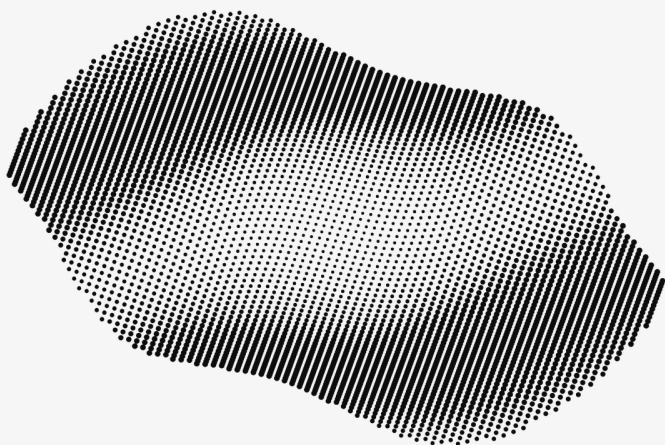




GUIDE 06 DOCUMENT ANALYSIS

Using LLMs to review case documents

How to turn a production into a text archive a language model can work on, ask questions that return sourced answers, and keep the human in charge of what is true.



ABOUT

A device production runs to tens of thousands of pages: design records, complaint files, logs, correspondence. A language model can read all of it in an afternoon. It cannot be trusted with any of it unless every answer points back to a page.

This guide describes a working method: how to build the archive, how to ask, which tasks repay the effort, where the models fail, and the rules that keep the output usable as the basis for expert work rather than a substitute for it. It names no products; the method is the same for any of them.

FOR

Legal counsel and their teams who must review a large production quickly and still vouch for every statement in a brief.

Manufacturers — complaint, quality, and regulatory teams with the same volume of records and the same duty of accuracy.

Hospitals — risk management and clinical engineering staff reviewing event files and device records.

NOTE

This guide is general technical information. It is not legal advice, and it does not address the rules of professional conduct, privilege, or protective orders that govern how a particular team may use these tools; counsel decides those first.

CONTENTS

In this guide

01	What a model can do, and cannot	04
	Figure: from production to sourced answer	05
02	Building the archive	06
03	Asking well	07
04	Three uses that pay	08
05	The limits	09
06	Working rules	10
	Checklist	12
	Sources	13
	How Alpha Devices helps	14
	Terms	15

SERIES This guide is one of fifteen practical guides from Alpha Devices. The full list is on the back cover; all are free at alphadevices.io/resources.html.

01

What a model can do, and cannot

A language model works on text. Given the text of a document, it can do a few things well and reliably; given a description of a document, it can only guess. The method in this guide rests on that distinction.

SUMMARISE	A document, a folder, or a thread, at a chosen length, with the page each point came from. Useful for triage, not for the final reading.
FIND	Every place in the record where a part number, a failure mode, a person, or a phrase appears, including passages that keyword search would miss because the wording differs.
EXTRACT	Dates, parties, part numbers, versions, and decisions into a table, each cell with its source. The raw material of a timeline (guide 15) or a request-response map (guide 05).
COMPARE	Two versions of a document, two complaint narratives, a protocol against its report: what changed, what is missing, and where they disagree.
DRAFT	Questions for a deposition, a request list, a first outline of a chronology, from the record and the guides, for an expert to correct.

NOTE

It cannot verify anything, know what should be in the record but is not, or be held responsible. Every output is a lead. The reading that makes it evidence is still done by a person.

FIGURE

From production to sourced answer

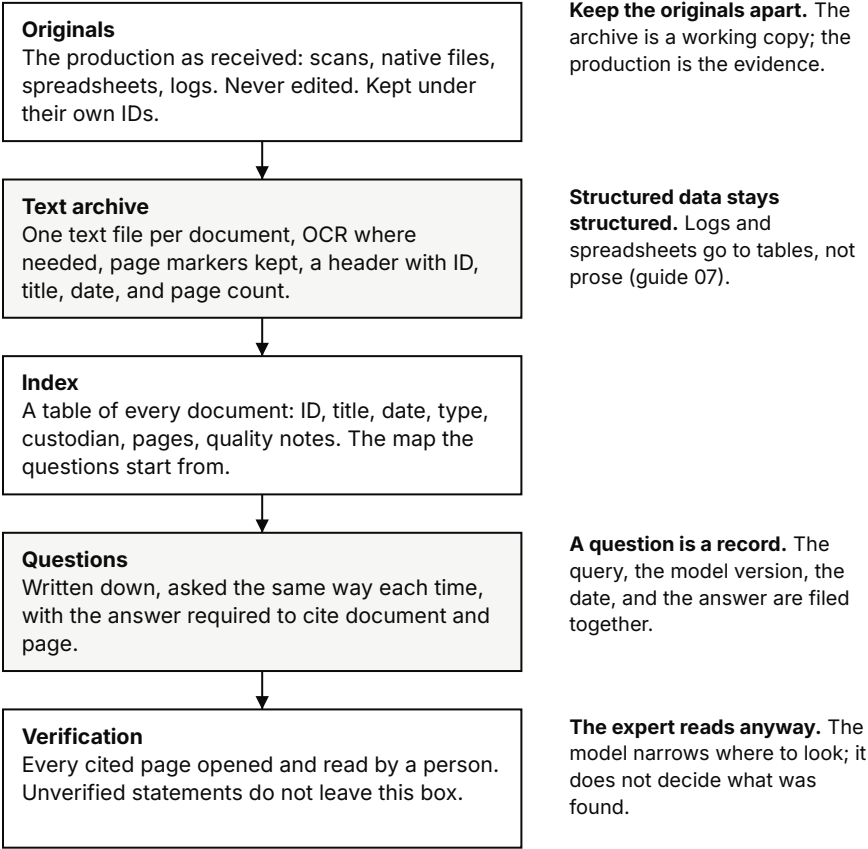


FIG 01 The archive is built once and queried many times. Most of the value is in the second and third boxes: a clean, indexed text archive makes every later question cheap and every answer traceable.

02 Building the archive

The archive is the part of the work that cannot be skipped. A model given a messy archive returns confident answers from the wrong page.

01 Inventory the production

Every file with its production ID, source, and format. Duplicates noted, not deleted. Families kept together: an email with its attachments, a report with its appendices.

02 Extract the text

Native files give text directly; scans need OCR. Keep a page marker in the text at every page boundary so a citation can name the page. Check a sample of OCR output against the image.

03 One file per document

Plain text, named by production ID, with a header: ID, title, date, author, type, pages, and a quality note (clean, OCR, partly illegible, redacted).

04 Build the index

A table from the headers. It is what the model is given first, and it is how a person checks that an answer drew on the right documents.

05 Keep data as data

Device logs, complaint databases, test results, and spreadsheets go into tables with their columns intact. A model summarising a spreadsheet in prose loses the numbers.

06 Record the gaps

Documents referenced but not produced, illegible pages, redactions. The model cannot see what is absent; the index can say so.

03

Asking well

The quality of an answer follows the question and what the model was given to answer it from. Four habits make the difference.

01

Give it the documents

Not a description of them. The relevant text files, or the archive with retrieval over it, are in front of the model when it answers. An answer from general knowledge of how complaint files usually read is worthless, and looks the same as a good one.

02

Require the source

Every statement in the answer carries the document ID and the page. An answer without a source is sent back. A source that does not say what the answer claims is a finding of the verification step, and it happens.

03

Ask narrow, ask again

"Summarise the CAPA file" returns a summary. "List every CAPA that names occlusion, with open date, close date, and root cause, as a table with sources" returns something that can be checked, and asked again when the production grows.

04

Ask what is missing

Given the request list (guide 05) and the index, which items have nothing in the production? Given the timeline, which months have no documents? The index makes this comparison possible.

KEY POINT

A question written down and asked the same way across the record is a method. A chat is not.

04

Three uses that pay

Most of the value in a device matter comes from three tasks that are large, repetitive, and exactly the kind of work a model does well under supervision.

01

The timeline

Every dated event in the record, extracted to a table: date, event, source document, page, and the stream it belongs to (design, market, field, unit; guide 15). The person then reads each source, corrects the entries, and resolves the conflicts between clocks. The model does the first pass over thousands of pages; the expert does the second over hundreds of rows.

02

Request against production

The request list from guide 05 on one side, the index on the other. For each requested item: what was produced, under which IDs, and what was not. The result is the schedule for the meet-and-confer, drawn from the record rather than from memory.

03

Reading the complaint file

Each complaint narrative classified by failure mode, component, and outcome, as the manufacturer should have classified it (guide 13), with the model's reasoning shown. The counts are then verified by reading, because a trend argued from a model's classification is a trend argued from a guess until someone reads the narratives.

NOTE

In each case the model's output is an intermediate: a table to be checked, not a conclusion to be cited.

05

The limits

These are the ways the method fails, and what each one means for the rules on the next page.

INVENTION	Models produce fluent statements with no basis, including citations to pages that do not say what is claimed, or do not exist. The only defence is reading every cited page.
OMISSION	A summary leaves things out, and the model does not know which things mattered. Summaries are for finding where to read, never for deciding what the record says.
SCALE	A model reads a bounded amount of text at once. Over a large production it works through retrieval, which finds what resembles the question and can miss what does not. Searches are repeated with different wording.
DISCLOSURE	Where the text goes, who can see it, whether it is retained or used for training, and what a protective order permits. Decided by counsel before a single page is uploaded; recorded with the archive.
DRIFT	Models change. The same question to a later version returns a different answer. The model, its version, and the date are recorded with every query that is relied on.

- **CAUTION**

- A model output is not evidence and not an expert's opinion. It is
- a pointer into the record. What is cited, relied on, or testified to
- is the record itself, read by the person who vouches for it.

06 Working rules

Seven rules that make the method defensible. They are short because they are meant to be kept.

01 **Originals are never edited**

The archive is a copy. The production, as received, is the evidence and stays as it is.

02 **Every statement has a source**

Document ID and page, or it is not used. No exceptions for the obvious.

03 **Every source is read**

By a person, before the statement goes into anything with a name on it.

04 **Counts are verified**

A number produced by classification is checked by reading the items it counts, or it is reported as an estimate.

05

Confidentiality is settled first

Terms of the tool, the protective order, and the client's instructions, reviewed and recorded before upload.

06

Queries are logged

Question, model, version, date, and answer, filed with the archive, so the work can be repeated and explained.

07

The expert reads the record

The model narrows the reading. It does not replace it. An opinion rests on what the expert read, and the expert must be able to say so.

NOTE

A team that keeps these rules can explain, for every statement, where it came from and who checked it. That is the whole point.

CHECKLIST

Document analysis at a glance

- ☐ Confidentiality, protective order, and tool terms reviewed by counsel and recorded
- ☐ Production inventoried with IDs, families, duplicates, and formats
- ☐ Text extracted with page markers; OCR sampled against the images
- ☐ One text file per document, with a standard header and quality note
- ☐ Index built and checked against the inventory
- ☐ Logs, databases, and spreadsheets kept as tables
- ☐ Gaps recorded: unproduced references, illegible pages, redactions
- ☐ Standard questions written down and reused as the production grows
- ☐ Every answer cites document and page; every cited page read
- ☐ Counts verified by reading, or reported as estimates
- ☐ Query log kept: question, model, version, date, answer
- ☐ Expert's own reading of the record completed and recorded

SOURCES

Sources and further reading

American Bar Association — Formal Opinion 512, Generative Artificial Intelligence Tools (July 2024)

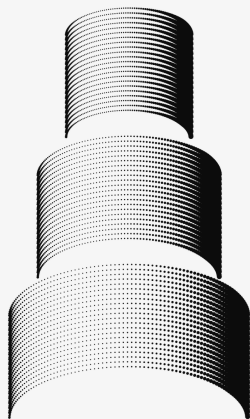
americanbar.org

NIST — Artificial Intelligence Risk Management Framework (AI RMF 1.0, 2023)

nist.gov

Alpha Devices — Guide 05, The manufacturer's record; Guide 13, Complaints, MDRs and CAPA; Guide 15, What did they know, and when?

alphadevices.io



SERVICES

We build the archive and read the record

We turn a production into an indexed text archive, run the standard questions across it, verify every answer against the page, and read the technical record ourselves.

01

The archive

Inventory, extraction with page markers, headers, index, and tables for logs and data, as a working set the whole team can query.

02

The three tasks

Timeline, request against production, and complaint classification, as verified tables with sources.

03

The engineering reading

Design, risk, software, and field records read by engineers who know what should be there, with the model's output as the map.

TERMS

Terms used in this guide

Language model

A program that produces text in reply to text, trained on large bodies of writing. It has no access to facts beyond what it was given and what it was trained on.

Index

The table of every document in the archive: ID, title, date, type, custodian, pages, and quality notes. The map the questions start from.

Retrieval

Finding, for each question, the parts of a large archive that most resemble it, and giving only those to the model.

Invention

A fluent statement by a model with no basis in the text it was given, including citations to pages that do not say what is claimed.

OCR

Optical character recognition: turning an image of a page into text. Imperfect on scans, stamps, handwriting, and tables.

Production

The set of documents a party produces in discovery, each with an identifying number.

Page marker

A line in the extracted text at every page boundary, so a citation can name the page of the original.

Query log

The record of each question asked: the question, the model and version, the date, and the answer.

CONTACT

Tell us about the device.

Send a short description of the device, the event, and any deadlines. We will tell you what to preserve now and recommend the next steps.

EMAIL info@alphadevices.io

WEB alphadevices.io

GUIDES

- | | | | |
|----|--|----|--------------------------------------|
| 01 | A device failed. What happens next? | 09 | How a device reaches the market |
| 02 | Searching MAUDE with purpose | 10 | Why the alarm did not help |
| 03 | Understanding waterfall design | 11 | When the software is the suspect |
| 04 | Reading an ISO 14971 risk analysis | 12 | Use error is a design question |
| 05 | The manufacturer's record | 13 | Complaints, MDRs and CAPA |
| 06 | Using LLMs to review case documents | 14 | Recalls, corrections and quiet fixes |
| 07 | What the device recorded | 15 | What did they know, and when? |
| 08 | The joint inspection | | |

Alpha Devices

2101 Brandywine Street, Suite 200
Philadelphia, PA 19130

Independent medical device engineering. © 2026 Alpha Devices.

