Prompt Cookbook and Best Practices Guide

User Guide

Fookes Software Ltd
Charmey, Switzerland
www.aid4mail.com

Table of Contents

Table of Contents 2
1. How to Use This Guide 6
2. Aid4Mail Prompt Library Basics 7
  2.1 Accessing the prompt library 7
  2.2 Library organization 7
  2.3 Library prompts are starting points, not production protocols 7
  2.4 INCONCLUSIVE and Review 8
3. Choosing the Right Aid4Mail AI Task 8
  3.1 Task comparison 8
  3.2 Use deterministic filters first 9
  3.3 Prefer Classify when every decision must be auditable 9
4. Core Prompt-Design Rules 9
  4.1 Day-one writability 9
  4.2 Screen the theme before drafting 10
  4.3 Prefer positive framing 10
  4.4 Use one operational test for the positive class 10
  4.5 Enumerate exclusions freely 11
  4.6 Keep prompts short enough to behave consistently 11
  4.7 Build decisive exclusions into the prompt 12
  4.8 Keep a readable source version and a single-line Aid4Mail version 12
5. Standard Prompt Anatomy 12
  5.1 Readable source version 13
  5.2 Aid4Mail single-line version 13
6. Issue Coding and Category Taxonomy Design 13
  6.1 Use the smallest useful taxonomy 13
  6.2 Make categories mutually exclusive 14
  6.3 Separate broad taxonomies into multiple passes when necessary 14
  6.4 Privileged, Exempt, and legal-specific labels 15
  6.5 Label naming rules 15
7. Reusable Prompt Templates 16
  7.1 Binary responsiveness template 16
  7.2 Semantic AI Filter template 16
  7.3 Restricted multi-category template 16
  7.4 Responsiveness plus privilege template 16
  7.5 FOIA/public-records responsiveness and holdback template 17
  7.6 Legal risk-level template 17
  7.7 Review-priority template 17
  7.8 Summary template 17

7.9 Translation template 18
7.10 Evidence-location analysis template 18
7.11 Entity and date extraction template 18
8. Cookbook Recipes by Practice Area 18
8.1 Digital forensics recipes 18
8.1.1 Insider threat / data exfiltration classification 18
8.1.2 Phishing and social engineering triage 19
8.1.3 Business Email Compromise analysis 19
8.1.4 Money laundering / illicit finance triage 19
8.1.5 Workplace violence threat triage 20
8.2 eDiscovery recipes 20
8.2.1 Discrimination or retaliation review 20
8.2.2 Antitrust and competition review 20
8.2.3 Securities litigation / insider trading triage 21
8.2.4 Privilege classification 21
8.2.5 Legal risk and priority workflow 21
8.3 FOIA/public-records recipes 22
8.3.1 General FOIA/public-records responsiveness 22
8.3.2 FOIA financial regulation and oversight 22
8.3.3 Law enforcement practices 22
8.3.4 Environmental regulation and impact 23
8.3.5 Travel and expense reports 23
9. Forwarded Threads, Attachments, and Source-Limitation Rules 23
9.1 Participant-authored evidence rule 23
9.2 When forwarded content may matter 24
9.3 Attachment strategy 24
9.4 Prompt language for attachments 24
10. Validation Workflow Before Production 25
10.1 Define acceptance criteria first 25
10.2 Build a representative sample 25
10.3 Review the right sets 26
10.4 Use the HTML portable viewer for prompt validation 26
10.5 Validation record 26
11. Prompt Iteration and Troubleshooting 27
11.1 False-positive analysis 27
11.2 Too many false positives 27
11.3 Too many missed items 27
11.4 Too many INCONCLUSIVE results 28
11.5 Unexpected labels or output format 28

11.6 Prompt-model fit 28
11.7 When to stop iterating 28
12. Overbroad Prompts and Corrected Versions 29
12.1 Overbroad forwarded-content prompt 29
12.2 Overbroad privilege prompt 29
12.3 Overbroad insider-threat prompt 29
12.4 Overbroad legal-risk prompt 30
13. Practical Sample Workflows 30
13.1 First-run prompt validation workflow 30
13.2 eDiscovery issue coding with privilege 31
13.3 FOIA/public-records responsiveness and exemption workflow 31
13.4 Digital-forensics exfiltration triage 31
13.5 Analyze workflow for summaries and translations 32
14. Defensibility and Preservation 32
14.1 Preserve the production record 32
14.2 Prompt version log 33
14.3 Defensibility language for review notes 33
15. Quality Control for Analyze Outputs 33
15.1 Summaries 34
15.2 Translations 34
15.3 Entity extraction 34
15.4 Risk notes and explanations 34
16. Known Limits and Cautionary Patterns 34
16.1 Subjective themes 34
16.2 Absence-detection themes 35
16.3 Model disagreement 35
16.4 Legal and policy limits 35
16.5 Sensitive investigations 35
Appendix A: Prompt Readiness Checklist 36
Matter objective 36
Prompt design 36
Validation setup 36
Appendix B: Validation Worksheet 37
Appendix C: Category Precedence Worksheet 38
Appendix D: One-Line Prompt Conversion Examples 38
D.1 Readable version 38
D.2 Aid4Mail single-line version 39
D.3 Conversion checks 39
Appendix E: Prompt Version Log Template 39

Appendix F: Source Map 39
Appendix G: Theme Inventory for Cookbook Planning 41
eDiscovery 41
Digital Forensics 41
FOIA/Public Records 41

Prompt Cookbook and Best Practices Guide

Using reusable prompts, issue-coding taxonomies, validation practices, and sample workflows for Aid4Mail AI-assisted email review

Audience: Review leads, investigators, litigation-support teams, FOIA/public-records reviewers, DFIR practitioners, and eDiscovery professionals using Aid4Mail Investigator or Aid4Mail Enterprise.

Scope: This guide focuses on prompt drafting, issue coding, validation, quality control, and review workflows. It does not cover provider credential setup, cloud region configuration, offline model installation, or detailed model-selection tables.

Primary source material: Aid4Mail AI Integration User Guide §8; Aid4Mail AI Email Review Workflow Guide; bundled Aid4Mail prompt examples and theme inventory; Aid4Mail benchmark methodology notes; Podesta corpus methodology note.

---

1. How to Use This Guide

Aid4Mail AI workflows are controlled by prompts: written instructions that tell the selected model what to decide, what to ignore, and what output label or field to return. This guide converts the project source material into practical reusable patterns.

Use it for four recurring tasks:

1. **Drafting prompts** for AI Filter, AI Classify, and AI Analyze tasks.
2. **Designing issue-coding taxonomies** such as `Responsive`, `Unresponsive`, `INCONCLUSIVE`, `Privileged`, `Exempt`, `High Risk`, or matter-specific issue labels.
3. **Validating prompts before production** using representative samples, known positives, false-positive review, and documented acceptance criteria.
4. **Running practical review workflows** for eDiscovery, digital forensics, FOIA/public-records review, and internal investigations.

Treat every template as a starting point. A prompt that is defensible in one matter may be overbroad, underinclusive, or legally inappropriate in another. Revise each prompt for the matter objective, custodians, languages, attachment profile, legal rules, and review protocol.

---

2. Aid4Mail Prompt Library Basics

Aid4Mail includes a library of pre-written prompts organized by **task** and **theme**. The prompt library is designed to prevent review teams from starting with a blank page, but the prompts still require matter-specific review and validation.

2.1 Accessing the prompt library

In Aid4Mail:

1. Open **Project Settings**.
2. Select the **AI** tab.
3. Find the **Prompt** field for the relevant task: **Filter**, **Classify**, or **Analyze**.
4. Click **Open** above the Prompt field to browse available prompts.
5. Select a prompt, adapt it, verify it, and test it on a small sample before using it in production.

2.2 Library organization

Aid4Mail prompt themes are grouped into three major families.

Theme family	Typical use	Examples
Digital Forensics	Evidence triage, incident investigation, threat detection, fraud indicators, suspicious communications	Data Exfiltration, Insider Threat, Business Email Compromise, Phishing and Social Engineering, Malware Distribution, Crypto Fraud, Money Laundering
eDiscovery	Litigation review, issue coding, legal risk triage, privilege and sensitivity workflows	Antitrust and Competition, Discrimination or Retaliation, Insider Threat, Privilege Classification, Securities Litigation, Trade Secret Misappropriation, Legal Risk Level
FOIA/Public Records	Public-records responsiveness, disclosure review, public-interest triage, exemption or privilege routing	Government Misconduct, Environmental Regulation and Impact, Law Enforcement Practices, Lobbying and Political Influence, Travel and Expense Reports, Surveillance and Data Privacy

Sensitive themes such as child exploitation, terrorism-related activity, extremist recruitment, non-consensual content distribution, or human trafficking may appear in the library for legitimate forensic workflows. Use those prompts only under the applicable legal, investigative, and organizational protocols. This guide lists such themes only at a high level and does not expand them into detailed recipes.

2.3 Library prompts are starting points, not production protocols

Before production, always check whether the selected prompt:

- Matches the matter objective.
- Uses the correct AI task: Filter, Classify, or Analyze.
- Uses output labels that match the review protocol.
- Defines both responsive and unresponsive conditions.
- Handles forwarded or quoted content.

- Handles ambiguity through `INCONCLUSIVE` or an equivalent review label.
- Fits the selected model and corpus language.
- Has been validated on a representative sample.

2.4 INCONCLUSIVE and Review

Aid4Mail commonly uses `INCONCLUSIVE` as the human-review or abstention label. Some prompt-library examples use `Review` where that label suited the context. For new review workflows, use `INCONCLUSIVE` unless the matter protocol specifically requires `Review`, `Needs Review`, or another human-review label.

Use the labels consistently. Do not mix `Review`, `INCONCLUSIVE`, `Inconclusive`, and `Needs Review` in the same restricted classification unless each label has a distinct documented meaning.

---

3. Choosing the Right Aid4Mail AI Task

Aid4Mail AI has three main task types: **Filter**, **Classify**, and **Analyze**. The same concept can often be expressed through more than one task, but the defensibility and output implications differ.

3.1 Task comparison

Task	What it does	Typical output	Best use	Caution
AI Filter	Determines whether an email passes a semantic condition	Include/exclude from downstream processing, often `True` / `False`	Reducing a corpus when the semantic rule is clear and the consequence of exclusion is controlled	Do not use as unreviewed culling for high-stakes privilege, exemption, redaction, or responsiveness decisions unless passed and rejected records are preserved and validated
AI Classify	Assigns one label to each email	Folder names or exported fields such as `AI.Classify` / `X-AI-Classify`	Responsiveness decisions, issue coding, privilege/exemption routing, triage categories	Categories must be mutually exclusive or governed by precedence rules
AI Analyze	Generates text or structured analysis	Summary, translation, entities, risk notes, evidence location, reasoning field	Reviewer assistance, summaries, translations, extraction, short issue notes	Outputs are less constrained and require additional checking for omissions or unsupported statements

3.2 Use deterministic filters first

Do not use AI for rules that deterministic Aid4Mail filters can apply faster and more cheaply. Use standard filters for:

- Date ranges.
- Known senders, recipients, custodians, or domains.
- Known folders.
- File types.
- Exact terms or simple Boolean logic.
- Known duplicates or nonresponsive system notifications.

Use AI when the decision depends on meaning, context, intent, indirect language, multilingual content, or distinctions that cannot be reliably expressed as keywords.

3.3 Prefer Classify when every decision must be auditable

For legal review, FOIA/public-records workflows, privilege review, sensitivity routing, and high-stakes culling, **AI Classify** is usually preferable to **AI Filter** because it can export a category for every email.

Use Classify when the record should show:

- Which emails were `Responsive`.
- Which emails were `Unresponsive`.
- Which emails were `INCONCLUSIVE` and routed to human review.
- Which emails were `Privileged`, `Exempt`, `High Risk`, or assigned to a matter-specific issue category.

AI Filter can still be appropriate, but preserve processing logs and filter decision records for both passed and rejected emails.

---

4. Core Prompt-Design Rules

The strongest prompts are short, operational, and testable. They describe the decision a qualified reviewer would make on day one of a matter, before seeing the positive examples.

4.1 Day-one writability

A prompt should be writable at the start of an investigation. It should express the review objective in general professional language, not in phrases mined from the target corpus.

Acceptable prompt material:

- General behavioral descriptions.
- Domain-standard exclusions.
- Matter objectives from the review protocol.
- Category definitions approved by counsel, agency staff, or investigators.
- Rules about source limitation, such as participant-authored text versus forwarded material.

Avoid:

- Phrases copied from positive examples after reading the corpus.
- Named entities discovered only through corpus review, unless they are part of the matter scope.
- Long lists of exact phrases that convert the prompt into a keyword query.
- Prompt revisions that simply encode known false positives or true positives without a general rule.

4.2 Screen the theme before drafting

Before writing a full prompt, ask whether the theme is suitable for AI classification.

Viability question	Strong sign	Weak sign
Can the positive class be expressed as one operational test?	"Does this email show unauthorized movement of company data?"	"Does this email match one of ten unrelated suspicious behaviors?"
Does the prompt detect presence rather than infer absence?	Detects explicit conduct, topic, sentiment, or request	Requires inferring what the sender deliberately avoided saying
Do expected positives share a semantic signature?	Common behavior, topic, or issue	Unrelated patterns held together only by case theory
Would two reviewers apply the rule consistently?	Clear boundary and examples of exclusions	Subjective theme with plausible disagreement on many emails
Would a keyword search fail for good reasons?	Conceptual, indirect, multilingual, or contextual evidence	Same result likely obtainable with a simple phrase list

If a theme fails several of these checks, narrow it, split it into multiple passes, or treat the output as exploratory analysis rather than a production classification.

4.3 Prefer positive framing

Models handle direct instructions better than nested negation. Write what each label means in positive terms.

Better:

```
Reply Unresponsive when the email is routine scheduling, logistics, a newsletter, a press roundup, an automated alert, or a forwarded article with no substantive participant-authored commentary.
```

Weaker:

```
Do not reply Responsive if the email is not non-routine and does not contain non-public non-logistical content.
```

The word `not` is sometimes unavoidable, but avoid using it as the central logic of the prompt.

4.4 Use one operational test for the positive class

The positive criterion should usually answer one question.

Prompt objective	Strong operational test
Insider threat	Does a participant appear to move, collect, disclose, or access data outside authorized channels?
Privilege	Does the email seek, provide, relay, or discuss legal advice or attorney work product?
FOIA/public records	Does participant-authored text directly concern the subject matter of the request?
Discrimination or retaliation	Does the email contain evidence of biased treatment, protected activity, adverse action, or retaliation?
Business Email Compromise	Does the email attempt to impersonate a trusted party to redirect funds, credentials, or sensitive information?

Avoid stuffing the positive definition with a long list of patterns. Lists in the positive class can make the model trigger on isolated words instead of applying the overall concept.

4.5 Enumerate exclusions freely

Lists work well in exclusion clauses because the intended logic is usually “if any of these apply, exclude.”

Good exclusion lists include:

- Routine scheduling.
- Logistics.
- Newsletters.
- Automated alerts.
- Spam and bulk mail.
- Press roundups.
- Forwarded articles without participant commentary.
- Boilerplate signatures and disclaimers.
- Routine business operations unrelated to the issue.

A strong prompt often has a concise positive test and a more explicit unresponsive/exclusion section.

4.6 Keep prompts short enough to behave consistently

As a drafting target:

- **60–90 words** is often enough for simple library-style prompts.
- **100–150 words** is a strong practical target for production review prompts.
- **Up to about 250 words** may be justified for multi-category classification, privilege/exemption logic, or forwarded-content rules.
- **Over 250 words** should trigger a redesign review. The theme may be too broad, or the prompt may need to be split into multiple passes.

Longer prompts can degrade consistency, especially on smaller offline models and broad taxonomies.

4.7 Build decisive exclusions into the prompt

A decisive exclusion is the rule that prevents the prompt from collapsing into keyword-style behavior.

Example:

```
Classify as Unresponsive when the relevant terms appear only in a forwarded
article, quoted thread, or third-party message and the actual email participants
add no substantive commentary of their own.
```

This is stronger than telling the model to “avoid false positives from forwarded content” because it defines the review rule.

4.8 Keep a readable source version and a single-line Aid4Mail version

Aid4Mail prompt fields use a single continuous prompt. For drafting and defensibility, keep two versions:

- **Readable source version:** line breaks, bullets, version notes, and reviewer comments.
- **Aid4Mail version:** one continuous paragraph with the same meaning and exact output labels.

Do not rely on visual formatting to carry meaning. When converting to one line, preserve logic with punctuation and explicit label names.

5. Standard Prompt Anatomy

A defensible classification prompt usually contains seven elements.

Element	Purpose	Example
Role	Frames the professional perspective	`You are a digital forensics investigator.`
Task	Defines the required decision	`Classify the email into exactly one category.`
Responsive criteria	Defines the positive class	`Reply Responsive if the email shows unauthorized data movement or insider-threat activity.`
Unresponsive criteria	Defines exclusions and ordinary-course content	`Reply Unresponsive for routine business, scheduling, newsletters, or content unrelated to the issue.`
INCONCLUSIVE threshold	Creates the human-review queue	`Reply INCONCLUSIVE if relevant signals exist but the content is too ambiguous for a reliable decision.`
Source limitation	Controls quoted/forwarded/attachment treatment	`Base the decision only on participant-authored content, not forwarded third-party text.`

Output format	Makes parsing auditable	`Reply with only one label: Responsive, Unresponsive, or INCONCLUSIVE.`

5.1 Readable source version

```
You are a digital forensics investigator. Classify the email into exactly one
category.

Reply Responsive if participant-authored content shows unauthorized data movement,
insider-threat activity, credential sharing, unauthorized access, systematic
collection of files before departure, use of personal storage or messaging to
bypass monitoring, or attempts to circumvent security controls.

Reply Unresponsive for routine business, normal IT support, general cybersecurity
awareness, newsletters, alerts, scheduling, logistics, or forwarded material where
the actual email participants add no substantive commentary.

Reply INCONCLUSIVE if participant-authored content suggests one of the Responsive
behaviors but lacks sufficient clarity or context for a reliable decision.

Reply with only one label: Responsive, Unresponsive, or INCONCLUSIVE.
```

5.2 Aid4Mail single-line version

```
You are a digital forensics investigator. Classify the email into exactly one
category. Reply Responsive if participant-authored content shows unauthorized data
movement, insider-threat activity, credential sharing, unauthorized access,
systematic collection of files before departure, use of personal storage or
messaging to bypass monitoring, or attempts to circumvent security controls. Reply
Unresponsive for routine business, normal IT support, general cybersecurity
awareness, newsletters, alerts, scheduling, logistics, or forwarded material where
the actual email participants add no substantive commentary. Reply INCONCLUSIVE if
participant-authored content suggests one of the Responsive behaviors but lacks
sufficient clarity or context for a reliable decision. Reply with only one label:
Responsive, Unresponsive, or INCONCLUSIVE.
```

6. Issue Coding and Category Taxonomy Design

Issue coding is the process of assigning review labels that are useful for legal review, investigative triage, production, escalation, or quality control. AI can apply these labels consistently, but only if the taxonomy is clear.

6.1 Use the smallest useful taxonomy

Prefer the smallest category set that supports the workflow.

Workflow need	Recommended taxonomy
Basic responsiveness	`Responsive`, `Unresponsive`, `INCONCLUSIVE`
Responsiveness plus privilege	`Privileged`, `Responsive`, `Unresponsive`, `INCONCLUSIVE`
FOIA/public records with legal holdbacks	`Privileged`, `Exempt`, `Responsive`, `Unresponsive`, `INCONCLUSIVE`
Early risk triage	`High Risk`, `Moderate Risk`, `Low Risk`, `INCONCLUSIVE`
Multi-issue misconduct triage	Matter-specific issue labels plus `Clean` and `INCONCLUSIVE`
Review priority	`High`, `Medium`, `Low`, `INCONCLUSIVE`

Do not add categories merely because they are interesting. Every label should change routing, review priority, reporting, or production handling.

6.2 Make categories mutually exclusive

Overlapping labels create inconsistent outputs. If categories overlap, define precedence.

Example precedence rules:

```
If an email fits both Privileged and Responsive, choose Privileged.
If an email fits both Exfiltration and Compliance, choose Exfiltration.
If an email fits both Financial and Corruption, choose Corruption only when
bribery, influence, kickbacks, or quid pro quo conduct is explicit; otherwise
choose Financial.
If none of the issue categories applies, choose Clean.
Use INCONCLUSIVE only when relevant signals exist but the category cannot be chosen
reliably.
```

6.3 Separate broad taxonomies into multiple passes when necessary

A single seven-label prompt can work for early-stage triage, but production review often benefits from focused passes.

Broad task	Safer decomposition
“Find all misconduct”	Separate passes for financial crime, corruption, discrimination/retaliation, compliance, and exfiltration
“Find all legally sensitive emails”	Separate privilege, legal-risk, confidentiality, and redaction passes
“Review all FOIA exemptions”	Separate responsiveness pass, then exemption/privilege/redaction pass on likely responsive items
“Find all cybercrime”	Separate phishing/BEC, malware, credential theft, unauthorized access, and data-exfiltration passes

Use a broad taxonomy for triage. Use focused prompts for defensible classification when the stakes are high.

6.4 Privileged, Exempt, and legal-specific labels

Privilege, work product, FOIA exemptions, privacy exclusions, redactions, and disclosure rules are jurisdiction- and matter-specific. This guide provides generic taxonomy patterns only.

Do not treat a generic prompt as a legal determination. Have counsel or the appropriate agency reviewer define:

- The legal privilege standard.
- Applicable FOIA/public-records exemptions.
- Redaction categories.
- Privacy and confidentiality rules.
- Category precedence.
- Escalation and human-review procedures.

6.5 Label naming rules

Use labels that are exact, short, and stable.

Good labels:

```
Responsive
Unresponsive
INCONCLUSIVE
Privileged
Exempt
High Risk
Moderate Risk
Low Risk
Clean
Financial
Corruption
Discrimination
Compliance
Exfiltration
```

Avoid:

```
Maybe Responsive
Probably OK
Needs a look
Sensitive-ish
Not sure
Relevant?
```

If using restricted classification in Aid4Mail, enter the allowed category list exactly as the prompt requires the model to return it.

7. Reusable Prompt Templates

The templates below are intentionally generic. Replace bracketed text with matter-specific criteria, labels, and exclusions. For Aid4Mail, convert the final prompt to a single line before saving it.

7.1 Binary responsiveness template

Use when the primary question is whether the email meets one defined issue.

```
You are an experienced [role]. Classify the email into exactly one category. Reply
Responsive if participant-authored content [single operational test for the issue].
Reply Unresponsive if the email [ordinary-course exclusions, domain-standard
exclusions, and nonresponsive examples]. When evaluating forwarded or quoted
content, base the classification only on comments written by the actual email
participants, not third-party material they merely forwarded or quoted. Reply
INCONCLUSIVE if participant-authored content suggests the Responsive behavior but
lacks sufficient clarity, specificity, or context for a reliable decision.
Otherwise, reply Unresponsive. Reply with only one label: Responsive, Unresponsive,
or INCONCLUSIVE.
```

7.2 Semantic AI Filter template

Use only when a pass/fail routing decision is appropriate and both passed and rejected decisions will be preserved as required by the matter.

```
Determine whether this email is potentially relevant to [issue]. Return True if
participant-authored content directly concerns [semantic condition], including
indirect or non-keyword expressions of that condition. Return False for routine
business, scheduling, logistics, newsletters, automated alerts, forwarded or quoted
third-party material without substantive participant-authored commentary, or
content unrelated to [issue]. Return True for ambiguous potentially relevant emails
only if they contain a concrete signal requiring review; otherwise return False.
```

7.3 Restricted multi-category template

Use for early-stage triage or issue coding with a fixed category list.

```
You are an experienced [role]. Classify the email into exactly one of the following
categories: [Category A], [Category B], [Category C], Clean, INCONCLUSIVE. Use
[Category A] when [definition]. Use [Category B] when [definition]. Use [Category
C] when [definition]. Use Clean when none of the issue categories applies. Use
INCONCLUSIVE only when the email contains relevant signals but lacks enough clarity
or context to choose one category reliably. Apply these precedence rules: [rules].
When evaluating forwarded or quoted content, classify based only on participant-
authored contributions. Reply with only one category name.
```

7.4 Responsiveness plus privilege template

Use when privilege must be routed before ordinary responsiveness.

```
You are an experienced eDiscovery reviewer. Classify the email into exactly one
category: Privileged, Responsive, Unresponsive, or INCONCLUSIVE. Choose Privileged
when the email seeks, provides, relays, or discusses confidential legal advice,
attorney-client communications, or attorney work product under the matter protocol.
Choose Responsive when participant-authored content clearly concerns [responsive
issue] and the email is not Privileged. Choose Unresponsive when the email does not
```

```
clearly concern [responsive issue], is routine business, scheduling, logistics, or
forwarded third-party material without substantive participant-authored commentary.
Choose INCONCLUSIVE when relevant signals exist but there is insufficient clarity
to classify confidently. If both Privileged and Responsive apply, choose
Privileged. Reply with only one category name.
```

7.5 FOIA/public-records responsiveness and holdback template

Use only after legal or agency reviewers define the relevant exemption and privilege standards.

```
You are an experienced FOIA/public-records reviewer. Classify the email into
exactly one category: Privileged, Exempt, Responsive, Unresponsive, or
INCONCLUSIVE. Choose Privileged when the email falls within the matter-specific
privilege standard. Choose Exempt when the email is responsive but contains
information covered by the matter-specific exemption or redaction protocol. Choose
Responsive when participant-authored content directly concerns [request subject]
and no Privileged or Exempt category applies. Choose Unresponsive for newsletters,
press roundups, automated alerts, routine scheduling, logistics, spam, general
operations unrelated to [request subject], or forwarded/quoted third-party material
with no substantive participant-authored commentary. Choose INCONCLUSIVE only when
participant-authored language suggests responsiveness or a holdback category but
lacks sufficient clarity or context. If both Exempt and Privileged apply, choose
Privileged unless the matter protocol says otherwise. Reply with only one category
name.
```

7.6 Legal risk-level template

Use for prioritization, not final legal conclusions.

```
Assess this email’s potential legal exposure based only on the email content. Reply
High Risk if it contains explicit admissions of wrongdoing, clear policy or
regulatory violations, threats of legal action, sensitive executive involvement, or
statements directly harmful in litigation. Reply Moderate Risk if it contains
ambiguous legal concerns, potential compliance issues, disputed facts, or content
requiring legal review. Reply Low Risk if it is routine business correspondence
with minimal apparent legal exposure. Reply INCONCLUSIVE if there is insufficient
content to assess the risk reliably. Reply with only one label: High Risk, Moderate
Risk, Low Risk, or INCONCLUSIVE.
```

7.7 Review-priority template

Use to sort an already relevant or likely relevant set.

```
Assess this email’s priority for human review. Reply High if it contains explicit
deadlines, imminent legal or regulatory action, senior executive involvement,
potential evidence of wrongdoing, privilege concerns, or time-sensitive decisions.
Reply Medium if it references ongoing legal, regulatory, investigative, or business
issues but no immediate deadline or decisive event. Reply Low if it is routine
correspondence with no apparent urgency. Reply INCONCLUSIVE if the available
content is insufficient to assign priority reliably. Reply with only one label:
High, Medium, Low, or INCONCLUSIVE.
```

7.8 Summary template

Use as an AI Analyze prompt.

```
Summarize the email in English using 150 words or fewer. Focus on the primary
topic, specific requests, action items, decisions, key facts, figures, deadlines,
and the sender's intent or desired outcome. Omit greetings, signatures,
boilerplate, and redundant material. If the email is not in English, translate the
summary into English. Do not add facts that are not supported by the email.
```

7.9 Translation template

Use when reviewers need English-language output but source tone and meaning matter.

```
Translate the provided email content into English. Preserve meaning, tone, intent,
informal language, euphemisms, and any coded or indirect references as accurately
as possible. Where a phrase has uncertain meaning, translate it literally and note
the uncertainty briefly. Do not summarize, omit, or add facts.
```

7.10 Evidence-location analysis template

Use when reviewers need a short explanation and location of evidence.

```
Analyze this email to determine whether it is relevant to [issue]. If clear,
specific participant-authored evidence is present, begin with RESPONSIVE: and
briefly explain the evidence, including where it appears: subject line, message
body, URL, attachment name, attachment content, metadata, or combination. If
potentially relevant evidence is present but ambiguous, begin with INCONCLUSIVE:
and explain what requires review. If the email does not match the criteria, respond
only with UNRESPONSIVE. Do not rely on forwarded or quoted third-party content
unless the actual email participants add substantive commentary.
```

7.11 Entity and date extraction template

Use for structured review support.

```
Extract the people, organizations, locations, dates, deadlines, monetary amounts,
and document names mentioned in the email. Return a concise semicolon-separated
list using these headings: People; Organizations; Locations; Dates/Deadlines;
Amounts; Documents; Notes. Include only information present in the email. If a
heading has no entries, write None.
```

8. Cookbook Recipes by Practice Area

The recipes below are not final production prompts. They are matter-ready starting points that should be adapted and validated.

8.1 Digital forensics recipes

8.1.1 Insider threat / data exfiltration classification

Recommended task: AI Classify
Labels: `Responsive`, `Unresponsive`, `INCONCLUSIVE`

```
You are a digital forensics investigator. Classify the email into exactly one
category. Reply Responsive if participant-authored content shows data exfiltration
or insider-threat activity, including unauthorized transfer of files or datasets,
```

```
use of personal accounts or storage to bypass monitoring, sharing proprietary
information with unauthorized parties, credential sharing, unauthorized access,
systematic data collection before departure, threats to delete or corrupt data, or
attempts to circumvent security controls. Reply INCONCLUSIVE if the content
suggests one of these behaviors but lacks enough clarity or context for a reliable
decision. Otherwise, reply Unresponsive. Reply with only one label: Responsive,
Unresponsive, or INCONCLUSIVE.
```

Validation notes: Include known or likely positives, benign IT support emails, routine file-sharing emails, departure logistics, and messages where attachment content may matter.

8.1.2 Phishing and social engineering triage

Recommended task: AI Classify or AI Filter
Labels for Classify: `Responsive`, `Unresponsive`, `INCONCLUSIVE`

```
You are a digital forensics investigator. Classify the email into exactly one
category. Reply Responsive if the email appears to be a phishing, spoofing,
credential-harvesting, malware-delivery, or social-engineering attempt, including
impersonation, suspicious links, suspicious attachments, urgent payment or
credential requests, sender anomalies, or deceptive requests to bypass normal
controls. Reply Unresponsive for ordinary business requests, legitimate security
notices, newsletters, training materials, or benign discussions of phishing that
are not themselves attack attempts. Reply INCONCLUSIVE if suspicious indicators
exist but the email lacks enough context for a reliable decision. Reply with only
one label: Responsive, Unresponsive, or INCONCLUSIVE.
```

Validation notes: Check sender-domain anomalies, link text, attachment names, and security-awareness newsletters that may otherwise create false positives.

8.1.3 Business Email Compromise analysis

Recommended task: AI Analyze after an initial Classify or Filter pass

```
Analyze this email for Business Email Compromise indicators. Begin with RESPONSIVE:
if participant-authored content attempts to impersonate an executive, supplier,
customer, or trusted party to redirect payment, change bank details, obtain
credentials, or disclose sensitive information. Briefly identify the evidence and
where it appears. Begin with INCONCLUSIVE: if suspicious indicators exist but
authenticity or intent cannot be determined from the email alone. Respond only
UNRESPONSIVE if no BEC indicators are present. Do not treat general fraud-awareness
training or news articles as responsive unless the email itself contains a
suspected BEC attempt.
```

Validation notes: Include false-positive examples such as legitimate invoice changes, finance-policy reminders, vendor onboarding, and cybersecurity training.

8.1.4 Money laundering / illicit finance triage

Recommended task: AI Classify

```
You are a digital forensics investigator. Classify the email into exactly one
category. Reply Responsive if participant-authored content shows potential money
laundering, illicit transfers, structuring, use of shell entities, suspicious
cryptocurrency movement, attempts to conceal source of funds, or coordination of
transactions that appear designed to avoid reporting or detection. Reply
Unresponsive for ordinary invoices, routine payments, legitimate accounting, market
commentary, newsletters, or forwarded articles without substantive participant
commentary. Reply INCONCLUSIVE if financial-crime indicators exist but the email
```

```
lacks sufficient clarity or context. Reply with only one label: Responsive,
Unresponsive, or INCONCLUSIVE.
```

Validation notes: Include ordinary finance operations and legitimate crypto or investment discussion to test precision.

8.1.5 Workplace violence threat triage

Recommended task: AI Classify with human escalation protocol

```
You are a digital forensics investigator. Classify the email into exactly one
category. Reply Responsive if participant-authored content contains threats of
physical violence, credible intimidation, aggressive intent toward coworkers or
workplace locations, or descriptions of planned or recent violent conduct. Reply
Unresponsive for ordinary workplace disputes, rude but non-threatening language,
general safety training, news articles, or forwarded third-party content without
substantive participant commentary. Reply INCONCLUSIVE if the language may indicate
a threat but lacks enough clarity or context for reliable classification. Reply
with only one label: Responsive, Unresponsive, or INCONCLUSIVE.
```

Validation notes: Align escalation thresholds with organizational safety procedures. Do not let AI output replace immediate human assessment where there is credible risk of harm.

8.2 eDiscovery recipes

8.2.1 Discrimination or retaliation review

Recommended task: AI Classify
Labels: `Privileged`, `Responsive`, `Unresponsive`, `INCONCLUSIVE`

```
You are an experienced eDiscovery reviewer. Classify the email into exactly one
category: Privileged, Responsive, Unresponsive, or INCONCLUSIVE. Choose Privileged
when the email seeks, provides, relays, or discusses confidential legal advice or
attorney work product under the matter protocol. Choose Responsive when
participant-authored content contains evidence of discrimination, harassment, bias,
protected activity, adverse action, retaliation, disparate treatment, hostile work
environment, or complaint-related consequences. Choose Unresponsive for routine HR
administration, scheduling, benefits, general policy distribution, or unrelated
business communications. Choose INCONCLUSIVE when potentially relevant signals
exist but the context is insufficient for reliable classification. If both
Privileged and Responsive apply, choose Privileged. Reply with only one category
name.
```

Validation notes: Include routine HR emails, performance-management emails, complaint references, legal communications, and borderline tone-only examples.

8.2.2 Antitrust and competition review

Recommended task: AI Classify or AI Filter

```
You are an experienced eDiscovery reviewer. Classify the email into exactly one
category. Reply Responsive if participant-authored content discusses potential
anti-competitive coordination, price fixing, bid rigging, market allocation,
customer allocation, competitor coordination, no-poach arrangements, competitively
sensitive information exchange, or attempts to influence market behavior
improperly. Reply Unresponsive for routine sales, ordinary competitive
intelligence, public market commentary, scheduling, logistics, newsletters, or
forwarded articles without substantive participant commentary. Reply INCONCLUSIVE
if competition-related signals exist but the email lacks sufficient clarity or
```

```
context for a reliable decision. Reply with only one label: Responsive,
Unresponsive, or INCONCLUSIVE.
```

Validation notes: Test against legitimate sales strategy and market-monitoring emails, which are common false positives.

8.2.3 Securities litigation / insider trading triage

Recommended task: AI Classify

```
You are an experienced eDiscovery reviewer. Classify the email into exactly one
category. Reply Responsive if participant-authored content suggests
misrepresentation to investors, undisclosed material information, misleading
financial projections, suspicious trading tied to non-public information, selective
disclosure, or attempts to conceal securities-related risk. Reply Unresponsive for
routine investor relations, public filings, scheduling, ordinary financial
reporting, newsletters, or market commentary without evidence of misconduct. Reply
INCONCLUSIVE if securities-related risk signals exist but the email lacks
sufficient clarity or context. Reply with only one label: Responsive, Unresponsive,
or INCONCLUSIVE.
```

Validation notes: Include public filing discussions, investor-call logistics, earnings-prep drafts, and ordinary forecasting emails.

8.2.4 Privilege classification

Recommended task: AI Classify
Labels: `Privileged`, `Not Privileged`, `INCONCLUSIVE`

```
You are an experienced privilege reviewer. Classify the email into exactly one
category: Privileged, Not Privileged, or INCONCLUSIVE. Choose Privileged when the
email seeks, provides, relays, summarizes, or discusses confidential legal advice,
attorney-client communications, legal strategy, or attorney work product under the
matter protocol. Choose Not Privileged for routine business communications, legal
department administration without legal advice, public legal updates, scheduling,
billing logistics, or forwarded legal news without confidential legal discussion.
Choose INCONCLUSIVE when legal-advice indicators exist but the participants,
confidentiality, or purpose of the communication cannot be determined reliably.
Reply with only one category name.
```

Validation notes: Counsel should approve the privilege standard and review samples from each category. Privilege prompts are jurisdiction-sensitive.

8.2.5 Legal risk and priority workflow

Recommended task: AI Classify after responsiveness culling

```
Assess this email’s priority for legal review based only on its content. Reply High
if it contains explicit admissions, threatened litigation, regulatory action,
senior executive involvement in a disputed issue, clear policy violations,
privilege concerns, or time-sensitive legal deadlines. Reply Medium if it contains
potential legal or compliance concerns but no immediate deadline or explicit
admission. Reply Low if it is routine business correspondence with minimal apparent
legal exposure. Reply INCONCLUSIVE if there is insufficient content to assign
priority reliably. Reply with only one label: High, Medium, Low, or INCONCLUSIVE.
```

Validation notes: Calibrate category distribution with the review lead. “High” should identify a manageable queue, not half the corpus.

8.3 FOIA/public-records recipes

8.3.1 General FOIA/public-records responsiveness

Recommended task: AI Classify
Labels: `Responsive`, `Unresponsive`, `INCONCLUSIVE`

```
You are an experienced FOIA/public-records reviewer. Classify the email into
exactly one category. Reply Responsive if participant-authored content directly
concerns [request subject], including decisions, analysis, communications, actions,
records, or substantive discussion within the scope of the request. Reply
Unresponsive for newsletters, press roundups, automated alerts, event invitations,
routine scheduling, logistics, spam, general operations unrelated to [request
subject], or forwarded/quoted third-party material where participants add no
substantive commentary. Reply INCONCLUSIVE if participant-authored content may
concern [request subject] but lacks sufficient clarity or context for reliable
classification. Reply with only one label: Responsive, Unresponsive, or
INCONCLUSIVE.
```

Validation notes: Test against forwarded articles, press clips, automated alerts, and routine agency operations.

8.3.2 FOIA financial regulation and oversight

Recommended task: AI Classify with matter-specific exemption protocol

```
You are an experienced FOIA/public-records reviewer. Classify the email into
exactly one category: Privileged, Exempt, Responsive, Unresponsive, or
INCONCLUSIVE. Choose Privileged when the matter-specific privilege standard
applies. Choose Exempt when the email is responsive but contains information
covered by the matter-specific exemption or redaction protocol. Choose Responsive
when participant-authored content substantively discusses financial regulation or
oversight within the request scope, such as regulatory examinations, audits,
filings, enforcement, compliance certifications, or oversight decisions. Choose
Unresponsive for routine scheduling, newsletters, press clips, automated alerts, or
forwarded material without participant commentary. Choose INCONCLUSIVE when
potentially responsive or holdback-relevant signals lack clarity. Reply with only
one category name.
```

Validation notes: Agency counsel or FOIA officers should define applicable exemptions and precedence.

8.3.3 Law enforcement practices

Recommended task: AI Classify

```
You are an experienced public-records reviewer. Classify the email into exactly one
category. Reply Responsive if participant-authored content substantively discusses
law enforcement practices within the request scope, including use-of-force policy,
arrest procedures, surveillance technology, body-camera policy, internal affairs
matters, detention practices, training, or oversight of policing activities. Reply
Unresponsive for routine scheduling, news clips, automated alerts, public press
summaries, general administration, or forwarded third-party material without
substantive participant commentary. Reply INCONCLUSIVE if potentially responsive
signals exist but the email lacks enough clarity or context. Reply with only one
label: Responsive, Unresponsive, or INCONCLUSIVE.
```

Validation notes: Include public news clips and automated alerts to confirm that only participant-authored substance triggers responsiveness.

8.3.4 Environmental regulation and impact

Recommended task: AI Analyze or AI Classify

```
Analyze this email to determine whether it is relevant to Environmental Regulation
and Impact within the request scope. Begin with RESPONSIVE: if participant-authored
content discusses environmental compliance, permits, enforcement, environmental
impact assessments, emissions, hazardous waste, environmental advocacy
communications, regulatory decisions, or related policy choices. Briefly identify
the evidence and where it appears. Begin with INCONCLUSIVE: if potentially relevant
signals exist but the email lacks clarity or context. Respond only UNRESPONSIVE if
the email does not match the criteria. Do not rely on forwarded or quoted third-
party content unless participants add substantive commentary.
```

Validation notes: Analyze format is useful when reviewers need both a label and a short evidence note.

8.3.5 Travel and expense reports

Recommended task: AI Filter for low-risk triage, AI Classify if every decision must be exported

```
Determine whether this email is relevant to Travel and Expense Reports within the
request scope. Return True if participant-authored content details official travel,
reimbursement, expense approval, travel policy compliance, lodging, transportation,
per diem, invoices, receipts, or payment of official travel costs. Return False for
routine scheduling without expense content, travel news, newsletters, personal
travel unrelated to official business, or forwarded material without substantive
participant commentary.
```

Validation notes: If the request has disclosure exemptions or privacy rules, use Classify rather than Filter.

---

9. Forwarded Threads, Attachments, and Source-Limitation Rules

Email corpora often contain forwarded articles, quoted threads, nested messages, attachments, disclaimers, and system-generated content. Without explicit rules, these materials can create false positives.

9.1 Participant-authored evidence rule

For responsiveness and issue coding, a useful default is:

```
When evaluating forwarded or quoted content, base the classification only on
comments written by the actual email participants. Classify the email as
Unresponsive when relevant language appears only in a forwarded article, quoted
thread, or third-party message and the actual participants add no substantive
commentary of their own.
```

Use this rule when the review question concerns what the mailbox participants discussed, planned, authorized, acknowledged, or requested.

9.2 When forwarded content may matter

Forwarded or quoted content may matter when the request or issue includes receipt, possession, forwarding, distribution, or awareness of the content itself.

Examples:

- A FOIA request seeks records sent to an agency, including attachments and forwarded materials.
- A leakage investigation asks whether confidential material was distributed, even without commentary.
- A harassment investigation includes forwarded images or offensive content.
- A malware investigation includes malicious attachments or links regardless of participant commentary.

When forwarded or attachment content matters, say so explicitly.

```
Classify based on participant-authored text, forwarded or quoted content, and
attachment text when any of those materials independently satisfy the Responsive
criteria.
```

9.3 Attachment strategy

Aid4Mail can include attachment names and, where configured, extracted attachment text. Attachment names are often useful even when full attachment text is not sent to the model.

Use this decision table.

Situation	Starting strategy
Issue likely appears in email body	Exclude full attachment text for first validation sample; include attachment names
Issue likely appears in attached documents	Include extracted attachment text and set a practical size limit
Attachment-heavy corpus with small local context window	Use attachment names first; validate whether full text materially changes outcomes
Privilege or FOIA review where attachments drive responsiveness	Include attachment text for validation; check truncation and context limits
Malware/phishing/BEC triage	Attachment names and URLs may be sufficient for first pass; include content only where useful and safe

Record whether attachment text was included, the attachment size limit, and any context-window limitations.

9.4 Prompt language for attachments

Use one of these rules.

Attachment names only:

```
Consider attachment names and email metadata, but do not infer the contents of
attachments that are not provided.
```

Full attachment text included:

```
Evaluate the email body and any provided attachment text. If attachment text is
truncated or insufficient, reply INCONCLUSIVE when the missing context prevents
reliable classification.
```

Attachments drive responsiveness:

```
Classify as Responsive when the email body or provided attachment text
independently satisfies the Responsive criteria.
```

---

10. Validation Workflow Before Production

Prompt validation is the difference between a useful AI workflow and an undocumented experiment. Every production prompt should be tested on a sample that represents the matter's risk profile.

10.1 Define acceptance criteria first

Before running the validation sample, record:

- Maximum tolerated missed responsive items in the sample.
- Acceptable false-positive burden.
- Acceptable `INCONCLUSIVE` queue size.
- Whether every `Responsive` and `INCONCLUSIVE` result will be reviewed.
- How `Unresponsive` results will be sampled.
- Whether a second model or senior reviewer is required.
- Who approves the prompt for production.

10.2 Build a representative sample

A validation sample should include more than random documents, especially in low-prevalence matters.

Include:

- Known or likely responsive examples.
- Likely unresponsive examples.
- Borderline or ambiguous examples.
- Major custodians.
- Major date ranges.
- Important folders.
- Representative languages.
- Forwarded or quoted threads.
- Emails with attachments.
- Emails likely to trigger privilege, exemption, redaction, or sensitivity handling.
- Known false-positive candidates, if available.

A practical first validation sample is often 100–500 emails, supplemented with targeted known-positive or likely-positive examples when prevalence is low.

10.3 Review the right sets

At minimum:

1. Review all `Responsive` results for false positives.
2. Review all `INCONCLUSIVE` results.
3. Spot-check `Unresponsive` results.
4. Oversample higher-risk areas, such as uncommon languages, large attachments, high-risk custodians, sensitive categories, and forwarded threads.
5. Record false positives, false negatives, category confusion, and reasons for prompt revisions.

For high-stakes review, also consider:

- A second model on all `Responsive` and `INCONCLUSIVE` results.
- A second model on a sample of `Unresponsive` results.
- Senior reviewer adjudication of disagreement sets.
- Separate recall-focused testing on likely positive examples.

10.4 Use the HTML portable viewer for prompt validation

For validation runs, exporting to HTML with **Include portable email viewer** enabled is often the fastest way to inspect results.

Useful viewer searches include:

```
class:responsive
class:inconclusive
class:privileged
class:exempt
has:attachment class:responsive
tag:yes class:responsive
tag:yes class:inconclusive
from:@company.com class:responsive
```

The viewer's Classification column lets reviewers inspect AI labels. Tags can mark false positives, false negatives, escalation items, or examples for prompt revision. Export viewer tags and archive the tag file if tags are used as validation, QA, escalation, or review-decision records.

10.5 Validation record

Preserve:

- Prompt version.
- Model and provider.
- Attachment inclusion setting and size limit.
- Category list.
- Sample size and sampling rationale.
- Known-positive or likely-positive examples used.
- Reviewer notes.
- False positives and false negatives.
- `INCONCLUSIVE` count and treatment.

- Prompt changes after validation.
- Named approver.
- Final production prompt.

---

11. Prompt Iteration and Troubleshooting

Prompt iteration should improve the general rule, not encode a list of observed mistakes.

11.1 False-positive analysis

When false positives occur, ask two separate questions:

1. **What did the model appear to trigger on?** This helps diagnose the error.
2. **What general rule distinguishes the false positives from true positives?** This belongs in the prompt.

Do not paste false-positive phrases into the prompt as a growing exclusion list unless those phrases express a general matter rule. Prefer a rule such as:

```
Classify as Unresponsive when the substance of the issue is fully described in
writing and the request for a call or meeting is merely logistical.
```

This is stronger than:

```
Do not flag emails that say call, meet, phone, discuss, circle back, sync, or chat.
```

11.2 Too many false positives

Tighten the prompt by:

- Narrowing the Responsive criterion.
- Adding decisive exclusions.
- Adding a forwarded-content rule.
- Requiring participant-authored evidence.
- Requiring clear and specific evidence.
- Defaulting to `Unresponsive` when criteria are not clearly met.
- Splitting broad themes into focused passes.

11.3 Too many missed items

Broaden or clarify the prompt by:

- Adding general behavioral indicators, not corpus-mined phrases.
- Including indirect language that a day-one investigator would reasonably expect.
- Removing overly strict wording.
- Checking whether attachment text is needed.
- Checking whether the selected model handles the corpus language.
- Reviewing whether deterministic pre-filters removed relevant items before AI processing.

11.4 Too many INCONCLUSIVE results

Clarify the threshold by:

- Defining when `INCONCLUSIVE` is allowed.
- Defining when ordinary ambiguity should default to `Unresponsive`.
- Adding category precedence rules.
- Separating a broad multi-category prompt into focused binary prompts.
- Using a stronger model only for gray-zone items.

11.5 Unexpected labels or output format

If output labels do not match the category list:

- Use exact label names in the prompt.
- Add “Reply with only one label.”
- Remove explanatory text from Classify prompts unless explanations are required.
- Check the restricted classification category list in Aid4Mail.
- Use stable capitalization, especially for `INCONCLUSIVE`.
- Validate again after any prompt change.

11.6 Prompt-model fit

A prompt validated on one model should be revalidated on any model intended for production. Models differ in how they handle ambiguity, category boundaries, long prompts, and multilingual content.

Do not tune a prompt solely until a smaller or faster model performs well. That may overfit to the weaker model’s quirks. Use fast models for iteration where useful, but validate on the production model and deployment settings.

11.7 When to stop iterating

Stop and redesign the theme when:

- Three or more substantially different prompt formulations fail to produce stable results.
- Different model families prefer incompatible prompt formulations.
- False positives remain high after clear exclusions.
- Recall remains weak despite a reasonable positive definition.
- The theme requires extensive caveats to explain what the model is really doing.
- The positive class depends on subjective interpretation that reviewers cannot consistently adjudicate.

A prompt-design failure can be a theme-design failure. Narrow the theme, split it, or use AI Analyze for exploratory review instead of treating the output as final classification.

12. Overbroad Prompts and Corrected Versions

These examples show common failure modes and stronger alternatives.

12.1 Overbroad forwarded-content prompt

Weak:

```
Flag emails about government misconduct.
```

Problem: The model may flag newsletters, press articles, quoted third-party allegations, or routine monitoring emails.

Stronger:

```
You are a FOIA/public-records reviewer. Reply Responsive if participant-authored
content directly discusses, reports, authorizes, investigates, or responds to
alleged government misconduct within the request scope. Reply Unresponsive for
newsletters, press clips, automated alerts, public commentary, routine scheduling,
or forwarded/quoted third-party material where the actual email participants add no
substantive commentary. Reply INCONCLUSIVE if participant-authored content suggests
misconduct but lacks sufficient clarity or context. Otherwise, reply Unresponsive.
Reply with only one label: Responsive, Unresponsive, or INCONCLUSIVE.
```

12.2 Overbroad privilege prompt

Weak:

```
Classify emails involving lawyers as Privileged.
```

Problem: Lawyer involvement does not necessarily mean legal advice or work product.

Stronger:

```
You are a privilege reviewer. Reply Privileged when the email seeks, provides,
relays, or discusses confidential legal advice, legal strategy, or attorney work
product under the matter protocol. Reply Not Privileged for routine scheduling with
lawyers, billing logistics, public legal updates, legal department administration
without legal advice, or business advice copied to counsel without a legal-advice
purpose. Reply INCONCLUSIVE when lawyer involvement and confidentiality are present
but the legal-advice purpose cannot be determined reliably. Reply with only one
label: Privileged, Not Privileged, or INCONCLUSIVE.
```

12.3 Overbroad insider-threat prompt

Weak:

```
Flag suspicious emails about files, USB drives, personal email, competitors, or
resignation.
```

Problem: It is close to a keyword list and will over-collect benign references.

Stronger:

```
You are a digital forensics investigator. Reply Responsive if participant-authored
content shows unauthorized movement, collection, access, disclosure, or planned
misuse of company files, credentials, datasets, or confidential information. Reply
Unresponsive for routine file sharing, approved IT support, ordinary resignation
logistics, normal collaboration, or general cybersecurity awareness. Reply
INCONCLUSIVE if the email contains a concrete insider-threat signal but lacks
```

```
enough context to determine whether the conduct is authorized. Otherwise, reply
Unresponsive. Reply with only one label: Responsive, Unresponsive, or INCONCLUSIVE.
```

12.4 Overbroad legal-risk prompt

Weak:

```
Mark any email with legal words as High Risk.
```

Problem: It treats legal vocabulary as risk rather than evidence.

Stronger:

```
Assess legal review priority. Reply High when the email contains explicit
admissions, threatened litigation, regulatory action, clear policy violations,
privileged legal strategy, or statements directly harmful to the organization's
position. Reply Medium for ambiguous legal or compliance concerns requiring review.
Reply Low for routine business, scheduling, ordinary contract administration,
public legal updates, or legal vocabulary without apparent risk. Reply INCONCLUSIVE
if there is insufficient content to assess priority. Reply with only one label:
High, Medium, Low, or INCONCLUSIVE.
```

13. Practical Sample Workflows

13.1 First-run prompt validation workflow

Use this workflow for any new prompt.

```
Define issue and output labels
        ↓
Select task: Filter, Classify, or Analyze
        ↓
Adapt a library prompt or draft from a template
        ↓
Create readable source version and Aid4Mail single-line version
        ↓
Select validation sample with likely positives, likely negatives, borderline items,
forwarded threads, attachments, languages, and key custodians
        ↓
Run sample in Aid4Mail
        ↓
Export to HTML with portable viewer if browser review is useful
        ↓
Review all Responsive and INCONCLUSIVE results; spot-check Unresponsive results
        ↓
Record false positives, false negatives, category confusion, and prompt changes
        ↓
Revise prompt and rerun sample if needed
        ↓
Approve production prompt and preserve validation record
```

13.2 eDiscovery issue coding with privilege

Goal: Classify emails for a discrimination/retaliation matter while routing privileged material.

1. Apply deterministic filters first: date range, custodians, domains, duplicates, and folders.
2. Configure AI Classify with labels: `Privileged`, `Responsive`, `Unresponsive`, `INCONCLUSIVE`.
3. Use a prompt that gives privilege precedence over responsiveness.
4. Validate on known complaints, HR files, performance-management emails, counsel communications, and routine HR administration.
5. Review all `Privileged`, `Responsive`, and `INCONCLUSIVE` results.
6. Spot-check `Unresponsive`, especially from key custodians and relevant time periods.
7. Export classification as a field and preserve logs, prompt, category list, validation notes, and final output.

13.3 FOIA/public-records responsiveness and exemption workflow

Goal: Identify records responsive to a request and route possible holdbacks.

1. Define the request scope and legal holdback categories with counsel or the FOIA/public-records officer.
2. Apply deterministic filters where appropriate: agency unit, date range, known custodians, and file types.
3. Configure AI Classify with labels such as `Privileged`, `Exempt`, `Responsive`, `Unresponsive`, `INCONCLUSIVE`.
4. Add a forwarded-content rule, unless forwarded content itself is within the request scope.
5. Validate on press clips, newsletters, automated alerts, meeting logistics, likely responsive records, and likely exempt records.
6. Review all `Privileged`, `Exempt`, `Responsive`, and `INCONCLUSIVE` records.
7. Use HTML viewer searches such as `class:responsive`, `class:exempt`, and `class:inconclusive` for validation.
8. Export and archive final output and any tag files used during sample review.

13.4 Digital-forensics exfiltration triage

Goal: Find potential unauthorized data movement.

1. Use deterministic filters for date range, relevant custodians, personal domains, external storage domains, or known endpoints where appropriate.
2. Avoid using issue keywords as the sole gate unless recall impact is tested.
3. Configure AI Classify with `Responsive`, `Unresponsive`, `INCONCLUSIVE`.
4. Decide whether attachment text is needed. Start with attachment names if the issue is likely visible in the email body; include full extracted text if attachments may contain the data movement evidence.
5. Validate on known benign file sharing, ordinary IT support, departure logistics, personal email use, and likely exfiltration examples.
6. Review all `Responsive` and `INCONCLUSIVE` results. Spot-check `Unresponsive` from high-risk periods and custodians.

7. Consider a second model or senior investigator review for `INCONCLUSIVE` and high-risk `Responsive` items.

13.5 Analyze workflow for summaries and translations

Goal: Add reviewer-assistance fields to a review set.

1. Run AI Classify first if only a subset needs analysis.
2. Configure AI Analyze with a focused summary, translation, or extraction prompt.
3. Set a practical output-token limit.
4. Export to a format that supports AI fields: HTML, PDF, CSV, TSV, XML, or JSON.
5. Validate summaries for omissions and unsupported statements.
6. Validate translations on key passages where possible.
7. Validate extractions for both missed and incorrectly added entities.
8. Preserve the final AI analysis field as reviewer assistance, not as an independent fact record.

---

14. Defensibility and Preservation

A defensible AI-assisted review workflow should be understandable, repeatable in design, and preserved in output. Exact reruns may not always reproduce every classification because providers update models, some models are nondeterministic, and ambiguous emails sit near decision boundaries. Preserve the actual output from the production run.

14.1 Preserve the production record

For each production AI workflow, preserve:

Item	Record to preserve
Matter identification	Matter name, number, client/agency, review objective
Corpus source	Collection source, custodians, date range, folders, file types
Deterministic filters	Pre-acquisition and post-acquisition filters
AI task	Filter, Classify, Analyze, or combination
Prompt	Readable source version and Aid4Mail single-line version
Prompt version	Version number, date, author, change notes
Category list	Exact allowed categories, if restricted classification is used
Model and provider	Model name, provider, deployment path, cloud region or offline host
Local model details	Model tag/name, quantization, context length, local tool/version, endpoint, digest or hash where available
Attachments	Whether attachment text was included, size limit, and any known truncation

Validation sample	Sample size, design, known positives, rationale, reviewer notes
Acceptance criteria	False-positive tolerance, false-negative tolerance, `INCONCLUSIVE` handling, approver
Production run	Date/time, settings, logs, errors, throttling, run interruptions
Output	Final exported classifications or analysis fields
AI Filter records	Passed and rejected decision records, if Filter was used
Viewer tags	Exported tag file, if HTML viewer tags were used
Second-pass review	Model comparison notes, disagreement review, adjudication notes

14.2 Prompt version log

Use a simple versioning pattern.

Version	Date	Author	Change	Reason	Validation result	Approved by
v0.1			Initial draft			
v0.2			Added forwarded-content rule	False positives from quoted articles		
v1.0			Production version	Accepted validation sample		

14.3 Defensibility language for review notes

Use neutral, factual wording in matter records.

```
Aid4Mail AI Classify was used as an automated triage and issue-coding step. The
production prompt, model, provider, category list, attachment settings, validation
notes, processing logs, and exported AI classification field were preserved.
INCONCLUSIVE items were routed to human review. Responsive and INCONCLUSIVE sample
results were reviewed before production, and Unresponsive results were spot-checked
according to the documented validation plan.
```

Do not describe AI output as a final legal determination unless the workflow includes the necessary human review and legal sign-off.

15. Quality Control for Analyze Outputs

AI Analyze outputs are more open-ended than Filter or Classify labels. Review them accordingly.

15.1 Summaries

Check for:

- Omitted action items.
- Omitted dates or deadlines.
- Incorrect sender intent.
- Unsupported interpretation.
- Failure to distinguish quoted/forwarded text from participant-authored text.
- Over-compression of long threads.

15.2 Translations

Check for:

- Mistranslated legal, technical, or colloquial terms.
- Loss of tone, sarcasm, euphemism, or coded language.
- Overconfident interpretation where literal translation is uncertain.
- Missing names, dates, numbers, or attachments.

15.3 Entity extraction

Check for:

- Missed people, organizations, dates, deadlines, and amounts.
- Incorrectly inferred entities.
- Duplicates caused by signatures or quoted threads.
- Confusion between sender, recipient, quoted author, and third party.

15.4 Risk notes and explanations

Check for:

- Unsupported statements.
- Legal conclusions beyond the source text.
- Confusion between allegation and fact.
- Overstated certainty.
- Evidence location that does not match the email.

Treat Analyze output as reviewer assistance tied to a source email, not as independent evidence.

16. Known Limits and Cautionary Patterns

16.1 Subjective themes

Themes such as “damage control,” “concern about optics,” “off-record intent,” or “candid internal assessment” can be useful for exploratory review, but they may not produce stable classification results across reviewers or models.

Use subjective themes cautiously:

- Define the operational test narrowly.
- Validate with human reviewers.
- Expect more `INCONCLUSIVE` results.
- Consider using Analyze notes instead of hard classification.
- Preserve caveats in the validation record.

16.2 Absence-detection themes

Prompts that ask the model to infer what an email deliberately avoids saying are harder than prompts that detect what the email says. If the theme requires absence detection, try to narrow it to a concrete observable signal.

Weaker:

```
Flag emails where the sender is hiding something from the written record.
```

Stronger:

```
Reply Responsive if participant-authored content explicitly asks to move a
substantive discussion from email to a phone call, in-person meeting, private
channel, or deletion/restricted-distribution practice, and the email leaves the
substantive topic unresolved in writing.
```

Even the stronger version should be validated carefully.

16.3 Model disagreement

Disagreement between models does not automatically identify which model is right. It may indicate:

- The prompt is ambiguous.
- The theme is subjective.
- Categories overlap.
- The model is capability-limited.
- The prompt is too long or too short for the model.
- The evidence is genuinely borderline.

For high-risk matters, review disagreement sets rather than averaging them away.

16.4 Legal and policy limits

AI can help route privilege, exemption, privacy, sensitivity, and redaction issues, but legal standards must come from the matter protocol. Do not use generic cookbook prompts as legal advice or agency disclosure policy.

16.5 Sensitive investigations

For highly sensitive subject areas, use the library prompts only under appropriate investigative authority, safety procedures, chain-of-custody controls, and reviewer-support protocols. Keep prompts focused on classification and evidence routing; avoid unnecessary reproduction of harmful content in prompt examples or reports.

Appendix A: Prompt Readiness Checklist

Use before running a validation sample.

Matter objective

- [ ] The review objective is defined.
- [ ] The AI task is appropriate for the objective.
- [ ] Deterministic filters have been applied first where practical.
- [ ] Issue-keyword filters are not the sole AI gate unless recall impact has been validated.
- [ ] Required languages are known or sampled.
- [ ] Attachment strategy is defined.

Prompt design

- [ ] Prompt can be written on day one without corpus-mined positive phrases.
- [ ] Positive class is expressed as a single operational test.
- [ ] Responsive criteria are clear.
- [ ] Unresponsive criteria and exclusions are clear.
- [ ] Forwarded/quoted content rule is explicit.
- [ ] Attachment treatment is explicit if attachments matter.
- [ ] `INCONCLUSIVE` threshold is defined.
- [ ] Output labels are exact and stable.
- [ ] Category list is mutually exclusive or has precedence rules.
- [ ] Prompt length is appropriate for the task.
- [ ] Readable source version and Aid4Mail single-line version are preserved.

Validation setup

- [ ] Representative sample selected.
- [ ] Known or likely responsive examples included.
- [ ] Likely unresponsive examples included.
- [ ] Borderline examples included.
- [ ] Major custodians, periods, folders, languages, and attachment types represented.
- [ ] Acceptance criteria defined.
- [ ] Reviewer and approver assigned.

Appendix B: Validation Worksheet

Field	Entry
Matter name / ID	
Prompt name	
Prompt version	
AI task	Filter / Classify / Analyze
Provider / model	
Deployment path	Direct API / Enterprise cloud / Offline
Attachment text included	Yes / No
Attachment size limit	
Category list	
Sample size	
Sampling method	Random / Stratified / Deliberate / Mixed
Known-positive examples included	Yes / No / N/A
Languages represented	
Forwarded-thread examples included	Yes / No
Attachment examples included	Yes / No
Responsive count	
Unresponsive count	
INCONCLUSIVE count	
False positives found	
False negatives found	
Category-confusion notes	
Prompt changes required	
Second model used	
Acceptance criteria met	Yes / No
Approved by	
Approval date	

Appendix C: Category Precedence Worksheet

Use this worksheet when categories can overlap.

Conflict	Precedence rule	Rationale	Reviewer notes
Privileged + Responsive	Choose Privileged	Privilege routing takes priority	
Exempt + Responsive	Choose Exempt	Holdback review takes priority	
Exfiltration + Compliance	Choose Exfiltration	More specific investigative category	
Financial + Corruption	Choose Corruption only if improper influence or bribery is explicit; otherwise Financial	Separates illicit finance from bribery/influence	
Multiple issue labels			
Issue label + Clean	Issue label overrides Clean	Clean applies only when no issue category applies	
Ambiguous issue category	INCONCLUSIVE	Human review required	

Appendix D: One-Line Prompt Conversion Examples

D.1 Readable version

```
You are an experienced eDiscovery reviewer. Classify the email into exactly one
category.

Reply Responsive if participant-authored content discusses potential anti-
competitive coordination, price fixing, bid rigging, market allocation, customer
allocation, no-poach arrangements, or improper competitor information exchange.

Reply Unresponsive for routine sales, ordinary market commentary, scheduling,
logistics, newsletters, or forwarded articles without substantive participant
commentary.

Reply INCONCLUSIVE if competition-related signals exist but the email lacks
sufficient clarity or context for reliable classification.

Reply with only one label: Responsive, Unresponsive, or INCONCLUSIVE.
```

D.2 Aid4Mail single-line version

```
You are an experienced eDiscovery reviewer. Classify the email into exactly one category. Reply Responsive if participant-authored content discusses potential anti-competitive coordination, price fixing, bid rigging, market allocation, customer allocation, no-poach arrangements, or improper competitor information exchange. Reply Unresponsive for routine sales, ordinary market commentary, scheduling, logistics, newsletters, or forwarded articles without substantive participant commentary. Reply INCONCLUSIVE if competition-related signals exist but the email lacks sufficient clarity or context for reliable classification. Reply with only one label: Responsive, Unresponsive, or INCONCLUSIVE.
```

D.3 Conversion checks

- [ ] No meaning was lost when line breaks were removed.
- [ ] Output labels are unchanged.
- [ ] Punctuation separates label rules clearly.
- [ ] Category precedence remains understandable.
- [ ] The prompt still fits the selected task.

Appendix E: Prompt Version Log Template

Version	Date	Author	Prompt summary	Change from prior version	Validation result	Approved for production
v0.1				Initial draft		
v0.2						
v1.0				Production version		

Appendix F: Source Map

This guide synthesizes the following project source material.

Source file	Material incorporated
`Aid4Mail_AI_Integration_User_Guide.md`, §8	Prompt-library access, organization by Filter/Classify/Analyze, theme families, and guidance to customize, verify, and test prompts on a small sample

`Aid4Mail_AI_Email_Review_Workflow_Guide.md`	Filter/Classify/Analyze distinctions, first-run workflow, prompt anatomy, three-label pattern, validation workflow, attachment strategy, HTML viewer validation, defensibility checklist, troubleshooting patterns
`Prompt_Development_Guidelines_for_Aid4Mail_Benchmark_Tests.md`	Day-one writability, foreknowledge-contamination guardrails, theme viability pre-screen, positive framing, length discipline, one operational test, decisive exclusions, prompt-model fit, false-positive analysis, and when to stop iterating
`AI-Prompt-Examples.md`	Classification, filtering, and analysis prompt examples for eDiscovery, digital forensics, and FOIA/public-records workflows
`AI Prompt Themes.md`	Theme inventory used to organize cookbook recipes and practice-area examples
`Benchmark_Methodology.md`	Benchmark-tested prompt shapes, including insider-threat binary classification, multi-category misconduct classification, Korean-language variants, and FOIA-style participant-authored evidence rules
`Podesta_Corpus_Benchmark_Methodology_Note.md`	Cautions about subjective themes, inter-model disagreement, absence-detection tasks, and the limits of using subjective naturalistic corpora as ground truth

`Aid4Mail Email Viewer - User Guide.md`	HTML portable viewer validation tactics, Classification column review, `class:` and `tag:` search syntax, and tag export/import for preserving validation or review decisions

Appendix G: Theme Inventory for Cookbook Planning

The prompt library contains many themes. Use this inventory to decide whether to adapt a library prompt, draft a new prompt, or split a broad issue into several focused passes.

eDiscovery

Antitrust and Competition; Breach of Contract; Compliance Violation; Corporate Espionage; Corporate IP Theft; Corruption and Bribery; Discrimination or Retaliation; Financial Crimes; Fraud Detection; HR Investigation; Insider Threat; Legal Risk Level; Legal Sensitivity; Mergers and Acquisitions; Privilege Classification; Regulatory Compliance Assessment; Securities Litigation; Sexual Harassment; Trade Secret Misappropriation; Unauthorized Access or Data Breach; Urgency and Priority Detection; Wage and Hour Disputes; Whistleblower Report; Wrongful Termination.

Digital Forensics

Blackmail and Extortion; Business Email Compromise; Credential Theft; Crypto Fraud and Illicit Transactions; Cybersecurity Threat; Data Exfiltration; Deepfake and Synthetic Media Crimes; Drug Trafficking; Email-Based Attack Vectors; Espionage; Evidence of Policy Evasion; Fake News and Coordinated Misinformation Campaigns; Financial Fraud; Human Trafficking; Identity Theft and Document Forgery; Insider Threat; IP Theft; Malware Distribution; Missing Person; Money Laundering; Online Radicalization and Extremist Recruitment; Phishing and Social Engineering; Romance and Financial Scams; Sextortion and Online Exploitation; Sexual Harassment; Spam and Malware; Stalking, Cyberbullying, and Digital Harassment; Suspicious Attachments or Links; Swatting and Online Threats; Terrorism-Related Activity; Victim Communication and Distress Signals; Workplace Violence Threats.

Highly sensitive digital-forensics themes should be used only under appropriate legal authority, evidence-handling protocols, and reviewer-safety procedures.

FOIA/Public Records

Education Policy; Environmental Regulation and Impact; Financial Regulation and Oversight; Government Contracting and Procurement; Government Misconduct; Health Policy and Public Health; Law Enforcement Practices; Lobbying and Political Influence; Personnel and

Hiring Practices; Public Interest; Scientific Misconduct; Surveillance and Data Privacy; Travel and Expense Reports; Whistleblower Communications and Retaliation.

Date of publication: May 15, 2026.