

Use of Large Language Models in GMP-Regulated Work

Document number	[SOP-0001]
Version	22Oct2026
Effective date	[DDMonYYYY]
Owner	[Role, e.g., Head of Quality]
Status	[Template]

Template notice (remove before use). This is a template, not advice. Your company must adapt it to its own products, systems and risks, and approve it through its own document control. The author accepts no liability for its use. Text shaded like this must be aligned with your own procedures before use.

License. [CC BY 4.0](#). Free to use and adapt. Attribution: *Joshua Radke, Radkeland Applied, radkeland.com*. No attribution is needed in your internal controlled copies; please credit the source if you publish or redistribute it.

Contents

Summary	3
I. Purpose	3
II. Scope	3
A. In scope	3
B. Out of scope	3
III. Principles	4
IV. Roles and Responsibilities	4
V. Definitions	5
VI. Procedure	5
A. Approved Tools	5
B. Classifying a Use	6
C. Data Handling	7
D. Reviewing and Verifying Output	7
E. Recording AI Assistance	7
F. Errors, Incidents and Deviations	8
G. Training	9
H. Periodic Review	9
VII. Records	9
VIII. References	10
IX. Revision History	11
Appendix A: Classification at a Glance	12
Appendix B: Use Cases, Good and Bad	12
Appendix C: Adaptation Guide (delete before use)	13

Summary

Large language models (LLMs) may assist GMP-related work, but they never own it. A qualified person reviews, verifies and takes responsibility for every AI-assisted output before it enters a GMP document or record, and AI never makes a GMP decision. Controlled data stays inside approved tools.

I. Purpose

Define how personnel may use LLMs and similar generative AI tools in work that touches GMP, so the company gets the productivity benefit without losing control of its quality decisions, records or data.

Why: In April 2026 FDA issued a warning letter to a drug manufacturer that used AI agents to write its specifications, procedures and master production records without quality unit review, and skipped process validation because the AI never mentioned it (Ref. 8). Nobody in that firm intended to break GMP. The tool was simply used without a procedure.

II. Scope

A. In scope

1. General-purpose LLM tools (chat assistants, writing assistants, AI features inside office software) used by any personnel for work that creates, informs or supports GMP documents, records or decisions.
2. Work outside GMP that uses company data, for the data handling rules in [VI.C](#) only. Refer to your Information Technology (IT) department policies for limitations.

B. Out of scope

1. AI or machine learning built into validated GMP computerized systems (for example, vision inspection, process control, predictive release tools). These are qualified and validated through the computerized system lifecycle, not this SOP.
2. AI used in drug discovery or clinical research, which falls under other procedures.

Why: A general-purpose LLM and a validated AI model are very different scopes. One is a drafting tool with no fixed behavior; the other is a system with a defined intended use and evidence that it performs. Mixing them in one procedure either over-controls the drafting tool or under-controls the system (Refs. 3, 9, 11).

III. Principles

Everything in [Section VI](#) traces back to one of these.

- A. **A person owns the output.** AI assists; a named, qualified person reviews, verifies and is accountable.
- B. **Use matches risk.** The more an incorrect output could affect product, patient or a GMP decision, the stronger the controls, up to not using AI at all.
- C. **AI never makes a GMP decision.** Release, disposition, investigation conclusions, specification acceptance and similar decisions remain solely with authorized people.
- D. **Controlled data stays controlled.** Confidential, proprietary and personal data go only into approved tools configured for it.
- E. **Use is traceable.** When AI materially contributed to a GMP document, the record says so.
- F. **Users are competent.** Nobody uses AI for GMP-related work without training on this SOP.

Why: These mirror the FDA and EMA joint principles for good AI practice (human-centric design, risk-based approach, clear context of use, data governance, documentation) translated into plain operating rules (Ref. 10). [III.A](#) and [III.C](#) restate long-standing quality unit responsibility under 21 CFR 211.22 (Ref. 1).

IV. Roles and Responsibilities

Role	Responsibility
User	Uses only approved tools; classifies each use (VI.B); follows data rules (VI.C); never presents unreviewed AI output as finished work.
Reviewer	A person qualified in the subject matter. Verifies AI-assisted content against first principles and its sources before it enters a GMP document or record (VI.D). The reviewer may be the user for Draft-tier work only where the document's normal approval chain still applies.
Process owner	Decides which uses are appropriate in their area; escalates unclear classifications to QA.
Quality Assurance (QA)	Approves this SOP and the approved tool list; resolves classification questions; reviews AI-related deviations; retains final approval of GMP documents per existing procedures.
System administrator	Configures approved tools per VI.A and maintains the approved tool list with QA.

Why: The warning letter in Ref. 8 states that AI output “must be reviewed and cleared by an authorized human representative” of the quality unit. Naming the reviewer role makes that a person, not a hope.

V. Definitions

Term	Meaning in this SOP
AI	In this SOP, shorthand for generative AI tools, most of which are built on large language models. Where the distinction matters, the SOP says LLM. Does not include AI built into validated systems (II.B).
AI-assisted content	Any text, table, summary or other output where AI generated or substantially changed the content. Spelling and grammar correction alone is not AI-assisted content.
Approved tool	An AI tool listed by QA and IT as permitted for company use, with its permitted data classes (VI.A).
Controlled data	Information classified as confidential, proprietary, or personal under the company’s data classification procedure, including batch data, customer and supplier details, and unreleased regulatory information.
LLM / generative AI	Software that produces text, code, tables or images in response to a prompt, based on statistical patterns rather than fixed logic.
Scratch, Draft, Decision	The three use tiers defined in VI.B.
Source	The authoritative material the content should reflect: the regulation, the data, the investigation notes, the current SOP.

VI. Procedure

A. Approved Tools

1. Use only tools on the approved tool list, within the data classes listed for each.
2. Approved tools are configured so company inputs are not used to train the vendor’s models and access is limited to company accounts.
3. Personal accounts and consumer versions of AI tools are not used for company work.
4. QA and IT review a tool before adding it, and when its vendor terms or major features change.

Why: The approved tool list is the single control that makes the data rules enforceable. Writing outcomes here (“inputs are not used for training”) rather than settings keeps this SOP true when the tools and their menus change.

B. Classifying a Use

Before using AI for a task, classify it into one of three tiers by asking what role the output will play.

Tier	The output is...	Examples	Controls
Scratch	Thinking material that never enters a GMP document or record.	Learning a topic, brainstorming, outlining, rewording an email.	Data rules (VI.C). No review record required.
Draft	Content that will become, or inform, a GMP document or record, but only through a person.	First draft of an SOP revision, deviation narrative from investigator notes, summary of a new guidance, training questions.	Data rules, full review against source (VI.D), normal document approval, AI assistance recorded (VI.E).
Decision	Something that would make or replace a GMP judgment, or generate GMP data, without a person in between.	Release or disposition, OOS invalidation, investigation conclusions, specification limits, issuing a master record.	Not permitted with general-purpose LLMs. Only through validated systems (out of scope, II.B).

1. When unsure, classify one tier higher and ask QA.
2. Severity, frequency and detectability decide borderline cases:
 - a. **Severity:** if this output were wrong and nobody caught it, what is the worst credible effect on product, patient or a GMP decision?
 - b. **Frequency:** how often will this use occur, and how often is the tool likely to get this kind of task wrong?
 - c. **Detectability:** how likely is the review to catch an error? Treat AI detectability as lower than it feels (see VI.D).
 - d. High severity with low detectability belongs in Decision, whatever the frequency.
3. A use that moves between tiers (for example, Scratch notes that end up pasted into a deviation) takes the controls of the higher tier at the moment it moves.

Why: The tiers sort uses by the role of the output, which is what drives risk. FDA's draft framework rates AI risk by "model influence" (how much the output drives the decision) and "decision consequence" (Ref. 9); the draft EU GMP Annex 22 bars generative AI from critical GMP applications while allowing non-critical use under qualified people (Ref. 11). Severity, frequency and detectability is the familiar ICH Q9 way to reason about the same thing (Ref.

12): severity is decision consequence, and detectability falls as model influence rises. The Ref. 8 firm effectively treated Decision-tier work as Scratch.

C. Data Handling

1. Controlled data goes only into approved tools permitted for that data class.
2. Never enter into any AI tool: passwords or credentials; patient or personal health information unless the tool is approved for it; material a customer or partner has restricted.
3. Remove or replace identifiers (customer, supplier, lot, product names) when the task does not need them.
4. Do not upload complete GMP records to Scratch-tier work.

Why: Once data leaves the company's control it cannot be recalled, and the loss may not be detectable at all. That makes data exposure a high-severity, low-detectability failure even when the task itself is low risk.

D. Reviewing and Verifying Output

For Draft-tier work, the reviewer:

1. Compares the content against its **source**, not against what seems reasonable.
2. Confirms every cited regulation, guidance, standard or reference exists and says what the content claims. AI tools can invent citations that look correct.
3. Checks numbers, units, limits and names character by character.
4. Looks for **omissions**: requirements, steps or caveats that should be present and are not. AI is notorious for not telling you what it left out.
5. When AI revised an existing document, compares old and new versions line by line so no requirement changed without intent. Where practical, have AI revise a tracked-changes copy so every change is visible to the reviewer.
6. Corrects or rejects the content. Accepting AI output is a decision the reviewer owns.

Why: AI output is fluent, and fluency triggers expectation bias: text that reads well is trusted more and checked less. Reviewing against the source breaks that bias. The omission check exists because of Ref. 8, where the firm skipped process validation partly because the AI agent "never told [them] it was required." 21 CFR 211.22(c) puts approval of procedures and specifications with the quality unit (Ref. 1), and 211.100 requires those procedures to be adequate (Ref. 4); neither is satisfied by a tool.

E. Recording AI Assistance

1. For Draft-tier use of AI, record: **"AI-assisted ([tool], [version, if shown]). Reviewed against source by [name], [date]."**
2. Put the statement in the record you are already completing:

When the AI-assisted content goes into...	Put the statement in...
A new or revised controlled document	The document change request (DCR)
A deviation or CAPA	The investigation report
An OOS or laboratory investigation	The investigation record
Training material	The training material's revision history
Anything else	The record's comment or notes field

[Adapt this table to your own quality management system.]

3. **Optional:** keep a company-wide **AI-Assistance Register**: an uncontrolled index listing each Draft-tier use by record type, record number, user initials, date and area. It points to the records; the records remain the evidence.
4. Keep the sources the reviewer verified against with the normal record.
5. Saving prompts and full AI transcripts is not required unless a procedure or investigation calls for it.

Why: An inspector's first question about AI will be "where did you use it, and who checked it?" A one-line entry answers that and supports traceability, which both the FDA and EMA principles call for (Ref. 10). Retaining every transcript adds volume without adding evidence; the reviewed document and its sources are the evidence. The finished document itself is a GMP record retained under 211.180 (Ref. 5), and Part 11 applies to it through the system it already lives in (Ref. 6). The optional register answers "where did you use AI?" in minutes, gets an inspector to the relevant records, and keeps the practice of evaluating AI use visible. It stays uncontrolled so it stays light, which is acceptable because it only points to controlled records.

F. Errors, Incidents and Deviations

1. If an AI-originated error is found in an approved GMP document or record, handle it through the deviation and CAPA system.
2. If controlled data was entered into a tool not approved for it, report it to QA and IT the same day or as soon as practicable.
3. Investigations do not stop at "the AI was wrong." Ask why the review did not catch it.

Why: "The AI got it wrong" is the new "human error": a description, not a root cause. The useful question is always about the review, the classification, or the procedure that let the error through.

G. Training

1. Personnel complete training on this SOP before using AI for any company work.
2. Training covers: the three tiers, the data rules, the review steps in [VI.D](#), and at least the use cases in [Appendix B](#).
3. Retraining follows significant revisions of this SOP, and any review under [VI.H.2](#) that changes how approved tools may be used, even when the SOP text does not change (for example, a tool that could only draft gains the ability to take actions).

Why: 21 CFR 211.25 requires personnel to have the training to do their assigned functions (Ref. 2). Using AI on GMP work is now one of those functions. Training also builds the habit the review depends on: distrusting fluent text until it is checked.

H. Periodic Review

1. QA reviews this SOP and the approved tool list at least every [\[12\]](#) months.
2. Review sooner when: a referenced guidance is finalized or changed (notably Refs. 9 and 11); an approved tool changes materially; or a deviation involves AI.

Why: Two of this SOP's key references are drafts, and AI tools change monthly. A fixed procedure for a moving technology turns into one of those documents that quietly stops matching reality.

VII. Records

Record	Where it lives
Approved tool list	Controlled document maintained by QA and IT
AI assistance entries (VI.E)	Within the affected GMP document or record
Training records	Training system
Deviations and data incidents	Deviation and CAPA system
AI-Assistance Register (optional)	Uncontrolled index maintained by QA; points to the records above

Records are retained per the company's record retention procedure and 21 CFR 211.180 (Ref. 5).

VIII. References

Each reference states where this SOP uses it. References marked **framework** inform the reasoning but are not applied directly.

#	Reference	Used in	How it's used
1	21 CFR 211.22, Responsibilities of quality control unit	III, IV, VI.D	Quality unit approves procedures and specifications; basis for human review and for AI never deciding.
2	21 CFR 211.25, Personnel qualifications	VI.G	Training before use.
3	21 CFR 211.68, Automatic, mechanical, and electronic equipment	II.B	Framework. Applies directly to validated computerized systems, which are out of scope; cited to mark the boundary.
4	21 CFR 211.100, Written procedures; deviations	VI.D	Procedures must be adequate regardless of who or what drafted them; cited in Ref. 8.
5	21 CFR 211.180, General requirements (records)	VI.E, VII	Retention of the finished documents and records.
6	21 CFR Part 11, Electronic records; electronic signatures	VI.E	Applies to AI-assisted documents through the controlled system they are stored in; AI tools themselves are not the record.
7	FD&C Act section 501(a)(2)(B)	I	Drugs not made under cGMP are adulterated; the legal basis cited in Ref. 8.
8	FDA Warning Letter to Purolea Cosmetics Lab, #320-26-58, 02Apr2026 (link)	I, IV, VI.B, VI.D	First cGMP warning letter about AI. AI-generated specifications, procedures and records without QU review; required process validation missed because the AI did not mention it; AI output “must be reviewed and cleared by an authorized human representative” of the QU.
9	FDA draft guidance, <i>Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products</i> , Jan 2025, docket FDA-2024-D-4689	II.B, VI.B, VI.H	Framework; draft, not for implementation. Its “model influence” and “decision consequence” risk concepts inform the tiers. Its scope is AI producing data for regulatory decisions, not drafting tools.

#	Reference	Used in	How it's used
10	FDA and EMA, <i>Guiding Principles of Good AI Practice in Drug Development</i> , 14Jan2026	III, VI.E	Section III translates its human-centric, risk-based, context-of-use, data governance and documentation principles.
11	EU GMP Annex 22, <i>Artificial Intelligence</i> , draft for consultation, Jul 2025	II.B, VI.B, VI.H	Framework; draft. Generative AI “should not be used in critical GMP applications”; non-critical use under qualified personnel. Basis for the Decision tier. Applies directly only to EU GMP.
12	ICH Q9(R1), <i>Quality Risk Management</i>	VI.B	Severity, probability (frequency) and detectability as the way to reason about borderline classifications.

Further reading (not used directly): FDA, *Computer Software Assurance for Production and Quality System Software* (final Sep 2025, updated Feb 2026), for risk-based assurance thinking; ISPE, *GAMP Guide: Artificial Intelligence* (2025), for validating AI inside GMP systems.

IX. Revision History

Version	Description	Author
22Oct2026	First issue of template.	Joshua Radke, Radkeland Applied

Appendix A: Classification at a Glance

1. **Will the output enter a GMP document or record, or inform a GMP decision?** No: **Scratch**. Yes: go on.
2. **Will a qualified person review it against its source before it's used?** Yes: **Draft**. No, or the output is the decision: **Decision**, not permitted.
3. **Unsure?** Go one tier up and ask QA.

Appendix B: Use Cases, Good and Bad

Theoretical examples for training. Each shows the tier and the lesson.

#	Situation	Tier	Verdict	Lesson
1	An investigator gives AI their interview notes and asks for a first-draft deviation narrative, then checks every statement against the notes and corrects two. Change history notes AI assistance.	Draft	Good	Source-based review, recorded.
2	A user pastes a full batch record, including customer names, into a consumer chatbot to “summarize what went wrong.”	Scratch, data rules broken	Bad	The task was low risk; the data exposure was not.
3	AI summarizes a new FDA guidance for leadership. The summary cites a CFR section that doesn't exist, and it goes out unchecked.	Draft	Bad	Verify every citation (VI.D.2). Fluent is not the same as true.
4	AI generates product specifications and a master production record, which are issued without QU review.	Decision	Bad	This is Ref. 8. Never permitted.
5	A team uses AI to brainstorm possible cause categories before a root cause session; the investigation itself is written by people from evidence.	Scratch	Good	AI widened the thinking; people did the investigating.
6	A user asks AI to “simplify” an SOP, likes the result, and routes it for approval. A required verification step was silently dropped.	Draft	Bad	Compare old and new line by line (VI.D.5). Look for omissions (VI.D.4).
7	An analyst asks AI whether an OOS result can be invalidated and records the answer as the rationale.	Decision	Bad	Investigation conclusions stay with people.
8	A trainer uses AI to write quiz questions for an SOP, then checks each answer against the SOP and fixes one wrong key.	Draft	Good	Low severity, but still reviewed against source.

Appendix C: Adaptation Guide (delete before use)

1. Replace bracketed placeholders: document number, owner, review interval.
2. Align the data classes in [VI.C](#) with your data classification procedure.
3. Create the approved tool list before this SOP becomes effective.
4. Rename roles to match your organization. Keep the reviewer distinct from “whoever happens to be nearby.”
5. If you change the tiers, keep the rule that AI never makes a GMP decision.
6. Before removing any control, read its **Why**. If the reason no longer applies to you, remove it on purpose and record why in your change control.
7. Check Refs. 9 and 11 for final versions; both were drafts at the time of writing.
8. After inserting, moving or removing sections or references, press **Ctrl+A**, then **F9** to update the cross-references and the table of contents.
9. Section numbers in the text are live fields; Ctrl+click follows them. Return to where you were with Alt+Left Arrow.