Privacy in the age of cheap intelligence

How data fusion changes personal security

Owl + Kestrel | 3 November 2026

Executive assessment

Meaningful privacy remains achievable. Preserving it can require much more effort, technical knowledge and coordination than ordinary people expect. A person can carefully avoid stating a sensitive fact and still publish enough evidence for someone else to infer it.

The relevant evidence is the combined record: photographs, recordings, writing, transactions, relationships and observations made by other people. An adversary can connect those records, use one to interpret another and revisit old material when new information becomes available. Judging every clue by whether it independently identifies an arbitrary person substantially understates this capability.

Five conclusions follow.

- **An ordinary post can be a measurement.** Its incidental physical and behavioral details may reveal more than its intended subject. Human inspection of a photograph or recording does not establish what a machine can extract from it.
- **Combining clues can change the answer.** Observations can narrow different dimensions of uncertainty, resolve ambiguities or supply a missing reference. Several individually inconclusive clues can jointly support a consequential inference.
- **The victim can be selected by the analysis.** An opportunistic adversary can concentrate on the informative cases and abandon the rest. Failure to identify most people does not establish safety for the subset it can identify or classify.
- **Cheaper analysis makes existing archives more consequential.** Published research already demonstrates inexpensive automated identification in particular settings. More capable analysis can expose implications of old records without any new disclosure by their subject.[13]
- **Protection must survive some disclosure.** Privacy controls should reduce unwanted exposure and selection. Account security, physical access controls and secure handling of valuable assets must also limit harm after someone becomes identifiable.

The resulting forecast is more accessible, detailed preparation for some forms of stalking, coercion and theft. Reported cases already show online research supporting physical crime.[16] The size and timing of a broader increase remain uncertain. That uncertainty is compatible with acting now on a clear mechanism.

Contents

1 What people actually publish
2 What makes an inference tractable
3 How combining clues changes the answer
4 What the physical examples establish
5 The person can be an output of the search
6 The past can become more revealing
7 What cheaper analysis changes
8 The physical crime forecast
9 What meaningful protection requires
10 How to evaluate the threat honestly
Conclusion
Appendix Technical models and interpretation
Sources

1 What people actually publish

A short recording of a conversation also records a room. It may retain reverberation, background machinery, a passing sound and traces introduced by the recording equipment. A photograph records patterns of light that include whatever happened to be behind its intended subject. The publisher's attention does not determine the information content of the file.

Electrical network frequency offers a concrete example. Small variations in mains frequency can leave traces in recordings. Researchers have used such traces to help establish recording time, synchronize media and investigate which power grid a recording came from. This requires a measurable signal and suitable reference recordings or a validated classification model; a common grid signal is not an address.[1]

The privacy implication extends beyond that particular technique. A background detail can become useful because an analyst has a comparison record, a physical model or other evidence that the publisher never considered. What sounded like irrelevant noise may help answer a different question.

The same process can involve ordinary posts. Imagine someone publishing restoration photographs of a valuable instrument under a hobby username. Their professional profile does not mention it. A friend's event post connects the hobby identity to the professional one. The combined record now associates a privately discussed possession with a named person. This fictional example requires no exotic sensor: the consequential information emerges from the connection.

A person reviewing a post usually asks whether it visibly contains a name, address, document or embarrassing detail. A capable analyst can ask what circumstances would produce the observed data. Those are different reviews. The second includes facts that were never deliberately communicated.

Here, privacy means maintaining chosen boundaries around what particular observers can learn about a person: identity, whereabouts, routines, associations or other personal circumstances. It is contextual. A professional identity can be public while a home address is private; a gathering can be public after it ends while its live location remains restricted.

This definition makes the central failure visible. A service can honor the intended audience of an individual post while leaving its contents easy to combine with other records. A person can exercise care on their own

account while another participant discloses a shared event. Privacy must therefore be assessed across records, observers and time.

2 What makes an inference tractable

An inference begins with an unknown situation and observations that might distinguish it from alternatives. The question is whether the available evidence supports a useful answer at an acceptable cost and error rate.

Five separate requirements matter.

Requirement	What an assessment must establish
Distinguishing information	Different candidate situations must produce measurably different observations under the relevant conditions.
Capture and survival	The useful difference must reach the sensor and survive recording, compression, editing and publication.
Context and reference data	The analyst must possess or infer the timing, comparison records, geometry or other context needed to interpret the observation.
Reliable estimation	The method must distinguish candidates despite noise, changing conditions and plausible alternatives outside the reference set.
Decision value	The resulting uncertainty must be low enough to change a relevant decision. Exact identification is only one possible outcome.

These requirements separate fundamental ambiguity from contingent difficulty. If two explanations produce indistinguishable data, more computation on those data cannot separate them. If the missing ingredient is an external timestamp or an additional observation, the obstacle may be removable. If the obstacle is specialist labor, automation can change the economics. If it is access to a private archive, access must be justified in the threat model rather than assumed.

A motivated adversary is also adaptive. It can change methods, revisit a hypothesis, seek another observation or stop investigating an unproductive case. An assessment that grants unlimited time, universal data access and perfect instruments becomes fiction. An assessment that confines an adversary forever to the first uploaded file also misses how investigation works. The useful model states its resources and lets it make choices within them.

Resolution is an error budget

The right question is what difference must be measured. For a timing measurement, uncertainty may come from the source, the recording clock, the propagation path and the detector. For a geometric measurement, it may come from image resolution, calibration, orientation and the reference frame. Increasing the nominal resolution of one sensor does not automatically improve every part of that budget.

As an idealized scale calculation, a position difference of 100 metres on an Earth of radius about 6.4 million metres corresponds to roughly three arcseconds of central angle. This converts one simplified positional requirement into an angular requirement. It does not say that a handheld photograph supplies that accuracy; the complete geometry and all other errors still matter.

Likewise, neither an audio sample interval nor a Fourier display's frequency-bin spacing is a universal bound on estimation accuracy. Models can estimate parameters between samples or bins, but uncertainty still depends on the signal, observation duration, noise and model fit. Finite noisy observations require an error analysis.[11]

Repeated observations help in two ways. Independent, unbiased measurements of the same quantity can reduce random error; under standard assumptions, averaging four reduces standard error by half. More diverse observations can do something more valuable: distinguish alternatives that the first measurement could not separate at all. A shared calibration error or a missing reference frame does not disappear merely because many measurements agree.

Demonstration is evidence rather than a ceiling

A published experiment establishes an achieved capability under its stated conditions. It is useful evidence about feasibility, inputs and failure modes. It does not establish the maximum capability of a differently resourced observer.

Conversely, a plausible physical mechanism does not establish a measured success rate. A serious assessment should be willing to derive a requirement, estimate whether it is attainable and propose a discriminating test. It should label that work as a derivation or hypothesis. The absence of an end-to-end public demonstration should leave an explicitly analyzed question, rather than an automatic finding of safety.

3 How combining clues changes the answer

Sensor fusion means estimating a shared situation using multiple measurements, with their uncertainties and relationships taken into account. Data fusion extends the same idea to records that need not come directly from physical sensors. NASA's statistical fusion work, for example, combines observations with different errors, coverage and sampling patterns to estimate common physical quantities.[5]

The basic intuition needs no specialist equipment. Imagine 64 cards containing every combination of four letters, four shapes and four numbers. Learning the letter leaves 16 possible cards. Learning the shape leaves four of those. Learning the number leaves one.

Each clue, considered on its own, is compatible with 16 cards. Together, they identify one. Saying that every clue is individually inconclusive is therefore compatible with a decisive joint result.

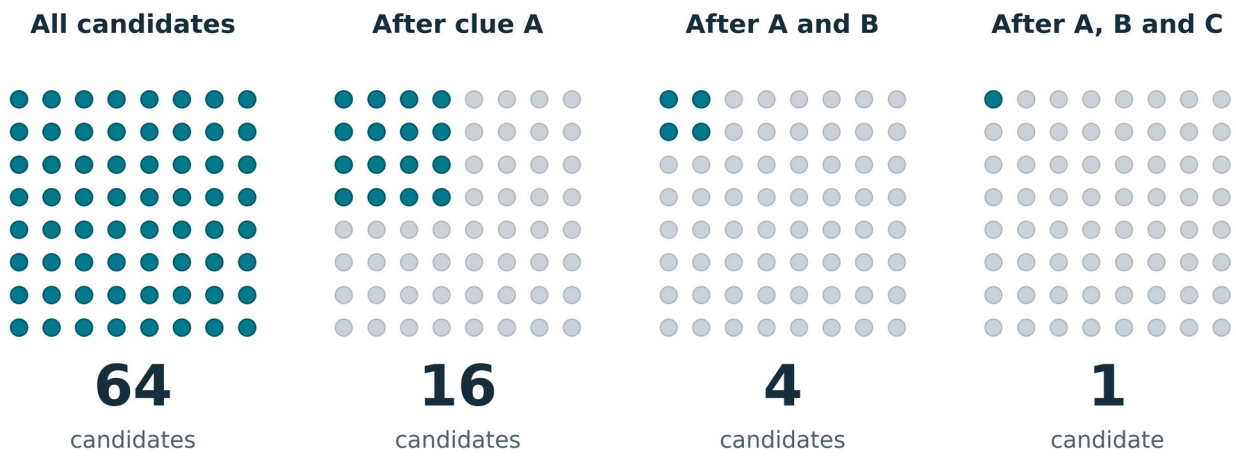


Figure 1. A constructed example with exact clues and a complete candidate set. The counts illustrate the mechanism; they are not measured identification rates.

In this example, each clue contributes two bits: enough to distinguish four equally likely alternatives. The uncertainty falls from six bits to zero. The information adds while the number of remaining candidates shrinks multiplicatively. Its practical value can change sharply when the remaining uncertainty becomes manageable.

Real clues often change relative plausibility rather than eliminate candidates. A sound may be more likely in one kind of space than another. An observation may narrow the time of a recording. A separate record may make one location more plausible. Good inference carries those uncertainties through the combined result.

Some information exists in the relationship

There is also a stronger form of interaction. Consider two independent fair coins and ask whether they match. Seeing either coin alone tells you nothing about that answer. Seeing both tells you completely. The information about the relationship is available only jointly. Information theory calls this kind of contribution synergy.[6]

For privacy, one record might contain an anonymous observation while another supplies the identity connection that makes it personal. Neither has the same meaning alone. A timing reference can make an otherwise ambiguous environmental signal interpretable. The new clue's value depends on what is already known.

This is the rigorous meaning behind the intuition that the combined result can exceed the apparent value of its parts. It does not mean that arbitrary confidence scores should be added or multiplied. Two websites repeating the same claim provide one underlying observation. Two sensors can share a bias. A model can make several confident claims from the same mistaken assumption.

Additional evidence reduces uncertainty on average under a correct model. A particular new observation can properly increase uncertainty or overturn an earlier conclusion. An inference system that always becomes more confident as material accumulates is not necessarily learning correctly.

Sparse clues can distinguish rich records

In a study of 1.5 million mobility traces, four randomly chosen points in time and space uniquely distinguished 95 percent of traces within the studied dataset, using hourly time resolution and locations at the carrier-antenna scale.[7] A uniquely distinguished trace still needs a connection to a person before it becomes a named identification. The result nevertheless demonstrates how a few observations can discriminate among extensive records.

Earlier work on the Netflix Prize dataset showed how auxiliary information could help connect anonymized rating records to outside identities.[8] Removing an explicit name does not remove every relationship that can reconnect a record to its subject.

4 What the physical examples establish

The following examples illustrate distinct ways that information becomes useful. Their dependencies belong inside the analysis: some are available from other records, some require additional collection, and some remain substantial barriers.

Stars can provide a reference frame

Astrometry.net demonstrated recovery of celestial pointing, image scale and orientation from astronomical image pixels.[2] A separate stellar positioning prototype combined star observations with local gravity and reported errors within 50 metres at three test locations during short intervals.[3]

These are different achievements. A bare starfield generally supplies celestial orientation; terrestrial position needs further constraints, including an Earth-linked direction and timing or an equivalent longitude constraint. The prototype's result cannot be transferred to an arbitrary bedroom photograph.

For a joint assessment, the next question is whether the rest of the record supplies the missing constraints with sufficient accuracy. That question requires geometry and an error budget. The starfield's contribution depends on the rest of the available evidence.

Echoes can constrain a space

Researchers have reconstructed room geometry from acoustic echoes using a known sound and a microphone array.[9] This establishes that acoustics can encode spatial structure. Separate research has estimated room acoustic properties from a 360-degree photograph.[10]

Combining incidental audio with candidate interior photographs is a plausible research hypothesis. The relevant test is whether the audio helps discriminate among realistic candidate spaces. It need not reconstruct an entire building to matter. Unknown sound-source position, furnishings, materials, recording-device behavior and incomplete photographs all belong in that test. The cited studies do not establish reliable home identification from arbitrary social-media clips.

This example shows why first-principles analysis matters. A coarse estimate of a room property may have little value during a global search and substantial value after other evidence leaves a few candidates. Its contribution must be measured in that conditional setting.

Background events can join separate records

A shared event can connect recordings made by different devices. A moving source can also supply geometric information through changes in its received signal. For example, a known trajectory and a

measured Doppler pattern can constrain a recorder's position under an appropriate model. The event, trajectory, timing and propagation uncertainties all matter.

This is a physical derivation, not a claim that every passing vehicle identifies its listener. Multiple observations help when they add distinct constraints. The useful capability may be associating two recordings, narrowing a time interval or rejecting a candidate location. The full address need not emerge from the acoustic measurement itself.

The consequential step is the join between incidental media and reference information. An assessment should measure how much that join reduces uncertainty and which additional constraints it requires.

Hardware can leave measurable signatures

Digital communication is implemented by physical hardware. Research on radio-frequency fingerprinting has used transmitter characteristics to distinguish enrolled devices. One study explicitly examined receiver changes and collaborative inference using ten LoRa devices and twenty receivers.[4]

A usable fingerprint must remain sufficiently stable and discriminating across relevant conditions. Physical differences alone do not guarantee that the published measurement preserves them. A fingerprint can also link observations without identifying an owner; an additional anchor supplies that connection.

Filtering needs a specified purpose

Removing unnecessary frequencies can reduce particular exposures. It is not a general certificate that an audio file has become private. Mains-frequency traces and harmonics can lie inside the nominal audible range.[1] Meanwhile, a signal can be inaudible to a listener because it is faint or masked, even when its frequency is within that range.

The defensible question is whether a transformation removes a specified source of information while preserving the intended content. Hearing no revealing background sound is a different test. A product that promises privacy should identify the channels it handles and validate the result against plausible inference methods.

5 The person can be an output of the search

A predetermined-target investigation asks what can be learned about a particular person. An opportunistic search asks which people or situations meet some criterion. The second can succeed while returning no answer for most of the population.

This changes the denominator. A method with limited coverage can still produce many consequential results when applied to a large pool. An adversary can discard ambiguous cases, concentrate attention on informative ones and then seek more evidence about the selected subset. Population screening and individual investigation can be consecutive stages of the same process.

Research on inferring personal attributes from social and behavioral relationships illustrates the statistical distinction. In one Google+ and Google Play study, selecting the more confidently predicted half of test users increased precision for predicting one city in a user's reported cities-lived history from about 57 percent to more than 92 percent.[12] Those figures describe that dataset and attribute. They do not establish a current location, address identification or prediction of household security habits.

The general lesson is that average accuracy can conceal a much more useful selected subset. An assessment should report coverage, precision among returned answers, false matches and the number of cases to which the method can reasonably be applied. Precision must also be checked in the population being searched:

when the sought condition is rarer, the same error behavior can produce a much larger share of false selections.

Selection can arise from an unexpected combination

The instrument example shows how this can happen. A hobby record establishes a possession; another record supplies the identity connection. Someone searching for that circumstance could become interested in the person only after the records were linked. The person did not have to be independently important enough to attract an investigation beforehand.

Similarly, a person's own account may reveal little about a forthcoming event while someone else's post supplies the missing association or timing. The relevant disclosure can occur outside the subject's direct control. Work on multiparty privacy has examined precisely this dependence on other people's sharing and relationships.[14]

These examples should not be converted into claims that any imagined behavioral classifier works. Predicting a particular habit from unrelated posts needs suitable ground truth, population testing and calibration. Intuitive plausibility is enough to motivate an evaluation, not enough to assert a success rate.

An attacker also need not be right to cause harm. A mistaken identification can direct harassment or suspicion at an uninvolved person. False positives reduce the reliability of the inference; they do not automatically eliminate its consequences.

Certainty depends on the next decision

Different actions require different levels of confidence and precision. An inference can be useful as a lead, as a way to rule something out or as a reason to concentrate attention. Proof suitable for a courtroom is a much higher standard than many informal decisions require.

Likewise, a coarse location can change the cost of further inquiry even when it does not identify a particular residence. Further inquiry has its own expense and remaining uncertainty; the assessment should model both.

The useful boundary is whether the inference materially changes what an observer can do within its resources. Requiring every individual signal to finish the entire task sets the wrong threshold.

6 The past can become more revealing

Records can acquire new significance after publication. A later identity connection can attach old writing to a person. A new reference recording can make an old environmental trace interpretable. A better method can extract relationships that earlier tools missed.

This is why deleting an explicit label or using a pseudonym may be insufficient. Stylometric research has studied linking authors across different writing platforms.[15] Such results do not mean that every writer can always be recognized. They show that writing behavior can become another source of evidence when identities are otherwise separated.

The important temporal asymmetry is that publication happens once while interpretation can happen repeatedly. A person may stop sharing and still face new inferences from existing copies. A collection of apparently innocuous fragments can become sensitive when a later record supplies the missing context.

The inference can also propagate through relationships. A record may first establish that two observations came from the same source, without revealing who that source was. A later anchor can identify the cluster.

Uncertainty must propagate too: a dubious link should not become a certainty merely because it is stored in a database.

Archives do not automatically remain useful forever. Places change, habits change and observations become stale. Retention and deletion policies can reduce future exposure, although removal from one service cannot recall every independently held copy. Both facts matter: stale evidence should lose weight, while durable records can preserve risk long after the subject stops thinking about them.

Third-party records complicate every individual control. Friends, employers, venues, customers and service providers may hold related information. Effective privacy therefore includes coordination and institutional decisions about collection, access and retention. A checklist addressed only to the subject cannot govern the entire record.

7 What cheaper analysis changes

The cost of an inference includes obtaining data, preparing it, extracting features, connecting records, estimating an answer and checking it. Some of that work once required scarce specialist time. Automation can reduce several components, and it can make methods accessible to people who could not previously perform them.

A 2026 USENIX paper reports an identification agent costing roughly US$1–4 to run per profile in its experiments. Its Hacker News evaluation achieved about 67 percent recall at 90 percent precision on a selected sample whose profiles had originally linked to LinkedIn; direct identifiers were removed for the test. The authors describe that population as easier than typical pseudonymous accounts.[13]

This is a measured capability in a bounded setting. It is enough to challenge the assumption that connecting scattered personal information always requires expensive manual investigation. The quoted operating cost is not the cost of obtaining every possible dataset or executing a physical crime.

Price a useful result

A model subscription price, the budget for a research prototype and the recurring cost of a verified inference answer different questions. A prototype may be expensive to build and cheap to reuse. A cheap model call may be expensive in practice if it creates large amounts of manual verification work.

The relevant unit is the cost of a sufficiently reliable output that changes a decision. It should include failures, abstentions, correction and the cost of the data actually used. Comparisons need an explicit baseline: a skilled analyst, a simpler automated system or no investigation at all.

Lower cost can make an existing investigation cheaper and make additional cases worth considering. The latter matters particularly for personal security: people who were never individually valuable enough to research may become visible during inexpensive screening. Better information can also improve decisions and reduce subsequent wasted effort. Its value need not be limited to the analyst hours it replaces.

Physical activity retains physical costs

If analysis accounts for 10 percent of an activity's total cost, making it ten times cheaper reduces total cost by 9 percent, holding everything else fixed. If analysis accounts for 90 percent, the same improvement reduces total cost by 81 percent. These are arithmetic scenarios, not empirical estimates.

That calculation isolates the price of analysis. Better information can also change which activities are attempted, how much uncertainty remains and how much subsequent work is wasted. Those effects can dominate the direct saving; 9 percent is not a ceiling on the value of better intelligence in the first scenario.

Travel, physical access, logistics, exposure to detection and converting stolen property into value do not vanish because analytical work becomes cheaper. The expected effect should therefore vary across activities. It can be substantial wherever discovering and understanding opportunities changes consequential decisions.

This distinction also explains why a single crime-rate multiplier would be misleading. Cheaper intelligence can change who is selected, how tailored the preparation is and which incidents become feasible without producing the same increase in every type of crime.

Information and computation must be distinguished

Analysis cannot create missing evidence merely by producing a convincing story. Generative reconstruction can fill gaps using learned expectations. Apparent detail is not proof that the missing detail was recovered.

Its real economic significance is that previously impractical implications of available evidence can become accessible. That can be a large change without requiring new physical laws, perfect sensors or universal surveillance access.

8 The physical crime forecast

Online information already contributes to physical victimization. The FBI has described organized burglary groups combining internet research with surveillance and commercially available technology.[16] That reporting establishes an existing connection between digital information and physical crime. It does not identify AI as the cause of the described activity.

My forecast is that cheaper extraction, linkage and inference will make detailed preparation accessible to a wider range of offenders and usable against a wider range of people. Some incidents will appear unusually well informed because their preparation no longer required the time and expertise observers assume.

The mechanism is strongest when useful information is already available and better interpretation changes which cases are worth pursuing or reduces consequential uncertainty. A limited subset of informative cases can be sufficient. It weakens when the necessary observations are inaccessible, the signal is absent or the remaining barriers are largely insensitive to better information.

There are several reasons to expect uneven adoption. Tools vary in reliability; access to useful data varies; some tasks demand tacit knowledge or local presence. Defenders also adapt. Those differences affect timing and scale, while leaving the underlying incentive intact.

Institutional recognition has already begun. The European Commission's summary of Europol's 2025 organized-crime assessment identifies AI as increasing criminal capability and accessibility.[17] Awareness is not the same as having effective controls, measurement or investigative capacity for every emerging use.

The forecast should be tested through changes in the cost, quality and reach of preparation. More individualized intelligence can become available for harmful uses while adoption and harm remain concentrated in particular activities. A uniform increase across all crime categories is unnecessary for the transition to matter.

Monitoring should therefore track how incidents were prepared, the sources and tools involved, the amount of human work required and whether the preparation depended on automated inference. Sophistication alone should not be treated as proof of AI involvement. Without such evidence, dramatic anecdotes will be poor measures of the trend.

9 What meaningful protection requires

Privacy and security address connected but different parts of the problem. A person may want to remain unlinked across contexts, avoid timely exposure of an activity, reduce unwanted selection or remain safe after an identity becomes known. Effective protection starts by naming the desired boundary.

The existence of an expert countermeasure does not make privacy an ordinary default. A precaution that requires signal-processing knowledge, constant account review and coordination with everyone nearby carries a real burden. That burden belongs in the assessment of whether protection is usable.

Begin with one important boundary: what you want to keep private, from whom and for how long. Review the public records you control for that boundary, reduce unnecessary live or routine disclosure and agree on sharing expectations with the people involved. Maintain protections that remain useful if someone learns the information anyway. Add technical media precautions when the particular exposure warrants them.

Reduce unnecessary disclosure

Choose what information needs to be released for the intended purpose. A text summary, a limited view or a smaller audience can sometimes serve that purpose with less incidental information than a raw recording. Consider timing as well as content: a disclosure after an event can have different consequences from the same disclosure while it is happening.

Metadata removal and media transformations can address particular channels, but their scope should be explicit. Retaining a visible identifying scene while removing coordinates does not remove the scene. Filtering a frequency band does not address unrelated timing, linguistic or visual clues.

The objective is proportionate minimization, not a demand that everyone manually inspect a spectrogram before posting. The unusual examples show how far human intuition can diverge from machine inference. Products should absorb as much of that technical burden as possible.

Preserve boundaries across contexts

Where separation matters, reduce unnecessary connections between accounts and audiences. Evaluate the accumulated record, including repeated details and shared relationships. Changing one surface feature should not be treated as a reliable barrier against all forms of linkage.

For shared activities, agree on audience and timing expectations with the people involved. Platforms can help by supporting meaningful controls for multiple affected people. Research on multiparty privacy makes clear why a subject's own settings alone may be insufficient.[14]

The practical standard should reflect what the person is trying to protect. Someone separating a sensitive support-group identity from professional life has a different need from an organizer limiting the live location of a private event. A useful system should help maintain the specified boundary without requiring constant expert interpretation.

Remain protected after exposure

Reducing selection is useful, but concealment cannot carry the whole security burden. Actual access controls, account protections and secure custody or handoff of valuable items still matter when a person's identity or general whereabouts are known. Postal Inspection Service guidance, for example, includes secure handling and prompt collection of deliveries as ways to reduce theft exposure.[19]

Defensive planning should ask what remains protected if a privacy measure fails. A public identity should not imply easy account takeover. A visible interest should not imply unprotected assets. A disclosed destination should not remove every relevant safety measure.

This also avoids a counterproductive response: trying to look like a low-value target to an imagined classifier while neglecting protections that directly prevent harm. Without validation, manipulating speculative behavioral signals is guesswork. Concrete controls can protect across several ways an adversary might select someone.

Put responsibility where control exists

Responsible party	Priority	What success would look like
Individuals and households	Choose important privacy boundaries and reduce unnecessary public detail; maintain protections after exposure.	A manageable routine that addresses the person's actual risks rather than an endless list of exotic precautions.
Friends and organizers	Coordinate the audience and timing of shared information.	One participant's disclosure is less likely to defeat another's stated boundary.
Platforms and product teams	Minimize unnecessary capture, retention and cross-context exposure; make protective settings understandable and effective.	Ordinary use requires less specialist work to preserve reasonable privacy.
Data holders	Limit access and retention according to purpose; audit combinations and exports as well as individual fields.	Risk assessments address what the available collection permits an observer to infer.
Researchers and public institutions	Measure realistic combined capabilities and preparation methods; support prevention and recovery.	Better evidence about who is exposed, which controls work and how emerging harms are actually enabled.

This allocation follows a broader security-by-design principle: manufacturers should take responsibility for protective defaults and user outcomes.[18] A warning that shifts an unmanageable analytical task onto the user is not an adequate substitute for a safer product.

10 How to evaluate the threat honestly

A useful assessment should begin with an outcome and a bounded observer. State what that observer wants to learn, which records it can access, what additional collection it can perform, what time and money it can spend and what counts as a useful result. Evaluate both a predetermined target and selection from a broader population when both are relevant.

The benchmark should then test combinations. Measure how much an observation helps alone, after other evidence is available and across repeated observations. Compare the combined method with simpler baselines. Preserve realistic compression, missing data, stale records and changes in equipment or environment.

Candidate sets must include realistic alternatives and cases for which the correct answer is absent. A system that always chooses the nearest enrolled match can appear successful while failing badly on new people or places. Allow abstention and measure whether its confidence corresponds to observed accuracy.

Report the full path from input to useful output: sample selection, data availability, costs, human verification, coverage, precision, false matches and consequential errors. Separate development cost from recurring cost. Do not multiply the best result from one experiment by the cheapest component from another and call it a measured capability.

Use consenting participants and controlled environments for evaluations. The purpose is to establish exposure and test protections. It does not require exposing private individuals or treating a plausible generated reconstruction as ground truth.

Finally, distinguish four kinds of claim. A **demonstration** reports what occurred under stated conditions. A **derivation** follows from a model with explicit assumptions. A **hypothesis** describes a plausible combination requiring validation. A **forecast** estimates how incentives and capabilities may change behavior. These distinctions make an assessment more forceful because the reader can see exactly what supports each conclusion.

Conclusion

The records people publish are measurements that can be combined, reinterpreted and connected to them over time. Their privacy depends on the resulting collection, the observer's resources and the protection that remains after an inference is made.

Cheaper analysis changes that relationship even when it remains imperfect. It expands the set of records worth examining and the set of people who can be selected from them. A serious assessment must account for that process instead of requiring every isolated clue to identify everyone.

Privacy remains attainable. The failure is making ordinary people perform a specialist's job to maintain boundaries they reasonably expected to have. Better defaults, less unnecessary retention and protections that survive partial exposure should carry much more of that work.

Appendix Technical models and interpretation

A Joint inference with nuisance variables

Let X represent the fact or situation of interest, E the observations, R auxiliary information and Z uncertain recording or environmental conditions. A general model is:

$$p(X, Z \mid E, R) \propto p(E \mid X, Z, R)\ p(X, Z \mid R)$$

The estimate of X must account for uncertainty in Z rather than silently fixing unknown conditions. Timestamps, orientation, equipment response and source behavior can all be nuisance variables. Different sources can remain dependent even when they appear on different websites.

Additional observations can constrain X, improve knowledge of Z or supply part of R.

B Conditional information

The information supplied by several observations obeys the chain rule:

$$I(X; A, B, C) = I(X; A) + I(X; B \mid A) + I(X; C \mid A, B)$$

The contribution of B after A is known can differ from B's contribution alone. Joint information can include synergy, while overlapping sources also introduce redundancy.[6]

C Coverage and consequential outputs

For N attempted cases, let g be the fraction receiving an answer and p the precision among those answers. Expected correct outputs are N × g × p; incorrect outputs are N × g × (1 − p). These are counts of analytical outputs under the model, not counts of crimes. Generalizing them requires a representative population and a stable evaluation procedure.

D Sensitivity to analytical cost

If analytical work is fraction f of an existing activity's total cost and becomes r times cheaper, while other costs remain fixed:

$$\frac{C_{new}}{C_{old}} = 1 - f + \frac{f}{r}$$

The model isolates a cost effect. It omits changes in which cases are considered, additional verification, defensive adaptation and changes in the value of an inference. Those effects need separate analysis.

E Protection after selection

For a class of harms requiring selection by an observer, a simple decomposition of expected harm is:

$$P(\text{selected})\ P(\text{harm} \mid \text{selected})\ E[\text{severity} \mid \text{harm}, \text{selected}]$$

This expression uses conditional probabilities; it does not assume the factors are independent. It distinguishes reducing unwanted selection, preventing harm after selection and limiting consequences. Privacy controls can contribute to more than one factor, but anonymity is not the only available protection.

Sources

Primary research and official material support the empirical claims below. Numerical findings retain the conditions of their original studies. The report's physical-crime forecast and illustrative calculations are analytical judgments, not reported incident rates. Sources were checked on 3 November 2026.

[1] University of Maryland, MAST. *Electrical Network Frequency research*. Recording-time inference, synchronization and grid-related inference. mast.umd.edu

[2] Lang, D., et al. 2010. *Astrometry.net: Blind astrometric calibration of arbitrary astronomical images*. arxiv.org/abs/0910.2233

[3] Woodbury, D. P., et al. 2010. *Stellar Positioning System Part II: Improving Accuracy During Implementation*. Institute of Navigation. ion.org

[4] Shen, G., et al. 2022. *Towards Receiver-Agnostic and Collaborative Radio Frequency Fingerprint Identification*. arxiv.org/abs/2207.02999

[5] NASA Jet Propulsion Laboratory. *Statistical Data Fusion for Earth Remote Sensing*. uq.jpl.nasa.gov

[6] Williams, P. L., and R. D. Beer. 2010. *Nonnegative Decomposition of Multivariate Information*. arxiv.org/abs/1004.2515

[7] de Montjoye, Y.-A., et al. 2013. *Unique in the Crowd: The privacy bounds of human mobility*. Scientific Reports. Author-institution record: media.mit.edu

[8] Narayanan, A., and V. Shmatikov. 2008. *Robust De-anonymization of Large Sparse Datasets*. IEEE Symposium on Security and Privacy. cs.cornell.edu

[9] Dokmanić, I., et al. 2013. *Acoustic echoes reveal room shape*. Proceedings of the National Academy of Sciences. pmc.ncbi.nlm.nih.gov

[10] *Room Acoustic Properties Estimation from a Single 360° Photo*. 2022. Proceedings of EUSIPCO. eurasip.org

[11] Rife, D. C., and R. R. Boorstyn. 1974. *Single-tone parameter estimation from discrete-time observations*. NASA record. ntrs.nasa.gov

[12] Gong, N. Z., and B. Liu. 2016. *You Are Who You Know and How You Behave: Attribute Inference Attacks via Users' Social Friends and Behaviors*. USENIX Security. usenix.org

[13] Lermen, S., et al. 2026. *Large-scale online deanonymization with LLMs*. USENIX Security. usenix.org

[14] Thomas, K., C. Grier and D. M. Nicol. 2010. *unFriendly: Multi-party Privacy Risks in Social Networks*. research.google

[15] Overdorf, R., and R. Greenstadt. 2016. *Blogs, Twitter Feeds, and Reddit Comments: Cross-domain Authorship Attribution*. Proceedings on Privacy Enhancing Technologies. petsymposium.org

[16] Federal Bureau of Investigation. *Inside the FBI: Intercepting the South American Theft Group Threat*. Transcript discussing online research and physical surveillance in organized burglary. fbi.gov

[17] European Commission. 21 March 2025. *New report shows how organised crime is changing and affecting EU security*. Official summary of Europol EU-SOCTA 2025. home-affairs.ec.europa.eu

[18] CISA and partners. 2023. *Shifting the Balance of Cybersecurity Risk: Principles and Approaches for Security-by-Design and -Default*. cisa.gov

[19] United States Postal Inspection Service. *Mail and Package Theft*. uspis.gov