in context or combined with other data. These "implicative data" are frequently overlooked by security governance programs that rely on static taxonomies and element-level classification.
This article introduces the Implicative Data Lens as a practical risk analysis framework for security professionals. The lens reframes security assessment away from static data categories and toward relational, inferential, and contextual risk. Drawing on prior research and applied security analysis, the paper illustrates how implicative data expand breach impact, enable inference-based attacks, contribute to policy drift, and amplify third-party and supply-chain risk.

Keywords:
security governance, risk management, inference attacks, data classification, AI security, privacy engineering, data governance

Introduction: Problem Framing & Relevance

AI models analyze and synthesize data at a heretofore unprecedented scale, which poses data governance challenges at every stage for an organization. From the evaluation and configuration of third party tools to compliance with data protection laws and security standards, the potentials for misuse are increasing. These potentials are complicated by implicative data, which are data that may not directly identify an individual, but have consequences for individuals, groups, or organizations due to inferences, contextual associations, profiling, or aggregated patterns resulting from linkability, singling out, or inference [1].

Governance programs often rely on data taxonomies and the classifications of individual elements to calculate risk and determine what data elements require protections. This can overlook a lot of data, or data combinations, that can represent risk to the organization. Derivations, inferences, aggregations and metadata can all have an impact on individual outcomes. Synthesized data can be connected to an individual's data constellation, becoming personal data over which an organization is obligated.

The considerations raised by implicative data are especially relevant to security. Data do not necessarily need to relate to a person to be implicative – non-personal internal organizational data can reveal patterns about an organization's finances, products, and internal processes, and the scope of what data can make that possible is rapidly increasing. Data considered nominally benign by organizations can be recombined with external data sets or analyzed in unanticipated contexts. This increases the scope and impact of breaches and attacks, and expands the risks of future harms [2].

Implicative data, and a failure to recognize it and adequately protect it, can put organizations at risk of data breaches, contract violations, and create previously undetected security vulnerabilities. This article will describe the concept of implicative data for security professionals, including some key security considerations, and discuss how adopting an approach inclusive of implicative data can help to mitigate these concerns and contribute to a more future proof security program.

Industry Context and Prior Work

In a regulatory context, several definitions of personal data exist. While the patchwork nature of data protection laws can create some dissonance for entities that operate globally, most regulatory frameworks are based on the definitions provided by the General Data Protection Regulation (GDPR) and the California Privacy Rights Act (CPRA):

Under the GDPR: "'Personal Data' means any information relating to an identified or identifiable natural person ('data subject');, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person [3]."

Under the CPRA, "Personal Data" is defined as: "information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household [4]."

Both definitions refer to "ANY" information that implies personal attributes or behaviors without explicit or even indirect identification, but which is linkable or inferred. However, most data governance programs rely on data taxonomies, which can create a false dichotomy between "personal" and "non-personal" data, especially given the potential of advancing data analysis capabilities. This approach also often overlooks the combinatorial risks of data, focusing instead on elements individually.

Applying the Implicative Data Lens closes the expanding gap between the programs designed to meet the obligations of organizations that process data, and the capabilities of data analysis augmented by emerging technologies.

Prior Work:

In our recent work for the IEEE Symposium on Privacy Expectations (ISOPE), Dr. Mathis and I explore the concept of implicative data. We define implicative data as:

"…Data that may not be directly or uniquely identifiable in and of themselves, but have relevance to constellations of data by supplying insights, derivations, profiles, or other data enrichment that has implications for persons, organizations or groups of people [1]."

The work of Helen Nissenbaum, specifically that on Contextual Privacy, has contributed significantly both to the formulation of the Implicative Data Lens, and the operationalization of the concept. Her work helped establish a framework in which the sensitivity of data is dependent on context [5].

The work of Dr. Latanya Sweeny has also been extremely influential in this conceptual framework, most notably her work on health data reidentification. In this work, it was discovered to be possible to reidentify individuals in a dataset that had been deidentified to beyond a HIPAA standard [6].

Risk Analysis Lens and Methodology

Implicative data is a level of abstraction above traditional data identifiers such as indirect or quasi-identifiers, derived data, or inference. A distinguishing factor of Implicative data is that it is defined not just by its relational connections to a primary data element or profile, but by its relational potential for influencing outcomes.

Implicative data can include otherwise benign data that when combined with external data can lead to conclusions, attribution or reidentification. In order to address these risks, organizations must shift the scope of their security and privacy assessments beyond traditional data categories to include data that have implicative potentials.

Shifting the scope of your evaluations has two basic steps. First and foremost, the most important thing an organization can do to shift the scope of risk to adequately encompass implicative data is to stop asking: Does this solution process personal data? And to start asking: "Does this solution process data at an individual level?" To practically implement this shift in scope, add "Individual Level Data" as a classifier to your existing taxonomy.

This allows security and privacy professionals to assess the datasets without relying on increasingly meaningless dichotomies such as "personal" and "non-personal" data, and to overlay the implicative data lens across an existing governance program.

Secondly, evaluate the context of the data processing. Context in this case refers to the specific social, cultural, or situational setting in which data are processed. (Nissenbaum)

The context of collection helps to set the legal basis for usage of the data, and the context of the analysis can suggest what types of risks an organization may be incurring. Knowing what the intended outcomes are can provide insight into what types of incidental information might be contained in or generated by the processing of these data sets.

Sample context questions:

- What is the sensitivity and type of the dataset?
- What are we trying to learn?
- What is the reidentification risk of the dataset?
- What techniques have been employed?
- Is the information intended to combine with other information? If so, what?
- Are the data being relationally connected and what is the strength and type of the connection?

Assessing the risks with additional layers of context allows organizations to more adequately determine the likelihood of each risk, and to ensure that the appropriate protections are applied to data that may represent potential harms to individuals, organizations or groups.

Moreover, this helps to limit the scope of the assessment. Since implicative data is context dependent, all data have the potential to be implicative in the correct contexts. The important thing is to identify the most reasonably likely risks to avoid untenable and unreasonable scope expansion.

Context operationalization is a key aspect of the Implicative Data Lens. A limitation of this lens is the potential for everything to represent implicative data, making the scope of potentially relevant data impossible to practically govern. To remediate this, the application of context helps us to better assess the likelihood of risks.

Technical operationalization refers to the process of defining privacy context in terms of a specific data model with associated technical procedures or systems that allow for consistent measurement and analysis at scale.

Information operationalization is the act of defining what "information" means in a specific context and determining how it can be communicated and exchanged, understood, and Interpreted. This underpins the establishment of norms and expectations.

One context for data use is represented by the expectations of the data subject. We can help set these expectations by providing data subjects with notice language that clearly describes how the data will be used and the protections that will be applied to the data. We can help align expectations by assessing how the purpose for continued use can be mapped back to the original basis for collection. Even if data have been deidentified, pseudonymized or abstracted, we can still ensure that the data are being used in a manner consistent with the original purpose.

Most risks from implicative data remain undetected because they don't represent a recognized risk or link to identifiable persons. However, deidentified and even anonymized data can be associated with an individual with enough context. The implicative data lens allows organizations to better assess and control for actual risks stemming from data generated by AI processing, and to expand their governance funnel practically to include the use cases that represent risk beyond the traditional triggers of "personal" or "sensitive" data. Being able to programmatically evaluate the practical and contextual risk of data use helps organizations innovate faster within a future-proof governance structure.

Comparison: Risks and Why They Go Undetected

Risk Type	Why It's Missed	Impact
Metadata	It's viewable as "administritive" rather than "content."	Reveals internal hierarchies and software versions.
Timing/Latency	Most monitors only check *if* a site works, not how *fast* it responds.	Reveals if a specific user exists in a database.
Telemetry	Considered "technical health data" for the manufacturer.	Maps out a user's physical habits and locations.

Applied Security Insights

Implicative data creates essential considerations for security professionals. As the threat landscape evolves, it is imperative that we test against the right threats. This means thinking about not just the known applications of a given dataset, but about how that dataset might be used or recombined in the future. An understanding of the value of implicative data can help identify emerging risks and, in many cases, prevent them. The following are some examples of how implicative data can influence the security landscape.

- **Expanded Blast Radius in Breach Scenarios**

Even when breached data is nominally "non-personal," implicative data increases the potential for downstream harms to individuals, since compromised data can be used in unexpected and unauthorized ways, including the recombination with external data sets. Depending on how the breached data are ingested or proliferated, secondary harms could emerge long after the incident and become increasingly difficult to track. This is especially common in health care breach scenarios, where even non-personal data tend to be highly specific and inferential.

When planning for breach responses and remediations, Incident response plans and breach classifications must also consider the future inferential harm of the data, not just the impact of the immediate exposure.

- **Inference Attacks as a Primary Security Vector**

Implicative data contributes materially to inference attacks, in which adversaries derive sensitive attributes without direct access to protected data. Using inferences as the basis of an attack vector means attackers don't need to infiltrate high security systems in order to gain access to high sensitivity data. Incidental or low-sensitivity datasets can generate highly sensitive inferences, contributing to higher impact outcomes. Traditional perimeter and access-control models may fail to capture risk when harm arises from cross-context inference rather than data exfiltration.

To address this, threat models must expand beyond unauthorized access and treat inference capability itself as an attack surface.

- **Policy Drift and Security Misconfiguration**

Because implicative data sits outside traditional data classifications, it is often overlooked in applying critical protections and can become a security vulnerability. By misclassifying or failing to classify implicative data, organizations may fail to meet contractual, sectoral or legal commitments such as encryption requirements, logging and anomaly detection, or secure data lifecycle controls.

- **Third-Party and Supply-Chain Amplification**

The risks inherent to implicative data in an organization, such as miscategorization, reidentification, data attribution, inadvertent data creation and sharing are compounded by complications to the data system such as third-party sharing and model training. In these cases, organizations lose visibility into how implicative data is recombined downstream. This increases the risk of future adverse events such as spurious or harmful data generation or disclosure. This can be especially dangerous when models surface or re-surface implicative insights in unanticipated contexts.

AI is already in your environment. Who's responsible when it leaks your data?

Most security teams didn't choose to become responsible for artificial intelligence (AI) risk. It happened anyway.

A business unit deployed an AI-powered tool to automate customer interactions. A vendor quietly added AI-driven analytics to a platform your organization already uses. Someone on the finance team started using a large language model (LLM) to summarize internal documents. None of it went through a formal security review or made it onto the risk register.

And when something goes wrong (a data leak, a manipulated output, a compliance violation) the question lands on your desk. Not the business unit's. Not the vendor's. Yours.

The governance gap nobody planned for

The challenge isn't that security professionals aren't capable of governing AI risk. It's that the credentials they were trained on weren't built for it.

Consider what Certified Information Systems Security Professional (CISSP) and Certified Information Security Manager (CISM) were designed to address:

- Systems that behave consistently unless deliberately changed
- An attack surface defined by infrastructure and applications
- Risk that can be identified, measured, and controlled through established frameworks

Now consider what AI systems actually introduce:

- Behavior that shifts based on new data, user interaction, or environmental drift
- Prompt injection attacks that manipulate outputs by embedding hidden instructions in content the system processes
- Model poisoning that corrupts training data to produce systematically wrong results over time
- Data leakage through pathways that conventional controls never anticipated

These are documented attack vectors observed in production environments. And according to ISACA's own research, 60% of security professionals are very or extremely worried that generative AI will be exploited by bad actors, yet only 35% say AI risks are an immediate priority for their organization to address.

Most security programs are running behind. The question is how far.

Why this matters now

AI adoption is accelerating, and it is not becoming less risky as it scales. Every new deployment, every vendor integration, every employee tool expands the attack surface your team is responsible for, whether or not they have the training to assess it. Organizations that don't close this gap soon won't get to choose when the consequences arrive.

Governing AI risk means understanding the full lifecycle of an AI system: how data is sourced and protected, how models are trained and validated, how outputs are monitored for drift or manipulation, and how third-party AI components introduce risk that can't be directly inspected. It also means translating an emerging regulatory landscape, including the EU AI Act and sector-specific guidance in financial services, healthcare, and critical infrastructure, into operational controls security teams can actually implement.

A credential built for this moment

ISACA's Advanced in AI Security Management (AAISM) is the first and only AI-centric security management certification. Designed for active CISM or CISSP holders, it covers AI governance and program management, AI risk management, and AI technologies and controls, from the perspective of the security manager accountable for those decisions, not the engineer building the systems.
Destination Certification offers AAISM preparation through a three-day live Bootcamp and a self-paced MasterClass. Our training is developed by a team that includes John Berti, co-author of the first official ISC2 CISSP study guide and contributor to ISC2's official curriculum.

Our MasterClass is built around a study system that adapts to what you already know, so you're not covering ground you've already mastered. Across our programs, 93.6% of MasterClass students pass on their first attempt, one of the highest first-attempt pass rates in the industry. Nearly all of those who need a second attempt pass on the next try.

ISSA members receive 10% off either the MasterClass or the Bootcamp.

Or if you want to get CISSP certified, you get the same 10% off our CISSP MasterClass and Bootcamp.

Security professionals should ensure that all third party risk assessments include language restricting or preventing inferential reuse and model training.

- **Algorithmic Bias and "Proxy" Discrimination**

Effective data governance is about the integrity of systems and the data processed therein. Biased data or outcomes that perpetuate harms or inequities speak to the character and reputation of the organization, in addition to creating compliance and civil liberties risks. Proxy data or aggregate datasets can be implicative when they contribute to bias in outcomes. This is especially risky when these data are a surrogate for protected characteristics like race, religion, or sexual orientation.
Data Governance and Data Science professionals have an obligation to review systems thoroughly and repeatedly for evidence of bias or degradation. Since all data have the potential to become implicative under the right circumstances, good data hygiene and quality can help prevent spurious conclusions that might affect individual outcomes.

Practical Implications for Security Practice

There are growing practical instances and implications of the implicative data lens for security professionals. The following are some examples of how implicative data can be used to subvert security protections.

- **Social Engineering via "Contextual Legitimacy"**

The Attack: Attackers don't need your password if they can imply they belong in your inner circle. By harvesting "boring" implicative data, like hobbies, routines, preferred brands or vocabulary, an attacker can build a persona that looks plausibly similar to the legitimate ones.

The Risk: This creates **High-Fidelity Phishing**. Emails from these profiles aren't designed to be blandly generic or unassuming. They are designed to be perfectly timed messages about a delay in your specific coffee shipment, or a "memo" from a department that only someone with your badge-in habits would find credible. Information that may seem completely benign can be used to create or enrich realistic profiles that represent a threat.

- **Power Consumption & Heat Signatures (Side-Channels)**

The Attack: The physical "exhaust" of hardware is implicative data. In high-stakes environments, this information can be used to generate inferences about what type of operations are being performed, when, and where.

The Risk: An attacker with proximity to a device (or access to a smart thermostat in a server room) can monitor power fluctuations or fan speeds. These fluctuations can imply the type of cryptographic operations being performed.

Most companies secure their data packets but forget that the heat generated by the processor is also a data stream that can leak encryption keys and other sensitive inferential data.

- **The "Negative Space" Leak**

The Attack: Sometimes, the absence of data is the most telling implicative signal. Examining the shape of missing data can yield insights into a person's health, employment or daily routines that can in turn suggest insights about a person's behavior. For example, If a high-level executive's fitness tracker suddenly stops syncing data at 2:00 PM every Tuesday, it may suggest a recurring, private commitment (like a medical appointment or a secret board meeting).

The Risk: Security tools look for additions to databases or outbound traffic. They rarely flag a "stop" in data flow as a vulnerability, yet it can signal a predictable window for a physical or digital breach, or create a basis for insinuations.

Limitations and Future Considerations

The Implicative Data Lens is intentionally expansive, which introduces several practical limitations. Most notably, because implicative data is context dependent, nearly any data can become implicative under the right conditions. Without careful scoping, this creates a risk of untenable assessment breadth and governance fatigue. To remain operationally viable, organizations must therefore focus on reasonably likely risks, rather than attempting to account for all possible future implications.

The lens also relies heavily on an organization's ability to define, interpret, and operationalize context consistently. While context is essential for distinguishing meaningful risk from theoretical possibility, it is not always stable or uniformly understood across teams, systems, or time. This dependence limits consistency and repeatability, particularly in large or distributed environments.

In its current form, the Implicative Data Lens is primarily qualitative. It does not yet provide standardized metrics for measuring relational strength, inference likelihood, or downstream impact. This constrains its integration into automated tooling, comparative risk scoring, or large-scale monitoring programs. This represents much of the future research in this area.

Finally, as the threat landscape and data analytics capabilities continue to evolve, new forms of implicative risk will emerge that are not fully captured by current models. Further interdisciplinary research is needed to develop quantitative methods, security-native mappings, and shared industry practices that can mature the lens and support broader industry adoption.

Conclusion

Implicative data reframes security analysis away from static data classifications and encourages risk evaluation based on relational, inferred and reasonable contextual risk by switching from a data classification based model to one based on context. This helps to reduce organizational blind spots and expand the scope of organizational governance to include data that can represent practical threats.

The lens deliberately expands the scope of governance to include data that traditional programs overlook. This does not mean that all data should be governed equally. Context, likelihood, and intended outcomes act as practical constraints that make implicative risk assessable without overwhelming security teams. The lens is intended to be an overlay to an existing data governance program, not a replacement. The value of the lens lies in helping practitioners see what matters, not everything that is theoretically possible.

As analytics capabilities and AI-driven synthesis evolves, implicative risks will continue to emerge and outpace our both ability to classify them and the speed at which we update our data taxonomies. Security programs that fail to account for AI/ML capabilities, relational data use, and downstream recombination will increasingly misjudge risks and threats, both internally and internally. The Implicative Data Lens provides a foundation for evolving security governance to meet this reality, even as quantitative methods and tooling mature.

References

1. Bane, E. and Mathis, C. "Beyond Personal Data: Addressing Risks and Governance of Implicative Data in Modern Data Processing" (In press, 2026) Proceedings of the 2025 IEEE
2. Yuan, Alda, "Derived Data: A novel privacy concern in the age of advanced biotechnology and genome sequencing" Yale Law and Policy Review Inter Alia 37.1 2018
3. Regulation 2016/679 (General Data Protection Regulation or GDPR) (2016).
4. California Government Code § 7920.000 (2023).
5. Nissenbaum, Helen Privacy in Context: Technology, Policy, and the Integrity of Social Life (2010) Stanford Law Books, ISBN 978-0-8047-5237-4
6. Sweeney L, Yoo JS, Perovich L, Boronow KE, Brown P, Brody JG. Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study. Technol Sci. 2017;2017:2017082801. Epub 2017 Aug 28. PMID: 30687852; PMCID: PMC6344041.
7. Amplitude: What is Data Taxonomy? Examples Included | Amplitude 2026
8. Information Commissioners Office, "Personal Information- What is it?" [Online]. Available:https://ico.org.uk/for-organisations/

About the Authors

Emmi Bane is a Trust and Privacy professional at HP Inc., working at the intersection of data governance, security risk, and emerging analytics. Her research focuses on implicative data, AI-ML based risk, and the limitations of traditional data classification models in modern security programs. Her work focuses on the public health aspects of data collection and application.

Carl Mathis is a Trust and Privacy professional at HP Inc., with expertise in data governance, AI-ML based risk, and security implications of advanced data analysis. His work explores how relational and implicative data challenge traditional security and privacy models.