The University of Virginia Archival AI Protocol

An AI training and access standard for archival organizations

Developed by Leo S. Lo
University of Virginia Library
Version 1.1, January 27, 2026

Short description

The UVA Archival AI Protocol (UVA AAIP) is a practical standard for how AI can and cannot use archival collections, built on a simple rule: No access without control. Irreversible models do not get access unless provenance and institutional control are real.

Definitions we've used in this protocol reflect terminology usage based on and adapted from the Society of American Archivists' *Dictionary of Archives Terminology* (https://dictionary.archivists.org/)

Archival collections
Holdings of an archival organization taken as a whole.

Archival organizations
Entities which hold physical and digital records, papers, artifacts, ephemera — a variety of material and formats — because they have enduring and ongoing value to society, democracy, culture, and individuals. Archival organizations can be standalone organizations, such as a state archives, historical society archives, or community archives; or part of a larger organization, such as a special collections library or archives department within a museum or business setting.

Irreversibility
The technical inability to remove specific source materials from a trained AI model.

Principle of reversibility
A principle borrowed from the conservation profession which states that archivists don't do what can't be undone now and in the future. If we can back out of a model, then we have reversibility.

Provenance
For this protocol, provenance refers to the ability to trace an AI output back to the specific archival items and collections used as evidence.

Part I. Purpose, scope, and core rule

Purpose

- The UVA Archival AI Protocol proposes how archival collections in any setting may be used in AI systems. It is designed to help archival organizations
    - Protect donors, communities, and institutional trust.
    - Avoid irreversible loss of control when AI models are trained on archival material.

  - Enable responsible use of AI for discovery, description, preservation, and research.
  - Negotiate AI partnerships from a clear, consistent position.

Scope

The Protocol applies to:

- Material owned by the organization through a deed of gift, purchase, or transfer through a scheduled records disposition, **including unprocessed and undigitized materials in any format**. Material on deposit is out of scope for this protocol. Material on deposit is out of scope for this protocol.
- Digitized or born digital collections held by an archive.
- Intellectual content such as finding aids, metadata, and detailed descriptions.
- Institutional repository content where the archival organization holds or manages the relevant rights.

The Protocol does not directly govern:

- Licensed third-party content where rights are controlled by publishers or aggregators.
- Public domain materials not governed by a donor agreement, community practices, or ethical restrictions, although access controls may still apply.
- AI use of general circulating collections, except where local policy chooses to extend similar rules.

Institutional benefit and mission fit

AI uses of archives must produce clear benefit to:

- The archival organization, its mission, and services.
- Researchers, scholars, students, or staff.
- The broader public, donors, or communities of origin.

Possible benefits include:

- Improved discovery, description, and preservation.
- Enhanced accessibility — for example, better transcription or description.
- High-quality research collaborations and outputs.
- Concrete support for digitization, infrastructure, or staff time.

Uses that mainly benefit external commercial products, with limited or unclear benefit to the organization and its communities, may be declined even if technically feasible.

Core rule of the UVA Archival AI Protocol

Because current AI model training is effectively irreversible, the Protocol is grounded in this core rule:

> **Irreversible models do not get access unless item level provenance and meaningful attribution can be demonstrated in practice, and the archival organization retains contractually enforceable control to stop further use.**

Appendix B defines the minimum expectations for provenance evidence.

By default, the Protocol supports:

- Retrieval-based AI services that keep source materials under the archival organization's control.
- Tightly scoped internal models that the archival organization can shut down or replace.

By default, the Protocol blocks:

- Training or fine tuning of broad commercial, general-purpose training, institutional foundation models using archival materials, where provenance and true institutional control cannot be guaranteed.
- Systems that absorb knowledge into model weights.
- Systems in which reversal is not possible and data is widely redistributed.

Part II. Foundational pillars

Pillar 1: Provenance and attribution

If the AI cannot cite its source, it cannot use the archival material. Where provenance and attribution cannot be maintained to the organization's standard, archival material will not be used for AI training or for public-facing AI services.

The archival organization will require:

- Mechanisms that record which items and collections are used for training, evaluation, or retrieval.
- Systems that can link AI outputs back to source items, or at minimum to specific archival holdings.
- Credit to the archival organization.

Pillar 2: Donor, community, and ethical responsibilities

The archival organization will honor all commitments made in deeds of gift, transfer documentation, and purchase agreements. Any memorandum of understanding or agreements with communities and partners will be upheld.

AI uses will not proceed where uses would:

- Violate deed of gift terms.
- Conflict with cultural protocols or community norms.
- Breach privacy expectations or ethical standards.
- Place vulnerable individuals or communities at risk.

Where obligations are unclear, the archival organization will decline AI uses that create unacceptable uncertainty and risk losing trust.

Pillar 3: Institutional control

Training large AI models transforms source materials into model parameters in ways that cannot realistically be reversed. Once a model has been trained, it is usually impossible to:

- Isolate the influence of specific source material.
- Remove that influence on demand.
- Verify complete removal through technical inspection.

To maintain institutional control, the archival organization asserts a “Right to Stop,” where the organization can order the cessation of materials ingest and demand the decommission or destruction of the model if permission is withdrawn.

The Protocol therefore gives priority to uses where the archival organization can still:

- Decide whether archival material is used.
- Change that decision later.
- See and influence where and how models that rely on its archival material are deployed.

Where this principle conflicts with short-term opportunities, control over archival collections takes precedence.

Transparency and accountability

The archival organization will:

- Maintain an internal AI project log for projects and tools that use archival materials, with detail proportional to risk and content sensitivity.

- Communicate key aspects of its AI stance to donors, community partners, and the public.
- Seek periodic review of AI agreements and projects through appropriate governance bodies.

Part III. Categories of AI use

The **UVA Archival AI Protocol** draws a sharp distinction between two families of AI uses.

- **Retrieval and controlled internal models**
  AI systems that retrieve from or analyze archival materials, where the organization can remove its data and shut down the system if needed.
- **Training or adapting general purpose models**
  AI systems that absorb knowledge from archival materials into model weights that are then used widely, where reversal is not realistic.

Internal organizational use

Internal projects may use AI with archival materials when all conditions below are met.

- The primary purpose supports discovery, description, preservation, or planning.
- The project uses either:
  - Retrieval-based access to digitized collections, or
  - Narrow models trained only or mainly on defined collections, which the organization can delete from the model or retrain if needed.
- Provenance and attribution are maintained to the organization's standard in Appendix B.
  - Appendix B is a minimum guideline during exploration and may be strengthened for public-facing or vendor-hosted uses.
- Any product derived through AI should identify that AI was used.
- Sensitive or restricted materials are handled according to existing policies.

Examples include:

- Assisted metadata creation and enhancement.
- Internal search, recommendation, or collection analysis tools.
- Preservation risk assessment and prioritization.
- Accessibility features such as descriptive text or transcripts, where appropriate.

Internal projects must be registered in the organization's AI project log and undergo periodic review by the Archives and AI Governance Group

Research and educational use

Academic researchers and educational partners may use archival material in AI-supported projects when:

- The primary aim is research, learning, knowledge creation, or public scholarship rather than commercial product development.
- A data management plan describing AI use is submitted to and approved by the organization's administrative body.
- Retrieval and controlled internal models are used.
- Any training or adaptation of models that will leave the organization's control or be widely redistributed is treated as commercial or external use and evaluated accordingly.

Tools or models that are released beyond a specific research context are subject to review and may require separate agreements.

Commercial and external model use

Commercial entities and external partners seeking to use archival materials in AI systems must enter into formal agreements that address:

- The materials in scope.
- Whether the proposed use is retrieval, training, evaluation, some combination, or something else entirely.
- Provenance, attribution, and logging requirements, at minimum those in Appendix B.
- The archival organization's control rights, including the ability to stop further use and, where possible, to decommission narrow models.
- Compensation or benefit sharing.
- Audit and compliance mechanisms.
- Security and data handling, including deletion or archiving of derived datasets.

As a default, training or fine tuning of general purpose commercial foundation models on archival collections are not permitted.

Part IV. Permitted and prohibited uses

Uses generally permitted in principle

The following uses are generally permitted, subject to rights, ethics, and technical review:

- **Retrieval based systems that search the collection at query time**
  - Source items remain stored and controlled by the archival organization.
  - Outputs must clearly and accurately cite the originating material and collections and may also hyperlink back to the originating material and collections.
- **Narrow models trained only or mainly on archival collections**
  - Used to power tools for discovery, description, teaching, or preservation.
  - Hosted under institutional control or under contracts that allow decommissioning on request.
- **Computational research over collections as data**
  - Uses retrieval or time limited training under archival organizational oversight.
    - Any model, embeddings, or derived dataset that leaves institutional control is treated as external use.
  - Does not release models as generic, widely reused AI services.

These uses require case-by-case review and must comply with the principles outlined in this protocol.

Prohibited by default are uses that:

- Train, fine-tune, or adapt general purpose AI models where item-level provenance and meaningful attribution cannot be maintained and demonstrated, per Appendix B.
- Hinder the organization from realistically requesting cessation of new uses of its content.
- Widely deploy, license, or repurpose materials beyond the specific project context.
- Violate deed of gift terms, community protocols, privacy agreements, or ethical standards.
- Enable surveillance, profiling, targeted harassment, or other harmful practices toward individuals or communities represented in the archival material.
- Misrepresent, obscure, or falsify the origins of materials,
- Generate content falsely attributed to the archival organization or its archives.
- Bypass, weaken, or contradict existing access controls or restrictions.

Exceptional approvals, if any, must be documented in detail, including a justification grounded in mission, ethics, and explicit risk mitigation.

Part V. Rights reserved by the archival organization

In any AI-related agreement that involves archival material, the archival organization reserves certain rights. Each right has a **baseline expectation**, which should apply in most agreements, and a **preferred level**, which the organization will seek where scope and leverage make it realistic.

Control over scope and collections

Baseline

- The organization retains the right to define which archival collections, subsets, or items are in scope for any AI use.
- The organization may exclude specific archival collections at its discretion, including but not limited to restricted donor collections, community-governed archives, confidential records, or sensitive materials.

Preferred when feasible

- For larger or multi-project partnerships, the organization may revisit and adjust scope as archival and institutional policies evolve.

Right to cease further use

Baseline

- The organization retains the right to require a partner to stop any further use of archival material in AI training, fine tuning, evaluation, or retrieval.
- Upon notice, the partner must:
    - Stop ingesting archival content into new training runs or models.
    - Stop using archival content in evaluation pipelines.
    - Remove archives from any retrieval or indexing services that the partner operates.

Right to decommission dependent models

Baseline

- For models trained solely or mainly on defined archival collections within a named joint project, the archival organization will seek the right to require decommissioning or destruction of those models if the organization withdraws permission, subject to reasonable notice.
- This right applies in particular to narrow models and tools clearly framed as built from the archival organization's materials. For example a model trained on a single archive for a specific research or access tool.

Preferred when feasible

- Where technical and operational conditions allow, the archival organization may seek similar decommission rights in other cases where models depend heavily on its collections.
- For large commercial foundation models that draw from many sources, the archival organization understands that full model destruction is unlikely to be granted. In such cases, the organization will instead rely on the rights that end new uses of archival materials and to remove its content from retrieval.

Right to audit and verify processes

Baseline

During this exploration period, the archival organization retains the right to verify partner compliance with contractual obligations relating to the archival organization's materials.

Verification will be reasonably scoped to the partner's use of the archival organization's materials and may rely on documentation, process descriptions, and available evidence, including:

- High level descriptions of relevant workflows and data flows — for example training, evaluation, or retrieval.
- Data handling, access controls, and security measures applicable to the archival organization's materials.
- Reports or evidence showing which archival materials were used and for what purpose, where feasible.
- How provenance, attribution, and access controls are implemented.

Written reports, attestations, and remote review are the default. On site review may be requested for higher risk uses or unresolved concerns.

Preferred when feasible

- The archival organization may rely on independent third-party audits or certifications, such as security or compliance reports, as part of verification.
- For projects that rely heavily on archival collections, the organization may request additional transparency regarding data flows and system behavior, subject to confidentiality protections.

Right to compensation and benefit sharing

Baseline

- The archival organization retains the right to seek financial compensation, cost recovery, or in-kind benefits for AI-related uses of its archival materials, especially in commercial or revenue generating contexts.
- Compensation may include fees, support for digitization, infrastructure, staff time, access to tools or services, or other contributions that strengthen the organization and its communities.

Preferred when feasible

- In projects where archival materials provide a significant competitive advantage or are a primary training source, the organization will seek more substantial benefit sharing. This may include revenue sharing arrangements, preferred or ongoing access to tools and models, and coauthored research outputs or other forms of scholarly recognition.
- The archival organization will develop internal guidance that distinguishes:
    - Mission-driven collaborations where in-kind benefits may be sufficient.
    - Commercial arrangements where financial or structured benefit sharing is expected.

Part VI. Implementation and governance

Phased implementation

The Protocol can be implemented in phases.

Phase 1: Immediate steps

- Adopt the core rule and make a short public statement.
- Require that all AI requests involving archival collections are routed through a simple review workflow.
- Add basic AI clauses to new deeds of gift and vendor agreements.
- Add a notice to institutional transfer documentation.

Phase 2: Near term enhancements

- Conduct a targeted review of high risk or high value archival collections and existing agreements.
- Create an internal AI project log for the archives.
- Form or assign a governance group to oversee AI uses of archival material.

Phase 3: Longer term development

- Invest in technical infrastructure for provenance, logging, and controlled retrieval.

- Develop more detailed fees and benefit sharing schedules.
- Integrate the Protocol into broader institutional AI readiness and ethics efforts.

Governance structure

The archival organization will assign governance responsibilities to an existing body or to a specific group, for example:

- An Archives and AI committee.
- A digital strategy or AI steering group with representation from archivists.

Key responsibilities include:

- Reviewing AI related requests that involve archival material.
- Applying the decision framework in Appendix A.
- Recommending approvals, conditions, or rejections.
- Reviewing and updating the Protocol at regular intervals.

Part VII. Stakeholder engagement

Donors, creators, records custodians, and communities

The archival organization will:

- Include AI language in advancement and curatorial discussions, transfers, and deeds of gift.
- Explain the core rule in accessible terms, for example:
  “We will not let your material disappear into AI systems we cannot control. We will only support AI projects that keep provenance clear and that we can change our mind about later.”
- Offer options for donors, creators, records custodians, and communities who wish to restrict AI uses more tightly.

Archival staff

The archival organization will:

- Train archivists and staff who work with archival material in the basics of the Protocol and the decision framework.
- Provide simple tools, such as intake forms and checklists, to route AI requests correctly.
- Encourage staff to surface concerns and new scenarios.

Institutional leadership and counsel

The archival organization will:

- Brief institutional leaders and legal counsel on the rationale behind the Protocol.
- Present the Protocol as a risk management, ethics, and leadership tool.
- Work with counsel to adapt sample clauses to local legal context.

External partners

The archival organization will:

- Share the Protocol with potential AI partners early in discussion to set expectations.
- Use the core rule and rights reserved as baselines in negotiation.
- Prefer partners willing to meet higher standards for provenance, transparency, and control.

Part VIII. Sample clauses and Public statement

Deed of Gift AI clause, standard

> The Donor/Creator grants to the Archives the right to use the archival materials, and any digital reproductions of them, in AI-supported tools and services that the archival organization controls, including search, description, preservation, and accessibility functions.
>
> The archival organization will not authorize the use of this collection to train or adapt general purpose AI models that cannot maintain item-level provenance and meaningful attribution, or that prevent the archival organization from exercising realistic control over continued use of those models.
>
> Any broader AI training use of this material will require a separate written agreement that specifies scope, conditions, and protections for the Donor/Creator and the archival organization.

Deed of Gift AI clause, restrictive

> The archival materials may not be used in any AI training, model development, or related computational process without the prior written consent of the Donor, Creator, or their representative. In the case of an institutional transfer of records, prior written consent of the administrative body is required.

The archival organization will not authorize the use of these materials in training general purpose AI models under any circumstances.

Any request for AI-related use must specify whether it is retrieval based or model training and must be approved in writing by both the Donor/Creator/Records Custodian (or representative) and the archival organization.

Vendor AI restriction clause

Vendors may not use any materials, images, metadata, or other content provided by or created for the archival organization to train, fine tune, or adapt general purpose AI models for Vendor's products or for services to third parties, unless a separate written AI training addendum is executed.

Any approved AI use of archival collections must either be retrieval based or involve narrow models that the organization can cause to be decommissioned if permission is withdrawn, as outlined in the Protocol.

Vendors will maintain documentation and evidence sufficient to verify scope of use, at minimum consistent with Appendix B, subject to content sensitivity and confidentiality.

Public statement template

Archives and AI at [Archival Organization Name]

[Archival organization] holds archival collections of historical, cultural, and scholarly importance and value. We are committed to making these collections available for research, teaching, and public engagement while protecting the rights and expectations of donors/creators/custodians, and communities.

Because AI model training is effectively irreversible, we follow the UVA Archival Artificial Intelligence Protocol, an AI training and access standard for archival organizations. We do not allow our archival materials to be used to train general purpose AI models that lack item-level provenance, meaningful attribution, or realistic institutional control.

We may support AI projects that use retrieval over our archival collections or that build controlled tools and models under our oversight, where these projects align with our mission and respect donor and community obligations.

All AI related uses of our archival material require prior authorization from the organization's administration. To discuss a potential project, please contact [unit or email].

Appendix A. Decision Framework for AI Requests Involving Archival Material

This decision framework can be implemented as a one-page flowchart or used as a checklist.

Step 1. Rights and ethics check

- Do we clearly hold the rights needed for the proposed use?
- Are there donor, creator, records custodian, community, or ethical constraints that would rule out this use?

If the answer is no or uncertain, decline or seek clarification before proceeding.

Step 2. Identify the type of AI use

Is the request mainly about:

- Retrieval or analysis of digital collections.
- Training, fine tuning, or adapting a model that will absorb knowledge from archival material into model weights.

For direct retrieval and analysis uses, proceed to Step 3.

For training or adaptation uses proceed to Step 4.

Step 3. Retrieval and controlled internal models

Confirm:

- The archival organization, or a trusted partner, can remove its data and shut down the system if needed.
- Outputs maintain clear links back to original items or collections.
- The project advances discovery, preservation, research, teaching, or community benefit.
- Sensitive or confidential materials are handled according to policy.

If all answers are yes, proceed in principle, then apply appropriate contractual and technical controls.

If any answer is no, revise the project or decline.

Step 4. Training or adapting general purpose models

Ask:

- Can the partner maintain item-level provenance and meaningful attribution in practice?
- Can the archival organization retain real control over continued use of its content?

If the answer to either is no, the default decision under the Protocol is not to allow use of the archival material.

If both appear possible, classify the project:

- Narrow model trained only or mainly on defined archival collections.
- Large mixed-corpus model where archival collections are one source among many.

For narrow models:

- Require strong decommission rights, detailed documentation, and appropriate compensation or benefit sharing.
- Evaluate mission fit and reputational risk.

For mixed-corpus models:

- Recognize that full model decommissioning is unlikely.
- At minimum, require:
    - No new training or evaluation runs that use organization's archival materials.
    - Removal of archival content from retrieval and indexes on request.
    - Clear documentation of how archival materials are used.
    - Appropriate benefit sharing.

If these conditions cannot be met in a way that governance and counsel accept, do not proceed.

Step 5. Final questions

Before agreeing, ensure that:

1. The use respects donors, creators, records custodians, communities, and ethical obligations.
2. The archival organization keeps real control, not just symbolic promises.
3. The project can be explained clearly to staff, donors, creators, custodians, and the public.
4. Benefits to the archival organization and its communities justify risks.
5. The decision and conditions are recorded in the AI project log.

If any answer is no, tighten conditions or decline.

Appendix B. Minimum Provenance Standard for Citations and Logs

1. Purpose

This standard defines the minimum requirements for provenance when an AI system retrieves and summarizes content from digitized archival collections. It applies to any tool that produces user-facing outputs and any internal workflow that generates descriptions, transcripts, metadata, or research summaries.

2. Core rule

If the system cannot produce the required citation fields, it should either decline to answer from archival sources or provide only the portions it can cite, especially for public facing outputs.

3. Definitions

1. **Source item:** The archival object used as evidence, for example a scan, photograph, letter, audiovisual file, or born digital file.
2. **Collection record:** The descriptive record that provides context, provenance, and arrangement.
3. **Citation:** A reference that allows a reader to locate the source item and verify the claim.
4. **Output:** Any text, transcript, caption, metadata, summary, finding aid content, or answer generated with AI assistance.

4. Minimum citation requirements for outputs

Every output that asserts facts, quotes, summarizes, or describes archival content must meet the requirements below.

1. **Granularity**
    a. Cite at the source item level whenever the claim depends on a specific item.
    b. Cite at the collection record level only when the claim is purely contextual and does not depend on a specific item.
    c. For multi-item synthesis, provide multiple item citations, not a single broad citation.
2. **Required fields in each citation**
    a. Collection ID and collection title
    b. Item ID, persistent identifier preferred

c. Pointer to the relevant part of the item, when available, for example page number, timestamp, frame, or internal segment ID tied to the retrieval chunks
d. Holding organization name
e. Access link, public URL when available, otherwise an internal resolver link for restricted materials

3. **Linking**
    a. When an item is publicly accessible, include a stable hyperlink to the item record or digital object.
    b. When restricted, include a stable internal link that staff can use to retrieve the item.
4. **Accuracy and faithfulness**
    a. Do not cite an item unless the cited item directly supports the statement.
    b. Do not cite a collection record to support a claim that requires item level evidence.
5. **Quotations and transcription**
    a. For direct quotes and transcriptions, cite the exact source item and the location within the item.
    b. If the transcription is uncertain, label uncertainty and avoid definitive claims.
6. **No citation behavior**
    a. If the system cannot produce the required citation fields, it must respond with one of the following:
        i. “Insufficient evidence in the retrieved archival sources to answer.”
        ii. A partial answer that includes only statements that can be cited, plus a clear note that other parts were withheld due to missing citations.

5. Minimum logging requirements

Keep enough information to support learning, troubleshooting, and basic accountability while workflows are still being developed.

What to capture when feasible

For each interaction, capture some or all of the following, depending on the tool and content sensitivity:

- Date and time
- System name and version
- User category, for example staff, researcher, public

- Query text, or a hashed query when sensitive
- Retrieved items, item IDs and collection IDs when available
- Citations shown to the user, when the system provides them
- Output text, when retention is permitted, otherwise note that output retention is not allowed

Content sensitivity and access

- Do not store restricted or sensitive content in logs unless permitted by policy and protected by access controls.
- Limit log access to staff with a clear need to review or troubleshoot.

Review and iteration

- Periodically review a small sample of interactions to identify citation gaps, restricted content risks, and recurring failure modes.
- As systems mature or move to public facing use, strengthen logging expectations in line with the archival organization's risk tolerance and governance review.