FIELD GUIDE · VERSION 1.0 · OCTOBER 2026

SHADOW [redacted] STUDIO

The Creator Privacy & Provenance Handbook

Redaction risks, consent decisions, approval workflows, provenance templates, analytics literacy and a printable public/private matrix.

All examples use invented people and data.

Sahil Tayal
sahiltayal.com

ORIENTATION

How to use this guide

This handbook is for independent creators who publish their work and live their lives in the same browser. It is built around one distinction: **transparency about the work is not exposure of the person**. It accompanies the interactive exhibition at shadowstudio.sahilverse.online but is designed to stand alone, on paper.

> Every person, project, address and number in the examples is invented. The recurring case is *The Tidewater Almanac*, a fictional zine by a fictional maker, Ira Vell.

Learning objectives

By the end you should be able to: (1) separate the work, the process, the person and other people when deciding what to publish; (2) spot the identifiers that automated redaction misses; (3) run a two-stage approval before anything leaves your workshop; (4) write an honest provenance note, including AI use; (5) read engagement numbers without mistaking them for impact; (6) keep a record of what you chose not to publish.

Contents

Page	Section	Tool you leave with
3	Front stage, backstage: key ideas	Glossary
4	The four boundary rings	Rings diagram
5	Redaction risks on invented data	Risk table
6	Worked example: one newsletter, redacted	Before/after
7	The consent decision tree	Decision tree
8	Printable public/private matrix	Matrix
9	Publication approval workflow	Two-stage checklist
10	Provenance templates	Origin card
11	Analytics literacy exercises	Chart + questions
12	Threat-model notes	Threat table
13	Disclosing AI collaboration	Disclosure ladder
14	Activities and worksheets	Two worksheets
15	Limits, objections and uncertainty	—
16	References and further reading	—

Statement types are marked as in the website: **Observation** (what can be seen), **Inference** (a reasoned interpretation), **Invented scenario** (fiction used for illustration) and **Empirical** (a claim from a cited study or document). Nothing here is legal advice.

KEY IDEAS

Front stage, backstage

Empirical. The sociologist Erving Goffman (1959) described everyday life using a theatrical frame: a *front region* where a performance is given to an audience, and a *back region* where it is prepared, where props are kept and where performers can relax the performance. Goffman did not treat the back region as a lie. It is where the performance is made possible.

Inference. A creator's public archive is a front region that never closes: searchable, permanent, linkable. Readers can mistake the completeness of the *record* for completeness of the *person*. The backstage still exists; the question is what crosses the rope, and who decides.

Empirical. Helen Nissenbaum's *contextual integrity* (2004) argues that privacy is about appropriate flows: information shared in one context carries that context's norms. Marwick and boyd (2011) showed how social platforms collapse separate audiences into one, so people write for an imagined audience that rarely matches the real one.

Glossary

Term	Working definition in this guide
Omission	Leaving information out without implying anything false about it. Legitimate by default.
Deception	Leaving out or shaping information so that a reader forms a false belief that matters to them.
Identifier	Any detail that, alone or combined with others, points to a specific person.
Quasi-identifier	A detail harmless alone (postcode, birth date, job) that identifies when combined.
Provenance	A record of where a work came from: who authorised it, what it rests on, what changed.
Context collapse	Separate audiences merging into one, so a message meant for some reaches all.
Redaction	Irreversibly removing content from a copy, not hiding it with formatting.
Cold read	Reviewing a draft after a pause, as the person most affected by it would.

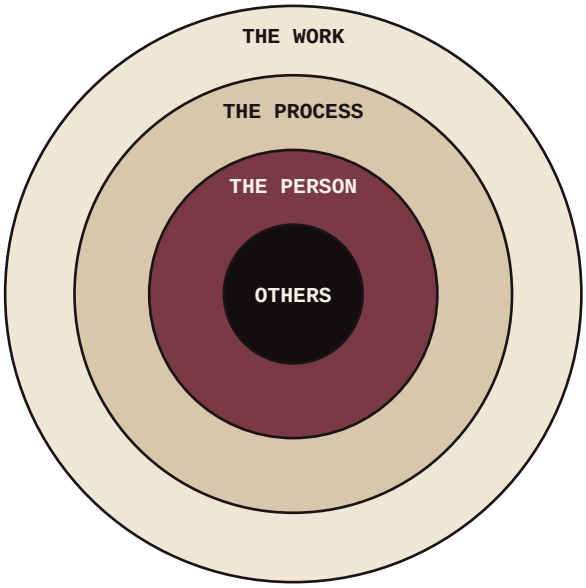
Omission versus deception: a test

Ask: *if the reader later learned what I left out, would they feel misled about the work, or merely reminded that I have a life?* Leaving out a health crisis from a launch post is the second. Claiming a project was "made in a weekend" when it took a year is the first.

THE MODEL

Four boundary rings

Before publishing anything, locate each element in one of four rings. The rings are a **model**, not a finding: they compress a messy reality into a decision aid, and you will meet cases that straddle two rings.



The work
Default public. Finished pieces, credits, method.

The process
Share deliberately: drafts, repairs, open questions.

The person
Default private: health, money, fears, relationships.

Others
Not yours to share: their consent, their context.

Figure 1. The four rings. Default publicity decreases toward the centre. "Others" sits at the core not because other people matter most to your work, but because their information is the least yours to release. Text alternative: four concentric circles labelled, from outside in, The work, The process, The person, Others.

How to use the rings

Ring	Default	Typical contents	Common mistake
The work	Public	Finished piece, credits, stated method, corrections	Hiding corrections to look flawless
The process	Deliberate	Drafts, failures, timelines, open questions, tools used	Performing struggle as content, or hiding it in shame
The person	Private	Health, money, home, relationships, beliefs, fears	Oversharing to seem authentic
Others	Not yours	Family, friends, clients, interviewees, readers who wrote privately	Assuming your openness covers them

Inference. Most regret after publishing comes from the inner two rings leaking out through the outer ones: a family name in a screenshot, a salary visible in a spreadsheet tab, a reader's email quoted in a thank-you post.

REDACTION

What automated rules miss

The website's Redaction Desk marks candidates with deliberately conservative rules: emails, phone-shaped numbers, long digit runs, currency, postcodes, numeric dates, @handles, and names directly after family words. That catches the obvious. The dangerous leaks are usually not obvious.

Empirical. Sweeney (2000) estimated that about 87% of the 1990 US population was likely uniquely identifiable from just five-digit ZIP code, gender and full date of birth. Narayanan and Shmatikov (2008) re-identified records in an "anonymised" movie-ratings dataset by matching them with a handful of public ratings. Combinations identify.

Risk (invented example)	Why rules miss it	Safer move
"Thanks to the only Parsi baker on Lantern Lane"	No number, no email: a unique description	Generalise: "a local baker"
Screenshot of a draft showing a browser tab "Dad - hospital forms"	Text inside an image	Crop, then re-check the crop at full zoom
Photo of the workbench with an invoice in the corner	Image; small text	Re-shoot, do not blur: blur can sometimes be reversed
File named ira-vell-salary-negotiation-v2.pdf	File names and metadata are not body text	Rename; strip document properties
"My ex, who you'll all recognise from Issue 2"	No name, but a pointer	Remove the pointer, not just the name
PDF with a black box drawn over a phone number	The text is still underneath	Use a true redaction tool; copy-paste test
Alt text: "Maya in her hospital bed laughing"	Accessibility fields are rarely reviewed	Review alt text with the same care as captions
Timestamps showing you post at 3 a.m. every night	Patterns, not items	Ask what a stranger could infer from the archive as a whole

Rule of thumb: **redact the copy, never the original**; then open the exported copy in a different program and try to select, search and copy the redacted area.

WORKED EXAMPLE

One newsletter, redacted

Invented scenario. Ira drafts a "behind the scenes" newsletter for Issue 2 subscribers. Below is the draft, then the decisions, then what was sent.

DRAFT (all details fictional)

```
Hi friends! Issue 2 is late because my sister Maya had surgery at City Hospital and I've
been her ride every day. Money is tight too: the printer quoted INR 48,500 and my account
is basically empty. Huge thanks to Dev from the ferry office (+91 98100 12345) who leaked
me the 1913 minutes even though he wasn't supposed to. Proofs are at 14 Lantern Lane if you
want to pop by.
```

Element	Ring	Rule caught it?	Decision and reason
Sister's surgery, hospital name	Others	Partly ("surgery", "my sister Maya")	Remove. Her health, her story. Say only "family commitments".
Printer quote, empty account	Person	Amount only	Remove balance. Keep a general note that printing is expensive, if useful.
Dev's name, role, phone number	Others	Phone only	Remove all three. "Leaked" exposes him to real consequences; name + office identifies him even without the phone.
1913 minutes	Work	No	Keep, but describe as "council minutes from 1913" and check they can be cited openly.
Home address for proofs	Person	Yes (address)	Remove. Offer a public pick-up point or post.

SENT

```
Hi friends! Issue 2 is running a few weeks late because of family commitments. Thank you
for waiting. Good news from the archive: council minutes from 1913 have settled one of our
open questions about the harbour. Printing is the expensive part this time; I'll explain
the new price on the order page. Proofs will be at the library table on Saturday for anyone
who wants an early look.
```

Notice. Three of the five risky elements were only partly caught by rules. The most serious one, exposing a source who broke a rule to help, contained nothing a pattern-matcher would flag except a phone number. Human judgement did the important work.

CONSENT

The consent decision tree

Use this for any single item: a photo, a quotation, an anecdote, a screenshot. It is intentionally cautious. Real decisions may justify departing from it, for example in public-interest journalism, which has its own ethics and is outside this guide's scope.

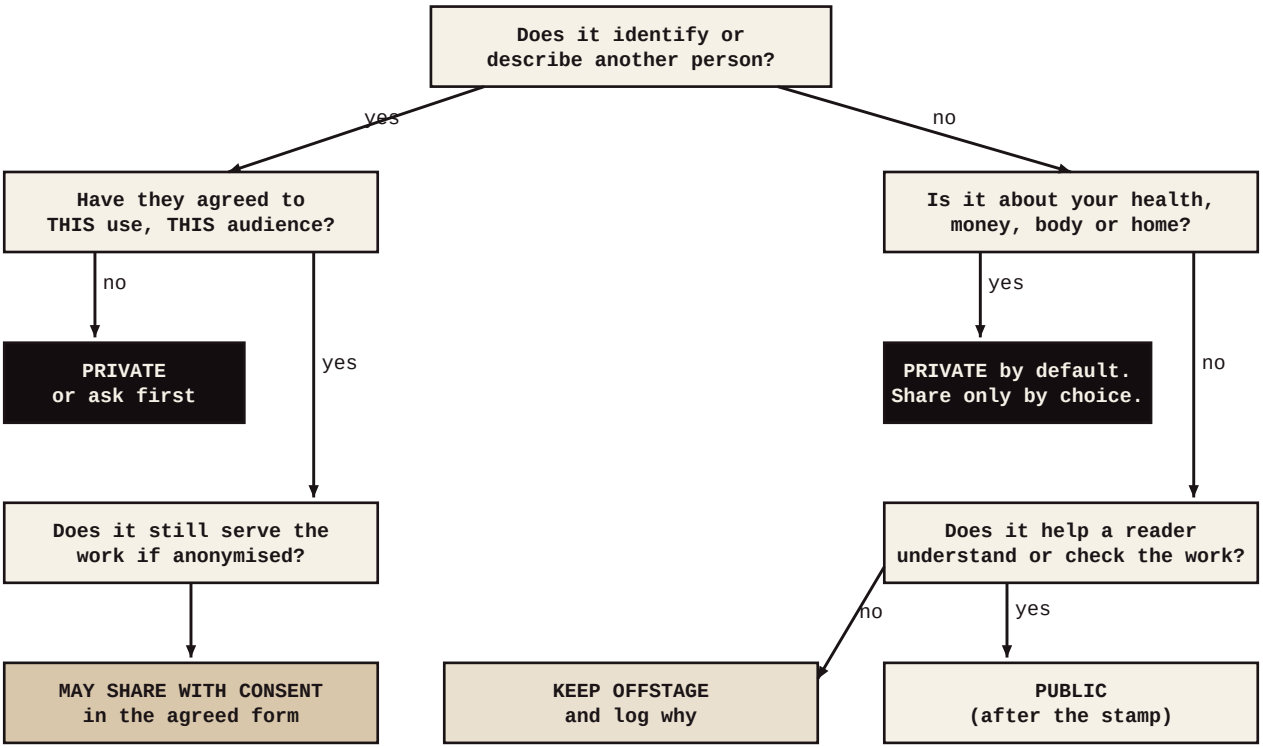


Figure 2. Consent decision tree. Text alternative: If an item identifies another person, ask whether they agreed to this use and audience; if not, keep it private or ask first; if yes, consider whether it still serves the work when anonymised, and share in the agreed form. If it does not involve others but concerns your own health, money, body or home, it is private by default and shared only by choice. Otherwise, if it helps readers understand or check the work, it can go public after the approval stamp; if not, keep it offstage and log why.

What counts as consent here

Specific (this item, this use), informed (they have seen it in context), current (not assumed from years ago), revocable (you will take it down if asked) and free (no pressure from friendship, money or power). A "sure, whatever" in a group chat rarely meets all five.

PRINTABLE

Public / private boundary matrix

Print this page. For your own practice, tick the cell that matches your default for each category, then compare with a collaborator. Disagreement is useful: it surfaces assumptions before a publication does.

Category	Private	Consent	Public	Notes / exceptions
My own health or diagnoses	[ ]	[ ]	[ ]	
Family members' health	[ ]	[ ]	[ ]	
Bank balances, debts, income	[ ]	[ ]	[ ]	
Project budgets (summary)	[ ]	[ ]	[ ]	
Romantic relationships, past and present	[ ]	[ ]	[ ]	
Family disputes	[ ]	[ ]	[ ]	
Home address, routines, location	[ ]	[ ]	[ ]	
Drafts and failed versions	[ ]	[ ]	[ ]	
Correction and repair history	[ ]	[ ]	[ ]	
Interview recordings	[ ]	[ ]	[ ]	
Readers' private messages	[ ]	[ ]	[ ]	
Photos with other faces	[ ]	[ ]	[ ]	
Tools used, including AI	[ ]	[ ]	[ ]	
Analytics and follower numbers	[ ]	[ ]	[ ]	
Client names and fees	[ ]	[ ]	[ ]	
Unreleased business plans	[ ]	[ ]	[ ]	
Beliefs: religious, political	[ ]	[ ]	[ ]	

Suggested defaults appear in the exhibition's Disclosure Threshold. **Empirical:** categories such as health, religious and political beliefs and sexual orientation are given special protection in some data-protection laws (for example GDPR Article 9 in the EU); India's Digital Personal Data Protection Act, 2023 governs digital personal data more generally. Laws apply to organisations differently from individuals; check current rules with a qualified adviser.

WORKFLOW

Publication approval in two stages

The approval stamp is a ritual, and rituals work by being boring and consistent. The exhibition's version deliberately cannot publish: approval and publication are separate acts, so the moment of pressing "publish" is never also the moment of deciding.

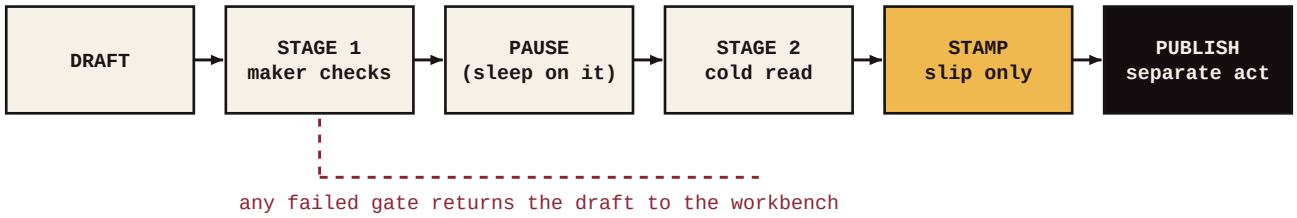


Figure 3. Six-step approval flow from draft to publication. Any failed gate in Stage 1 or Stage 2 returns the draft to the workbench. Publishing happens in a separate tool after the stamp.

Stage 1 · the maker checks	Question to answer in writing
Consent	Who is identifiable? Did each agree to this use, here, now?
Accuracy	Which claims did I re-check, against what?
Verifiability	Can a reader tell evidence from opinion from fiction?
Accessibility	Alt text, contrast, reading order, captions done?

Stage 2 · the cold read (after at least one night)	Question to answer in writing
Most-affected reader	Read it as the person in it with the most to lose. What changes?
Hidden channels	File names, metadata, comments, alt text, image corners, version history?
One-sentence reason	Why should this be public? If the sentence is weak, wait.

Inference. The pause is the most valuable step and the one most often skipped. Many publishing regrets arise in the gap between how a piece felt to write and how it reads to someone who was not in the room.

PROVENANCE

Origin card templates

A provenance note tells readers how a work came to be, so they can weigh it. It should be transparent about *method* and can be completely silent about the *person*. Where a field is omitted, say so: "omitted" is honest; "none" when there were some is not.

Empirical. The Coalition for Content Provenance and Authenticity (C2PA) publishes a technical specification for cryptographically signed provenance metadata ("Content Credentials") attached to media files. The cards below are hand-written and unsigned; they build reader trust through specificity, not cryptography.

Template A · short (for posts and captions)

> Made by [name]. Sources: [types, not private details]. Tools: [specific tasks]. Changed since first publication: [list or "nothing yet"]. Corrections: [public contact].

Template B · full origin card (fill in)

Field	Your entry	Show?
Title		[ ]
Authorised creator		[ ]
Publish date		[ ]
Tool contribution (be specific)		[ ]
Sources (types and public ones)		[ ]
Revisions		[ ]
Open questions		[ ]
Licensed assets		[ ]
Corrections contact (public)		[ ]
Deliberately omitted (list)		[ ]

Filled example (invented)

The Tidewater Almanac, Issue 1. Authorised by Ira Vell, 14 Sept 2026. Sources: 12 interviews (2 anonymised), town council minutes 1908-1915. Tools: AI-assisted transcription and copy-edit suggestions; all text written and checked by the author. Revisions: 2nd printing corrected a street name and captioned the map "simplified". Open: harbour relocation year, 1911 or 1913. Omitted: printing costs, personal circumstances.

ANALYTICS LITERACY

Thermometer, not diagnosis

Invented scenario. Twelve weeks of a fictional creator's numbers. Each line is scaled to its own maximum so their shapes can be compared; the absolute numbers are very different.

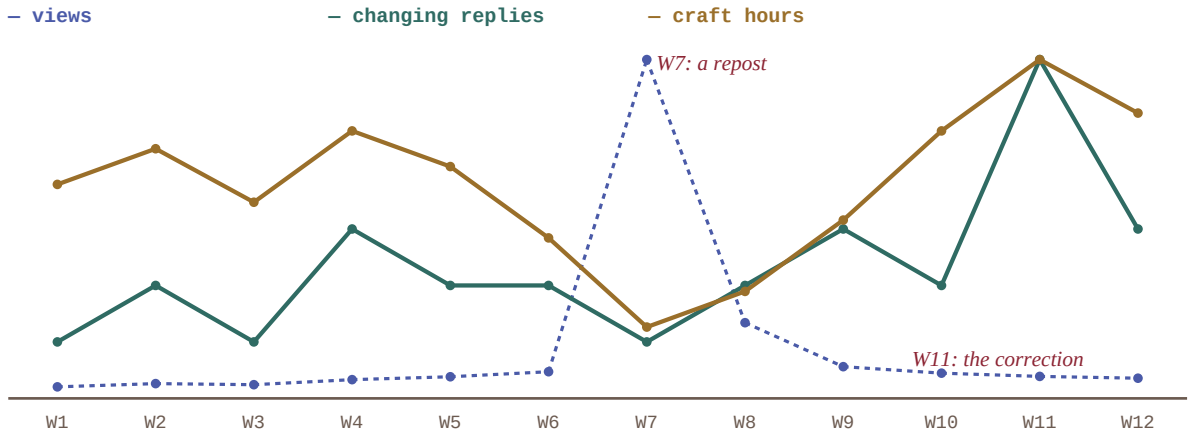


Figure 4. Views (dashed) spike to 9,400 in week 7 after a repost, while replies that changed the work stay at 1 and craft hours fall to 4. In week 11, views are 610, a reader's correction leads to the biggest revision (6 changing replies) and craft hours peak at 19. Data are invented.

Empirical. Campbell (1979) warned that quantitative indicators used for decisions become subject to corruption pressures and distort the processes they monitor. Strathern (1997) gave Goodhart's law its popular form: when a measure becomes a target, it ceases to be a good measure. Muller (2018) gathers cases across sectors.

Exercises

1. Which week would a dashboard celebrate? Which week improved the work most? Write one sentence on why they differ.
2. List three things that happened in week 11 that no analytics tool could record.
3. Choose one number you check often. What decision has it actually changed in the last month? If none, consider checking it less.
4. Design one "slow metric" you could record by hand weekly (e.g. corrections received, conversations that changed your mind).

Your answers

THREAT MODEL

Notes for a one-person studio

A threat model asks: what am I protecting, from whom, and what is a proportionate response? For most independent creators the realistic threats are mundane: their own haste, context collapse, an angry acquaintance, opportunistic scraping. This table is a starting point, not an assessment of your situation.

Asset	Plausible threat	Proportionate mitigation
Home location	Inferred from photos, routines, delivery posts	Delay location posts; check backgrounds and metadata
Family details	You, in a warm moment, oversharing	Rings model; the cold read
Accounts	Reused passwords, phishing	Password manager; two-factor; separate admin accounts
Drafts	Accidentally public links or repositories	Review sharing settings; keep drafts out of public folders and deploy directories
Private dashboards	Screenshots in "how I work" posts	Use invented demo data; never screenshot live dashboards
Browser storage	Shared device; malicious script	Treat localStorage as unencrypted; clear on shared devices
Text pasted into AI tools	Disclosure to a third-party service	Paste only what you would be willing to send to that company
Interviewees	Re-identification from quotes	Change identifying details with their agreement; let them check

Build hygiene for static sites

Before deploying any site: search the build folder for private hostnames, API keys, e-mail addresses and personal names; check source maps and HTML comments; confirm no private exports were ever copied into the project. The Shadow Studio build includes such a scan, and its security policy forbids scripts from making network connections at all (`connect-src 'none'`).

AI COLLABORATION

A disclosure ladder

Observation. Readers increasingly ask what a tool contributed to a work. "Made with AI" says too little; a full prompt log says too much and may leak private context. The useful disclosure names tasks, parts and checks.

0	No tool beyond ordinary editing software	`nothing to disclose`
1	Mechanical help: transcription, spellcheck, formatting	`optional note`
2	Research or summarising help; author verified every claim	`disclose task + check`
3	Drafted passages or code; author rewrote and tested	`disclose which parts`
4	Generated images, audio or voice	`label at point of use`
5	Output published largely as generated	`prominent label; reconsider`

Figure 5. Six rungs of tool involvement, from none (0) to output published largely as generated (5), with a suggested level of disclosure for each. This ladder is the author's proposal, not an industry standard.

Two ethics, not one

Output ethics: be specific about what the tool did. **Input ethics:** what you paste into a tool is itself a disclosure to its operator. A client's records, a friend's messages or a family member's medical letter do not become yours to share because the destination is a chat box.

> A workable line: disclose any tool whose output a reader might reasonably mistake for your own judgement, research or voice.

ACTIVITIES

Two worksheets

Worksheet 1 · Before you post (10 minutes)

Take one piece you are about to publish. Fill each line in writing; do not do this in your head.

Prompt	Answer
Who is identifiable, directly or by description?	
Which ring does each risky element sit in?	
What did automated checks miss when I looked by hand?	
Who is the most-affected reader, and what would they change?	
One sentence: why should this be public?	
When will I do the cold read?	

Worksheet 2 · The not-publishing log

Choosing not to publish is an editorial decision. Record it so it is a decision, not a drift. Revisit entries on the date you set.

Date	Item (described safely)	Why it stays offstage	Revisit on

Group activity (30 minutes, 2-6 people)

Each person fills the matrix on page 8 alone. Then compare one row at a time, starting with "Readers' private messages". The goal is not consensus; it is for each person to name the assumption behind their tick.

LIMITS

Objections, uncertainty, what this guide is not

Counterarguments worth taking seriously

Objection	Response
"Building in public" keeps creators accountable and helps others learn.	True for many. The guide argues for deliberate sharing of process, not against sharing it.
Withholding can protect the powerful and silence people who were harmed.	The hardest case, and unresolved here. Withholding to spare yourself and withholding to protect someone vulnerable look alike from outside.
Sometimes the personal is the method: illness shaped the work.	Then the creator may choose to share it, at a level they choose. The default stays private.
Numbers are the only feedback many creators get.	Agreed: use them as thermometers. The error is treating reach as impact.
Too much caution produces bland work.	Caution here is about other people and your own safety, not about ideas. Bold claims can be fully public.

Epistemic limits

The rings, decision tree, disclosure ladder and approval workflow are **authored models**, offered as decision aids. They have not been tested as interventions and no claim is made that using them reduces harm. The empirical claims cited come from specific studies in specific settings: Bernstein's (2012) finding that shielded factory workers experimented more, for example, comes from one factory and does not establish anything about creators. The redaction rules are heuristics with known false negatives; they will miss things.

What this guide is not

Not legal advice; not a security audit; not guidance for journalism, whistleblowing or research ethics review, each of which has established frameworks of its own; not a substitute for asking the people in your work what they want.

SOURCES

References and further reading

Bernstein, E. S. (2012). The transparency paradox: A role for privacy in organizational learning and operational control. *Administrative Science Quarterly*, 57(2), 181-216.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67-90.

Coalition for Content Provenance and Authenticity (C2PA). Technical specifications. https://c2pa.org/specifications/

European Union. (2016). General Data Protection Regulation (EU) 2016/679, Article 9: Processing of special categories of personal data.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday Anchor.

Government of India. (2023). The Digital Personal Data Protection Act, 2023 (No. 22 of 2023).

Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114-133.

Muller, J. Z. (2018). *The Tyranny of Metrics*. Princeton: Princeton University Press.

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. *Proceedings of the IEEE Symposium on Security and Privacy*, 111-125.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), 119-157.

Strathern, M. (1997). "Improving ratings": Audit in the British university system. *European Review*, 5(3), 305-321.

Sweeney, L. (2000). *Simple Demographics Often Identify People Uniquely*. Data Privacy Working Paper 3, Carnegie Mellon University.

Related public work by the author

Offline (offline.sahiltayal.com) · The Archive of Attempts (attempts.sahiltayal.com) · The Responsible Builder (thebuilder.sahiltayal.com) · Am I Successful? (scoreboard.sahiltayal.com). Each was reachable on 28 October 2026.

Colophon

Shadow Studio Field Guide, version 1.0, 28 October 2026. Written and edited by Sahil Tayal. Interactive companion: shadowstudio.sahilverse.online (preview; production address pending approval). More work: sahiltayal.com. Corrections welcome at sahil@sahiltayal.com. Set in Times and Courier. All examples invented; any resemblance to real people is unintended.