

the
**ResearchOps
Review**

RESEARCH

PRIVACY & ETHICS

IN THE AGE OF AI

A Ten-Point Diagnostic
Worksheet

SPONSORED BY

RALLY

As generative AI becomes embedded in enterprise tooling, research operations face a new set of governance challenges. Before generative AI, organisations typically relied on network defences, access controls, encryption, and secure infrastructure, supplemented in some cases by ethics policies or playbooks. These measures still matter, but generative AI creates and amplifies risks they were not designed to address, while established guidance for managing those risks remains limited.

If you build, manage, or govern research systems, this worksheet will help you assess your organisation's current maturity, identify systemic vulnerabilities in data governance and research ethics, and establish an operational roadmap for responsible AI-augmented research.

Who This Worksheet Is For

Created specifically for the user experience (UX) research profession, the worksheet's principles apply across qualitative and quantitative research more broadly. The worksheet is relevant to research functions of any size—from small teams to enterprise operations—that wish to establish system-level safeguards rather than rely on individual, ad hoc interventions. It will help you:

- assess current maturity across ten areas;
- identify gaps in both privacy controls and research ethics;
- prepare for a productive discussion with legal, information security, and procurement professionals; and
- distinguish between individual responsibility and system-level responsibility.

This worksheet is designed to help you make a meaningful start on a complex and evolving area of practice. Note that it is a diagnostic and meeting-preparation tool, not a compliance certification or a substitute for formal legal, security, or ethics reviews.

How to Use It

To maximise the strategic value of this framework, you should follow a structured, collaborative assessment process:

- 1. Complete a systems-level pre-assessment.**
Review the ten criteria across the systems, vendors, workflows, and data flows used in your research practice. Assign provisional maturity levels based on the evidence available, and record unknowns, assumptions, and questions for follow-up.
- 2. Convene a cross-functional group.**
Bring together representatives from research, ResearchOps, procurement, information security, legal, and relevant data or system-owner teams. Use the pre-assessment to focus the discussion on areas of uncertainty or concern.
- 3. Confirm the maturity scores.**
Validate each provisional score using evidence appropriate to the criterion, then compile the results into an overall diagnostic scorecard. Record any critical Level 0 findings separately.
- 4. Build a remediation roadmap.**
Use the standard operating protocols provided for low-scoring areas to assign responsibility, initiate process-design sprints, and track progress.

How the Scoring Works

Traditional compliance audits often rely on rigid yes/no checklists. While clean on paper, binary audits can fail to capture the operational realities of adopting AI tools. This worksheet takes a different approach, using ten criteria to help cross-functional teams surface systemic vulnerabilities, prepare for conversations, and create a prioritised to-do list.

After completing the worksheet, turn to the scorecard page to calculate your total, interpret your result, and record the areas that need attention first. Each of the ten challenges is scored from 0 to 3, giving a possible total of 0 to 30 points.

0–20 points | Critical Gaps

This score indicates material weaknesses in privacy, ethics, or operational controls. Escalate affected workflows to ResearchOps, legal, and information security professionals, pause specific data flows where exposure is suspected, and agree on a remediation plan.

21–26 points | Prioritise Improvements

The organisation has a foundation in place, but low-scoring areas need focused remediation. Assign owners, run targeted improvement work, and use human review for higher-risk AI outputs until the relevant controls are strengthened.

27–30 points | Operational Excellence

The organisation has a strong, evidenced socio-technical approach to responsible AI-augmented research. Document the assessment, maintain an audit trail, assign owners for ongoing maintenance, and review practices as systems, vendors, and AI use change.

Note. A high overall score doesn't override a critical weakness in an individual area. Any Level 0 finding involving participant consent, raw-data exposure, vendor processing, or unapproved AI use requires attention.

THE WORKSHEET

Use this workbook to assess the systems, vendors, workflows, and controls your organisation uses to manage research data and AI processing. It's designed to help you shift responsibility for ethical data management from individual practitioners to well-designed systems. Complete it at least annually and after any material change to the systems or data flows it covers.

01

02

03

04

05

Participant Data and Model Training

API Retention

Feature Updates

Raw Asset Access

Shadow AI

Why You Should Care

Some generative AI models are trained on large and varied datasets, which may include material scraped from the web. If a research tool's terms and settings don't clearly restrict the use of customer data, participant data may be processed, retained, or used for model improvement in ways the organisation and participants didn't expect. In some cases, this can allow sensitive participant insights to be memorised or inferred and reproduced in later model outputs, leaking them beyond the original research context.

Assessment Question

Do our active vendor contracts, data-processing agreements, and product configurations clearly define and restrict how participant data may be used, including for model training or improvement, and do we have evidence that these controls cover all approved integrations and subprocessors?

Many enterprise AI APIs retain prompts, uploaded data, and outputs, or associated safety and abuse-monitoring logs, for up to thirty days by default. The scope and enforceability of that retention period depend on the provider, product, endpoint, account type, contract, and configuration. Deleting a project folder in a research tool may therefore leave participant data in a separate third-party system for longer than the research team expects. Where zero data retention isn't available, the exception and any compensating controls need to be understood and governed.

Do our contracts and vendor settings confirm whether each external API retains participant data for up to thirty days by default, whether zero data retention applies, and how deletion of prompts, uploads, outputs, and safety logs is verified?

Cloud research tools can add AI features, change data flows, or introduce new subprocessors after procurement. If an organisation reviews a tool only at purchase or renewal, a new AI feature may be enabled before its privacy, security, or research-ethics implications are assessed.

Do our vendor-management and change-control processes identify new AI features, data flows, or subprocessors before they process participant data, and can we disable or restrict unapproved features while they're being reviewed?

Research repositories may contain recordings with identifiable voices, unredacted transcripts, and other raw assets that reveal participant identities or sensitive context. If users can download these files into personal AI tools or generic organisational tools that haven't been vetted for research data, they may create data flows outside approved purpose, retention, access, and consent controls.

Do repository permissions, download controls, and endpoint protections restrict access to raw research assets to authorised users and prevent transfer to AI tools not approved for the relevant research data and use case, while giving other stakeholders access to approved, appropriately aggregated insights?

Employees with legitimate access to research systems may still use personal or unmanaged AI accounts on work devices and paste raw participant data or survey files into them. If those services aren't approved for research use, the data may fall outside the organisation's controls for retention, access, vendor processing, and model use.

Do our network, identity, and endpoint controls prevent staff from using personal or unmanaged AI accounts to process research data on corporate devices, while providing an approved alternative for legitimate work?

Level 0
Unmitigated

Score
0

Level 1
Basic

Score
1

Level 2
Managed

Score
2

Level 3
Optimised

Score
3

Your Score

Level 0
Unmitigated

Score
0

Level 1
Basic

Score
1

Level 2
Managed

Score
2

Level 3
Optimised

Score
3

Level 0
Unmitigated

Score
0

Level 1
Basic

Score
1

Level 2
Managed

Score
2

Level 3
Optimised

Score
3

Level 0
Unmitigated

Score
0

Level 1
Basic

Score
1

Level 2
Managed

Score
2

Level 3
Optimised

Score
3

Level 0
Unmitigated

Score
0

Level 1
Basic

Score
1

Level 2
Managed

Score
2

Level 3
Optimised

Score
3



THE WORKSHEET

06

07

08

09

10

Consent and Erasure

Reidentification

Preserving Participants' Voices

Equitable Participation

Inclusive Recruitment

Why You Should Care

Traditional consent and data-management practices assume that research data can be located as discrete records and deleted at the source. Legacy research data may later be added to AI tools or repositories for purposes that weren't covered by the original consent. Regardless of when the data were collected, AI systems may retain information from them in ways that aren't easy to find or remove, even after the source record is deleted, particularly where the data have been used for model training or persistent memory.

Assessment Question

Do our consent and data-governance processes explain how AI will be used with new data, assess whether proposed AI uses of legacy data are compatible with the original purpose and participant expectations, and make clear what withdrawal or deletion can and can't achieve, while recording choices about distinct uses where feasible?

Research shows that LLM-based agents may reidentify some pseudonymous profiles by combining distinctive transcript details with public information. Other research artefacts may contain biometric data, including voices and faces, or other identifying information, such as names and on-screen details. In the age of AI, removing direct identifiers doesn't guarantee that data cannot be reidentified.

Do our data classification and handling policies treat pseudonymous data as potentially identifiable and require proportionate safeguards before AI tools process it?

AI-generated synthesis can misrepresent participant experiences, flatten disagreement, erase minority perspectives, or produce fabricated or misattributed quotations. Without links to source evidence, researchers can't readily check, challenge, or correct those interpretations before they're shared.

Do our research systems link AI-generated claims, themes, and quotations to relevant source material, such as passages or timestamps, so evidence can be manually verified and contextualised before findings are shared?

Controls designed to prevent fraud or verify identity can create barriers for participants with varying accessibility, literacy, language, or technology needs. Weak screening can pollute samples, compromise data quality, and divert incentives, but over-aggressive screening can exclude legitimate participants, including people with accessibility or literacy needs, further reducing the representation of groups already underrepresented in research.

Do our recruitment and survey systems use controls to flag AI-generated, duplicate, or otherwise suspicious participation, while monitoring for disproportionate exclusion across participant groups and routing uncertain cases for review?

Traditional panel recruitment often matches potential participants using demographic and quota criteria. Some AI-enabled platforms now rank, filter, or match people using hard-to-inspect rules and patterns from past recruitment data. If those systems favour speed or repeat past patterns, they may leave some groups out of samples and make findings less representative of the people affected.

Do we monitor panel composition and recruitment outcomes across relevant participant groups, whether we source participants from a vendor panel or manage our own panel with vendor-supplied software, and take corrective action when representation falls short?

	Score	
Level 0 Unmitigated	Consent materials don't explain AI processing, third-party services, retention, future or secondary use, or what withdrawal and deletion can and can't achieve. The organisation may apply AI to legacy data without checking whether the original purpose or consent covers the use, and ownership is unclear.	0
Level 1 Basic	Consent materials mention AI in broad terms, but bundle different uses and don't explain relevant data flows, retention practices, vendors, or the limits of withdrawal and deletion. Teams don't consistently assess whether new AI uses of legacy data align with the original purpose or participant expectations; staff rely on generic wording and manual follow-up.	1
Level 2 Managed	Internal templates distinguish relevant AI uses. Participant-facing materials explain the uses, choices, and limits on withdrawal or deletion that matter to participants. Named owners link consent records to relevant data and document whether legacy data uses align with the original purpose and participant expectations, including safeguards and exclusions.	2
Level 3 Optimised	Consent materials are tested with participants to ensure explanations of AI use and available choices are clear. When workflows or data uses change, including new uses of legacy data, the organisation reviews relevant data-governance decisions. Traceability is maintained across connected systems, and withdrawal requests are coordinated where feasible.	3
Your Score		

	Score	
	People conducting research send raw data to AI tools without removing direct identifiers, or treat pseudonymous data as anonymous once obvious identifiers are removed. They don't assess whether people could be identified from the remaining content, and no reliable controls govern the processing.	0
	People conducting research remove obvious direct identifiers, such as names and locations, using manual or basic automated tools, but rely on individual judgement. They don't systematically assess distinctive details, available public information, linkability, or the AI tool's retention and deletion limits.	1
	Policies classify pseudonymous data as potentially identifiable, and the organisation uses that classification to set standard safeguards for AI processing. Approved tools, data-minimisation and redaction controls, access restrictions, and defined retention and deletion settings protect routine processing without requiring a fresh review for every use.	2
	The organisation applies a repeatable risk review to higher-risk data and significant changes in AI use. It assesses plausible Reidentification pathways, considers potential harms to vulnerable groups, records residual risk and known deletion limits, and adjusts redaction, access, or processing decisions when circumstances change.	3

	Score	
	People conducting research use AI tools to generate summaries, themes, or quotations with no way to link the outputs back to the original data. Findings may be shared without a reliable way to verify what participants said or how their experiences were interpreted.	0
	AI tools provide broad citations or links, but not to the specific passages, timestamps, or records needed for verification. People rely on manual searching and individual judgement, and no consistent review is required before findings are shared.	1
	Research systems provide source-linked citations for AI-generated insights, with links to relevant passages, timestamps, or records. Established workflows require people to review those links and confirm that quotations are verbatim and interpretations are supported before findings are shared.	2
	Teams use existing research-quality or governance reviews to sample AI-generated findings for source accuracy, missing context, misattributed quotations, and flattened disagreement. They record recurring problems and update tools, workflows, or review guidance when needed.	3

	Score	
	People can submit AI-generated or heavily AI-assisted responses to qualify for incentives, and there are no reliable controls to detect or manage cases where this affects eligibility or data quality. The organisation doesn't know how screening decisions affect sample composition.	0
	People doing research manually inspect responses for signs of AI generation, such as repeated phrasing, implausibly fast completion, or inconsistent answers, and flag or remove responses based on individual judgement. They don't record which groups are flagged or assess how exclusions affect representation.	1
	The organisation offers accessible alternatives for different literacy and technology needs. Because fraud-detection controls can be privacy-invasive, the organisation limits telemetry to what is necessary. It uses documented criteria to flag AI-generated, duplicate, or otherwise suspicious participation while recognising that AI assistance may support access, and routes uncertain cases for review.	2
	The organisation monitors false positives and differential exclusion, calibrates thresholds using reviewed samples, and assesses whether screening changes demographic representation. It records the reasons for flags or exclusions and updates controls as AI-assisted participation patterns change.	3

	Score	
	AI-enabled matching or screening influences who is invited or excluded, but the organisation doesn't know where AI is used, what data it relies on, or how it affects representation. It accepts vendor or software outputs without an audit or a defined owner.	0
	The organisation knows AI is involved and relies on vendor assurances or occasional manual checks of panel composition. It has little evidence about subgroup invitation, screening, or exclusion, and no consistent process for correcting AI-driven representation gaps.	1
	A named owner sets representation goals and monitors how AI matching and screening affect invitation, screening, and completion rates across relevant participant groups. The organisation uses targeted outreach, additional sources, or works with vendors to adjust AI settings when gaps emerge.	2
	The organisation routinely evaluates AI matching and screening for differential exclusion and changes in representation over time. It adjusts vendor settings, matching rules, sourcing, quotas, or human-review pathways when AI-driven outcomes reduce representation and documents the effects.	3



YOUR SCORE & WHAT IT MEANS

Once you've scored the ten criteria, add the results to calculate your total. Use the result to identify gaps and strengthen your data and ethics safeguards. A high overall score doesn't override a critical weakness in an individual area. Any Level 0 finding involving participant consent, raw-data exposure, vendor processing, or unapproved AI use requires attention.

Your total score: / 30

0–20 Points | Critical Gaps

This score indicates material weaknesses in privacy, ethics, or operational controls. Escalate affected workflows to ResearchOps, legal, and information security professionals, pause specific data flows where exposure is suspected, and agree on a remediation plan.

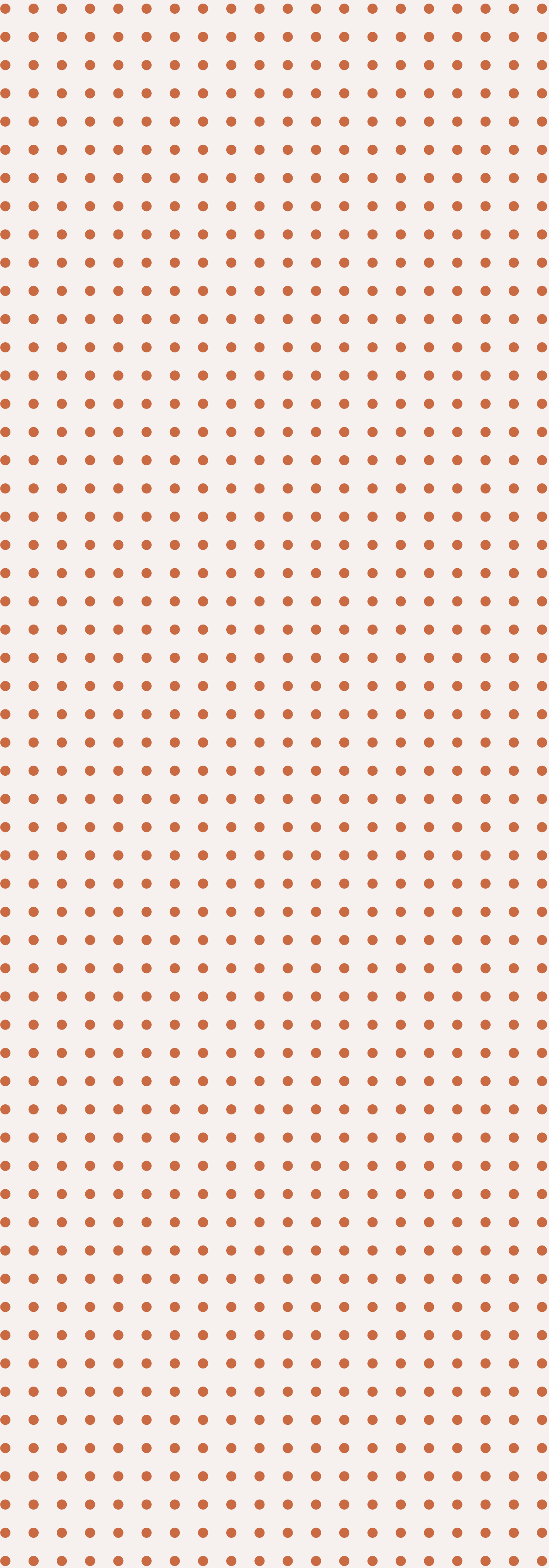
21–26 Points | Prioritise Improvements

The organisation has a foundation in place, but low-scoring areas need focused remediation. Assign owners, run targeted improvement work, and use human review for higher-risk AI outputs until the relevant controls are strengthened.

27–30 Points | Operational Excellence

The organisation has a strong, evidenced socio-technical approach to responsible AI-augmented research. Document the assessment, maintain an audit trail, assign owners for ongoing maintenance, and review practices as systems, vendors, and AI use change.

Notes



GLOSSARY & CONTRIBUTORS

AI-Assisted Participation

AI-assisted participation occurs when a participant uses an AI tool to help complete a survey, screener, or research activity. Whether this is acceptable depends on the study rules and its effect on data quality.

AI Model Improvement

AI model improvement is the use of data to evaluate, fine-tune, update, or otherwise improve an AI model or service.

AI Processing

AI processing is any use of an AI system to collect, transform, analyse, classify, summarise, generate, or otherwise act on research data.

Application Programming Interface (API)

An API is a structured way for one software system to send data to, and receive data from, another system.

Cloud Access Security Broker (CASB)

A CASB is a security service that helps an organisation monitor and control how people and systems access cloud applications and transfer data to them.

Data-Processing Agreement

A data-processing agreement is a contract that sets out how a service provider may process an organisation's data, including requirements for security, retention, deletion, subprocessors, and confidentiality.

Data Minimisation

Data minimisation is the practice of collecting and processing only the data necessary for a defined purpose.

Derived Representations

Derived representations are created from source data, such as embeddings, model parameters, summaries, classifications, or generated outputs. They may retain information or associations from the original data without preserving the original record in a directly retrievable form.

Differential Exclusion

Differential exclusion occurs when a screening or selection system excludes people from some participant groups at higher rates than others.

Pseudonymous Data

Pseudonymous data have had direct identifiers removed or replaced with codes but may still be linked to a person using additional information or distinctive details.

Reidentification

Reidentification is the process of linking pseudonymous, de-identified, or apparently anonymous data back to a specific person by combining remaining clues with other information.

Shadow AI

Shadow AI is the use of AI tools outside an organisation's approved systems, policies, contracts, and security controls, including personal or unmanaged accounts used for work purposes.

Subprocessor

A subprocessor is a third party engaged by a service provider to process data on its behalf, such as a company providing hosting, storage, moderation, or model infrastructure.

Tenant Restrictions

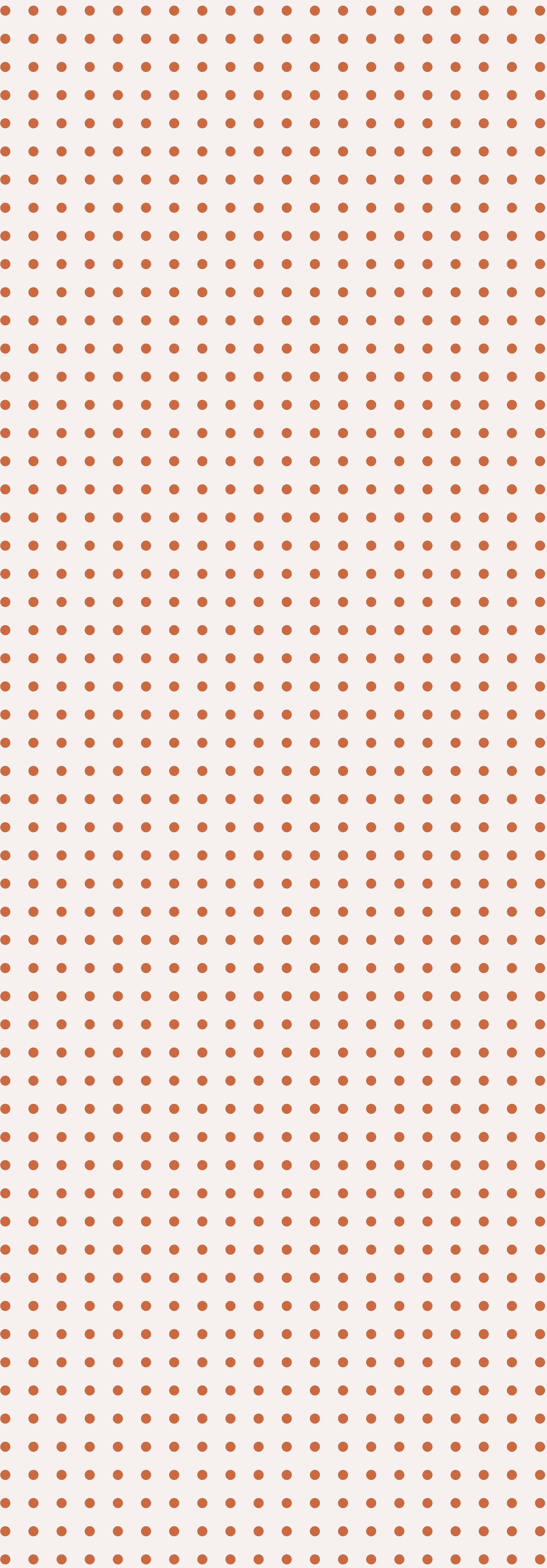
Tenant restrictions are technical controls that limit access to approved organisational accounts or environments, preventing work devices from connecting to personal or unmanaged instances of a service.

Zero Data Retention

Zero data retention is a service arrangement in which a provider doesn't retain specified inputs or outputs after processing. Its scope, eligibility, exclusions, and technical enforcement vary by provider and product.

Contributors

Thank you, everyone who shared their time, experience, and perspectives: Heather Ashley, Adalida Baca, Emma Baird, Kristina Bardavelidze, Stacy Bebie-Laiderman, Jordan Brinkman, Georgiana Livia Cruceanu, Alice Del Fabro, Arturo Diaz, Anna Drennan, Addison English, Michelle Gachette, Graham Gardner, Leila Hogarth, Emma Horrell, Boshra Javaheri, Alyssa Kreer, Austen Lazarus, Anne Lawrence, Colette Li, Mei Lie, Yuliya Martinavichene, Liza Meckler, Mads Monsen, Hyejin Nah, Romana Rajput, Nitya Rao, Lilian Rega, Allison Robins, Lidia Sambito, Karline Segan, Sufyan Selamet, Lucy Sutton, Christina Tan, Tina Tan, Annemieke Van Ruiten, Emily Wever, Rachel Wigen-Toccalino, Morgan Williams, Ryan Wolff, Pilar Zaragoza, and Changying Zheng.



REFERENCES

Standards, Regulatory, and Governance Frameworks

Alderman, Joseph E., Joanne Palmer, Elinor Laws, Melissa D. McCradden, Johan Ordish, Marzyeh Ghassemi, Stephen R. Pfohl, et al. "Tackling Algorithmic Bias and Promoting Transparency in Health Datasets: The STANDING Together Consensus Recommendations." *The Lancet Digital Health* 7, no. 1 (January 2025): e64–e88. [https://doi.org/10.1016/S2589-7500\(24\)00224-3](https://doi.org/10.1016/S2589-7500(24)00224-3).

European Parliament and Council of the European Union. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. Official Journal of the European Union, 2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>.

International Organization for Standardization. *Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems*. ISO 9241-210:2019. Geneva: ISO, 2019. <https://www.iso.org/standard/77520.html>.

Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard, and WCG Clinical. *Framework for Review of Clinical Research Involving AI*. June 2025. <https://mrctcenter.org/resource/framework-for-review-of-clinical-research-involving-ai/>.

National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg, MD: U.S. Department of Commerce, January 2023. <https://doi.org/10.6028/NIST.AI.100-1>.

Schwartz, Reva, Apostol Vassilev, Kristen Greene, Lori Perine, Andrew Burt, and Patrick Hall. *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. NIST Special Publication 1270. Gaithersburg, MD: National Institute of Standards and Technology, March 2022. <https://doi.org/10.6028/NIST.SP.1270>.

Academic and Theoretical Foundations

Greene, Kristen K., Mary F. Theofanos, Craig Watson, Anne Andrews, and E. Barron. "Avoiding Past Mistakes in Unethical Human Subjects Research: Moving from Artificial Intelligence Principles to Practice." *IEEE Computer* 57, no. 2 (February 2024): 53–63. <https://doi.org/10.1109/MC.2023.3327653>.

Lermen, Simon, Daniel Paleka, Joshua Swanson, Michael Aerni, Nicholas Carlini, and Florian Tramèr. "Large-Scale Online Deanonymization with LLMs." *arXiv:2602.16800v2*, February 25, 2026. <https://arxiv.org/html/2602.16800v2>.

McCradden, Melissa D., James A. Anderson, and Randi Zlotnik Shaul. "Accountability in the Machine Learning Pipeline: The Critical Role of Research Ethics Oversight." *The American Journal of Bioethics* 20, no. 11 (2020): 40–42. <https://doi.org/10.1080/15265161.2020.1820111>.

Nissenbaum, Helen. "Privacy as Contextual Integrity." *Washington Law Review* 79, no. 1 (2004): 119–58. <https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10>.

Sweeney, Latanya. "k-Anonymity: A Model for Protecting Privacy." *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems* 10, no. 5 (2002): 557–70. <https://doi.org/10.1142/S0218488502001648>.

UX Research and Practitioner Literature

Galanis, Jesse. "The Future of UX Research: Navigating AI-Powered Tools and Methods." *UXmatters*, May 4, 2026. <https://www.uxmatters.com/mt/archives/2026/05/the-future-of-ux-research-navigating-ai-powered-tools-and-methods.php>.

Liu, Feifei, and Kate Moran. "AI-Powered Tools for UX Research in 2023: Issues and Limitations." Nielsen Norman Group, July 2, 2023. <https://www.nngroup.com/articles/ai-powered-tools-limitations/>.

Mollick, Ethan. *Co-Intelligence: Living and Working with AI*. New York: Portfolio, 2024.

Russell, Daniel M. "The Challenges of Synthetic Users in UX Research." *ACM Interactions* 33, no. 1 (January–February 2026): 7. <https://doi.org/10.1145/3779007>.

Stokes, Thomas, and Lawton Pybus. "The PAIRED Framework for AI in UX Research." *Depth by Drill Bit Labs*, February 26, 2023. <https://depth.drillbitlabs.com/p/the-paired-framework-for-ai-in-ux>.

