

AI Responsible Use Policy

September 10, 2026

What counts as AI for this policy

For the purposes of this policy, Artificial Intelligence (AI) is software that generates text, images, audio, code, or other content, or that makes predictions, recommendations, or decisions in ways that could substitute for human judgment or human authorship. This includes tools like ChatGPT, Claude, Copilot, and similar systems, whether standalone or integrated into other software, and any successor tools that perform the same functions regardless of what they are called.

Preamble

This policy neither prohibits nor endorses the use of AI. It applies to how you use these tools and what you are accountable for, not to the underlying technology or which specific tools exist. It applies whenever you are using AI tools for work purposes or with institutional data.

The Folger's communications and work products reflect human judgment and require human accountability. Three principles apply to everything in this policy, at every level, without exception: professional standards; authorship; and accountability.

Professional standards. Your work must meet the Folger's professional standards for your role, whether or not you use AI.

Authorship. The Folger is a humanities institution, which centers the work of human beings. Our work and our mission builds on their voices, their judgment, and their creative and intellectual effort. The role AI may play in authoring an artifact depends on the kind of artifact:

- **Construction.** Functional artifacts: reports, code, structured documentation (SOPs, checklists, etc.), charts and graphs. The value derives from accuracy and fitness for purpose. AI may build these with oversight from a person, who is accountable for accuracy and appropriateness.
- **Conversation.** Communications that address a person or audience as a person: emails, letters, messages, notes, statements addressed from someone or some part of the Folger to someone else. The value derives from the relational work the communication does between persons. Replacing person-to-person communication with machine-generated text is corrosive to the very humanities the Folger exists to strengthen and support. Therefore, the Folger does not send

communications from machines pretending to be people. AI may suggest edits, refinements, or corrections at the person's direction, but may not originate, choose words, or impose structure on a person's behalf.

- **Composition.** Creative and scholarly work: research and scholarship, educational content, curatorial interpretation, theatrical and artistic work. The value derives from the intellectual and creative effort of the people who produce it. The work originates from the people who create it and must be the product of their unmediated knowledge, judgment, and craft. AI involvement is limited to identifying mechanical issues (spelling, grammar, formatting) for the person to address.

Accountability. If you used AI to help produce something, any errors, fabrications, and omissions in it are yours. "AI produced it" is not an excuse.

Risk management

This policy manages three categories of risk that arise when AI is used in institutional work.

1. **Data and information security risk** is the risk that institutional data (donor records, personnel files, published and unpublished research, credentials, and similar) is exposed, retained, or used in ways the Folger has not authorized. **Level 1** addresses this risk.
2. **Operational risk** is the risk that AI output is wrong, incomplete, or applied in ways that cause harm to the Folger's work, its people, or the people it serves. AI errors often appear plausible and are difficult to detect without careful review and domain expertise. **Level 2** addresses this risk.
3. **Reputational and values risk** is the risk that a use of AI, if known, damages how scholars, donors, peer institutions, or the public regard the Folger's authenticity, legitimacy, or scholarly credibility. It also includes the risk that a use undermines the Folger's commitments to diversity, equity, inclusion, and accessibility. **Level 3** addresses this risk.

These risks are not mutually exclusive. The three levels are not steps in a sequence, and they are not a ranking of severity. They are three separate checks. A single use may require more than one, and each level applies on its own conditions regardless of the others.

Level 1: Secure the data

Before using any AI tool for work purposes, one question must be answered:

Is this tool on the Folger's approved list? If no, contact IT before proceeding.

All approved tools must meet the following requirements by default:

- The vendor may not train on submitted data. Exceptions require explicit institutional approval at the executive level.
- Data retention must be limited, explicitly defined, and appropriate to the sensitivity of the data. Indefinite retention is not permitted.
- The vendor may not access or review prompt content, except where required by security or legal obligations explicitly disclosed in the agreement.
- Access must be managed and auditable by IT.

Approval at **Level 1** confirms a tool is safe to use with institutional data. It does not authorize any specific use. **Level 2** and **Level 3** apply regardless of whether a tool is approved.

Sensitive institutional data – including donor information, personnel records, financial records, published and unpublished research, credentials and API keys, security and infrastructure logs, donor intent and restricted gift documentation, copyrighted texts held under license, and collection records containing non-public information – requires a formal agreement with the vendor confirming these requirements before use.

General internal data – internal correspondence, routine administrative records, and publicly available information the Folger has compiled – may be used with approved tools without a formal vendor agreement, provided the tool otherwise meets the requirements above. General internal data does not include correspondence or records that contain sensitive data types listed above. When in doubt, treat the data as sensitive.

A companion checklist defines the full minimum requirements for vendor approval.

Level 2: Assess the stakes

Department heads and domain experts determine whether a proposed use is appropriate for their area, based on evidence including: what does the Folger gain, what does it lose, what error rates are acceptable, and what expertise is required to evaluate the output.

The stakes of a use come from two ratings: **Impact** (how big the problem would be) and **Remediation** (how hard it would be to undo). Rate each one separately. Your stakes level is the more severe of the two ratings. We manage risk conservatively, so a severe rating on either one is enough.

1. **Impact.** How hard would the problem hit us?
 - **Feather.** Affects only your own work.
 - **Pebble.** Affects a small, bounded group of people or a specific set of work.
 - **2x4.** Affects the institution, the collection, or people outside the Folger.
2. **Remediation.** How hard is it to undo the problem?
 - **Hat.** Easily undone. The fix is immediate or low-cost.
 - **Haircut.** Fixable, but it takes time and you live with the consequences in the interim.

- **Tattoo.** Largely permanent. Reversing is costly, painful, or impossible.

Stakes. Your stakes level is the more severe of the two ratings.

- **Low.** Both axes are at the lowest level (feather and hat). Proceed on your own judgment, no notification needed. *Examples: AI-suggested phrasing in your working notes; using AI to summarize a public article for your own reference; grammar-checking your own document.*
- **Medium.** The higher axis is at the middle level (pebble or haircut). Inform your supervisor before or as you proceed. No formal sign-off required. *Examples: AI-drafted first pass of an internal report; AI transcription for an internal meeting where attendees are notified; AI-drafted training materials for a small team.*
- **High.** Either axis is at the top level (2x4 or tattoo). Get documented sign-off from your department head before proceeding. Record the tool used, data class, purpose, approver, date, and data retention setting in the central register. Validation methods must be appropriate to the domain and risk. *Examples: adopting handwriting text recognition for manuscript transcriptions; AI-generated text on a public-facing webpage; pre-screening fellowship applications with AI; deploying an AI chatbot responding to public or researcher inquiries.*

Higher stakes require proportionally higher confidence in accuracy, including verification of key facts, citations, and outputs before use.

A note on validation for scholarly and public-facing work. For High Stakes uses that produce scholarly outputs, catalog records, finding aids, or public-facing content, validation must include verification against original sources directly, not against AI-generated summaries or outputs. AI-generated citations must be independently verified against original sources before inclusion in any such output.

Level 3: Protect institutional integrity

Some uses carry implications that go beyond operational risk: they touch how the Folger is regarded by the people it serves and the communities it works with. Three questions identify when Level 3 applies. Between them they trigger two assessments: a reputational assessment and a bias and equity assessment. A use may trigger one assessment, both, or neither.

- Would a reasonable scholar, donor, peer institution, or member of the public, knowing about this use, revise their opinion of the Folger downward? Concerns here include the Folger's authenticity, legitimacy, and scholarly credibility, and extend to scholarly outputs, theatrical and artistic productions, collection stewardship, and communications with donors, fellows, and peer institutions. If yes or uncertain, the **reputational assessment** applies.
- Could this use affect decisions about individuals? This includes hiring, screening, evaluation, fellowship review, access decisions, and similar uses that change

outcomes for specific people. If yes or uncertain, the **bias and equity assessment** applies.

- Could this use introduce or amplify bias in how the Folger's collections, scholarship, or programs represent or serve different communities or groups? If yes or uncertain, the **bias and equity assessment** applies.

Reputational assessment. When triggered, escalate to the Executive Team for a decision. Consider how the use would be perceived by the constituencies it touches, what precedent it sets for the institution's relationship with those constituencies, and how it sits with the Folger's stated commitments and scholarly standards.

The Executive Team determines one of three outcomes:

1. **Proceed.** The use does not require disclosure to maintain institutional integrity.
2. **Proceed with disclosure.** The use is acceptable but requires transparent acknowledgment of AI involvement in the relevant output or context. Disclosure must be clear, proximate to the output, and in plain language.
3. **Do not proceed.** The use fails the integrity check regardless of disclosure.

The determination and its rationale must be documented.

Bias and equity assessment. When triggered, escalate to the Executive Team for a decision. Consider what is known and knowable about how the system works, its training data, and its decision logic.

The Executive Team determines one of three outcomes:

1. **Proceed.** The use does not present meaningful bias or equity risk.
2. **Proceed with conditions.** The use is acceptable but requires specific measures to mitigate bias or equity risk. Conditions must be documented and their implementation verified before the use proceeds at scale.
3. **Do not proceed.** The use presents bias or equity risks that cannot be adequately mitigated.

Both assessments must be proportionate to the visibility and consequences of the use, and the determination and its rationale must be documented in either case.

Staff training

All staff will complete AI literacy training once a year. New staff will also complete it when they are hired. Training will cover how AI systems work, their capabilities and limitations, algorithmic bias, training data provenance, and environmental considerations. This training is the assumed foundation for Low Stakes and Medium Stakes decisions. High

Stakes decisions require demonstrated familiarity with these concepts as a precondition for sign-off authority.

Evaluating new tools

The approved tools list is not fixed. If you want to use a tool not currently on the list, contact IT. IT will work with you to evaluate the tool against Level 1 requirements and, where appropriate, arrange a pilot use with non-sensitive data while the full evaluation is underway. When evaluating tools that are otherwise equivalent, the Folger prefers vendors with transparent policies on energy use and environmental impact.

Who this applies to

This policy applies to all staff, interns, and volunteers whenever they use AI for Folger work or with institutional data.

It also applies to anyone else, including contractors and vendors, in either of two cases: when their work would lead a reasonable person to attribute it to the Folger, or when they are handling proprietary institutional data. The first covers work presented under the Folger's name or in its voice, whether or not the Folger directed how it was made. The second covers non-public institutional data (donor records, internal imaging, operational data, and similar) as distinct from data the Folger has made publicly available.

Violations and consequences

Violations of this policy fall into two categories.

1. **Procedural violations** include failing to follow the assessment process, using an unapproved tool with general internal data, or failing to notify a supervisor for a Medium Stakes use. These are subject to progressive disciplinary action as described in the Folger Employee Handbook.
2. **Serious violations** include using an unapproved tool with sensitive institutional data, proceeding with a High Stakes use without documented sign-off, and presenting AI-generated creative or scholarly material as the product of human ideas, argument, and judgment when it is not. These may result in immediate disciplinary action up to and including termination, depending on severity and impact.

Violations should be reported to IT and the Director of Talent and Culture. The full disciplinary process is described in the Folger Employee Handbook.

Incident handling

If an AI-related error or data exposure occurs: report immediately to IT and your supervisor, stop using the tool, and contain the exposure.

Detailed incident procedures are maintained by IT.

Policy ownership and review

IT is the operational owner of this policy. The Executive Team is the decision authority for edge cases and **Level 3** escalations. This policy will be reviewed annually. Questions about specific uses go to IT for **Level 1** questions and your department head for everything else.