The Law Firm AI Vendor Questionnaire: 38 Questions, Acceptable Answers, and Red Flags

This article provides general educational information, not legal advice. Firms should apply the professional conduct rules of the lawyer's licensing jurisdiction, applicable privacy and security laws, court orders, client instructions, and contractual requirements.

An AI Vendor Cannot Assume the Lawyer's Ethical Duty

Law firms often ask whether an AI company is "ABA compliant." That question is understandable, but it assigns responsibility to the wrong party.

The ABA Model Rules govern lawyer conduct. They serve as models for the rules adopted by individual jurisdictions. An AI provider can offer security controls, contractual protections, audit evidence, and appropriate system architecture. It cannot certify the lawyer's use of the product as ethical in every matter.

Model Rule 1.6 requires reasonable efforts to prevent unauthorized access to or disclosure of information relating to a representation. Model Rule 1.1 and its technology-competence comment require lawyers to understand the benefits and risks associated with relevant technology. Model Rule 5.3 requires appropriate supervision of nonlawyers and outside service providers.

Formal Opinion 512 makes the vendor-review obligation particularly concrete. Among other things, it advises lawyers to investigate an AI tool's security, reliability,

retention practices, proprietary-rights claims, contractual protections, breach-notification procedures, and exposure to cyberattack. It also notes the importance of confidentiality agreements, vendor credentials, security protocols, conflicts safeguards, and an available legal forum for enforcing the agreement.

The practical standard is therefore not, “Did the vendor say it was secure?”

The standard is, “Did the firm conduct and document a reasonable investigation before placing client information at risk?”

Confidentiality Is Broader Than Attorney-Client Privilege

Attorney-client privilege is an evidentiary doctrine. Model Rule 1.6 confidentiality is broader. It applies to information relating to a representation, whatever its source, and also covers disclosures that could allow another person to infer protected information.

An AI deployment can therefore create several separate concerns:

- Professional confidentiality under the applicable ethics rules
- Attorney-client privilege and work-product protection
- Privacy and data-security laws
- Protective orders, sealing orders, and litigation holds
- Client engagement terms and outside counsel guidelines
- Contractual confidentiality obligations
- Sector-specific requirements involving health, financial, employment, or government information

A vendor questionnaire should be reviewed through all of those lenses.

How to Use This Questionnaire

Require written answers. A sales call is not enough.

Ask the vendor to identify the document supporting each answer, such as a system architecture diagram, permission manifest, data-processing addendum,

subprocessor list, retention policy, SOC 2 report, ISO certificate, penetration-test summary, incident-response plan, or sample audit log.

A strong answer has four characteristics:

1. It is specific.
2. It can be technically verified.
3. It appears in the contract or incorporated security documentation.
4. It applies to all relevant subprocessors, including the underlying AI model provider.

A vague answer such as “we follow industry best practices” should be treated as an unanswered question.

I. Architecture and Data Flow

1. Does the system process email inside Microsoft 365, or does information leave the tenant?

A strong answer: The vendor provides a complete data-flow diagram identifying every point at which information is collected, transmitted, processed, temporarily stored, logged, backed up, or made available for support. If information leaves Microsoft 365, the answer identifies each receiving system and its purpose.

Red flag: “The data remains secure in our cloud,” without identifying the cloud, model provider, temporary storage, or logging systems.

2. Exactly what information is transmitted?

A strong answer: The vendor separately identifies email bodies, subjects, headers, sender and recipient information, attachments, contact information, calendar data, mailbox-folder structure, prompts, responses, summaries, and system metadata. The system transmits only what the approved function requires.

Red flag: The agent collects the entire mailbox because doing so is simpler for the vendor, even though the assigned task requires only a subset.

3. Which AI provider, product, model, and commercial service tier are used?

A strong answer: The vendor identifies the exact model endpoint, product tier, account type, and applicable provider terms. Firm information is not routed through a consumer chat account. The vendor discloses whether it can change models automatically and under what conditions.

Red flag: The vendor refuses to identify the downstream provider, calls the information proprietary, or reserves the right to route data through any model without notice.

4. Where is information processed, stored, backed up, and accessed?

A strong answer: The contract identifies the permitted countries or regions for primary processing, backups, support, and subprocessors. Transfers outside the agreed region require advance notice and, where appropriate, firm approval.

Red flag: "We use a global infrastructure," with no ability to identify where client information travels.

5. Does the system create summaries, embeddings, indexes, caches, or persistent memory?

A strong answer: All derived information is treated as customer data. It is tenant-isolated, subject to the same confidentiality and security obligations as the original email, included in retention schedules, and capable of being deleted.

Red flag: The vendor claims that embeddings, classifications, or summaries belong to the vendor or are not confidential because they are "derived data."

II. Microsoft 365 Permissions and Identity Security

Microsoft advises application developers to request the least-privileged Graph permissions needed for the specific function. Exchange Online Application RBAC can also limit an application's access to a defined set of mailboxes.

6. What exact Microsoft Graph or Exchange permissions does the application request?

A strong answer: The vendor supplies the full permission manifest and explains why each permission is necessary. Permissions are mapped to specific product functions.

Red flag: The vendor requests global administrator rights or broad mail permissions “to make installation easier.”

7. Can access be limited to a named mailbox or an approved mailbox group?

A strong answer: Yes. The system supports Exchange Online Application RBAC or an equivalent technical control that denies access to all mailboxes outside the approved scope. A pilot can be limited to one mailbox.

Red flag: The vendor insists that tenant-wide mailbox access is technically unavoidable without demonstrating why mailbox-level scoping cannot be used.

8. Which actions can the application perform?

A strong answer: Read, move, label, modify, delete, send, forward, and mailbox-rule permissions are separated. The application receives only the actions required for the approved workflow.

For an email-cleanup pilot, the preferred configuration is read-only. If write access later becomes necessary, it should be limited to approved actions such as moving messages into a review folder. Sending, forwarding, changing rules, and permanent deletion should remain disabled.

Red flag: All rights are bundled into one permission set.

9. How does the application authenticate?

A strong answer: The connector uses a managed identity, service principal, or certificate-based application identity. It does not use a shared administrator account or store a global administrator password. Credentials are protected in a secure secrets system, rotated, monitored, and promptly revocable.

Red flag: The vendor asks the firm to create a permanent user account with administrator rights and a password that never expires.

10. How are human and application identities protected?

A strong answer: Human administrators use single sign-on and multifactor authentication. Application identities are governed through workload-identity controls, certificate management, location or network restrictions where available, token monitoring, and periodic access reviews.

Service principals do not authenticate in the same manner as human users, so asking only whether "MFA is supported" is incomplete. Microsoft provides separate Conditional Access capabilities for workload identities.

Red flag: Human support accounts lack MFA, or the firm has no way to revoke the application's access immediately.

11. Does the system honor the firm's internal security and information-governance controls?

A strong answer: The application respects mailbox permissions, restricted-user groups, ethical walls, retention policies, litigation holds, journaling, and relevant data-loss-prevention or sensitivity controls. It cannot bypass those controls through its own storage or indexing system.

Red flag: The product copies all messages into a separate index that ignores the firm's restricted matters, legal holds, or Microsoft 365 retention controls.

III. Retention, Deletion, and Secondary Use

The ABA does not prescribe a universal number of hours or days for AI retention. For sensitive law firm email, zero retention or the shortest operationally necessary period is the preferred baseline. Any retention should be known, justified, protected, and enforceable.

12. How long is each category of information retained?

A strong answer: The vendor provides exact maximum periods for original content, attachments, prompts, responses, summaries, embeddings, indexes, application logs, security logs, error records, support tickets, and backups.

Red flag: “We keep data only as long as necessary,” without defining who decides what is necessary or identifying a maximum period.

13. What does deletion actually remove?

A strong answer: Deletion propagates to active storage, indexes, caches, replicas, and derived data. The vendor identifies the separate expiration schedule for backups and explains whether deleted data can ever return to the active environment during a restoration.

Red flag: Deleting an account merely removes the user interface while the vendor continues to retain the underlying information.

14. Is firm information used to train, fine-tune, improve, test, or evaluate any model?

A strong answer: No, unless the firm separately and affirmatively opts in. The prohibition is contractual and applies to the vendor and every model provider or subprocessor.

The answer should address training, fine-tuning, model evaluation, quality review, product improvement, benchmark creation, abuse monitoring, and human review separately.

Red flag: “We do not currently train on your data,” “we may use data to improve our services,” or an opt-out that applies only to the vendor but not to the underlying model provider.

15. Does the vendor use aggregated, anonymized, or de-identified information?

A strong answer: Any permitted telemetry is limited to non-content operational metrics, does not identify a client, matter, lawyer, correspondent, or mailbox, and cannot reasonably be reidentified. The contract prohibits sale and reidentification.

Red flag: A broad license allows the vendor to use “aggregated data” without defining what is removed or whether email metadata and matter patterns remain.

16. What happens when the agreement or pilot ends?

A strong answer: The firm can revoke access immediately, export required records and logs, request deletion, and receive written confirmation when deletion is complete. The vendor identifies when residual backup copies will expire.

Red flag: The vendor retains customer content or embeddings indefinitely after termination or cannot certify deletion.

IV. Human Access, Subprocessors, Ownership, and Segregation

17. Which vendor personnel can access firm information?

A strong answer: There is no routine employee access. Exceptional access is role-based, time-limited, approved, associated with a support or security ticket, and logged. Personnel receive security and confidentiality training and, where lawful and appropriate, screening.

Red flag: Any engineer or support representative can query customer content in a production database.

18. Which third parties and subprocessors receive or can access the information?

A strong answer: The vendor provides a complete list covering the AI model provider, cloud host, logging and monitoring platforms, customer-support systems, analytics providers, backup services, and affiliated companies. The list identifies the function, data involved, and processing region.

Red flag: The vendor names only its primary cloud provider while omitting the model API, logging service, and offshore support organization.

19. Are subprocessors bound by equivalent obligations?

A strong answer: The vendor performs due diligence, contractually imposes comparable confidentiality, security, retention, deletion, and incident-notification obligations, and remains responsible for subprocessor performance.

Red flag: The contract states that the vendor has no responsibility for the practices or security failures of third-party providers it selected.

20. Who owns the input, output, summaries, and derived data?

A strong answer: The firm retains all rights in its data. The vendor receives only a limited license to process information for the contracted service. It claims no ownership, lien, training right, or continuing license over prompts, outputs, summaries, embeddings, or indexes.

Formal Opinion 512 specifically advises lawyers to determine whether a provider retains submitted information or asserts proprietary rights over it.

Red flag: A perpetual, irrevocable, sublicensable license covering customer content or derived information.

21. How does the system prevent disclosure across customers or matters?

A strong answer: Customer data, vector stores, indexes, memory, encryption boundaries, and access controls are tenant-isolated. The vendor tests for cross-tenant retrieval and leakage. Within the firm, the system supports restricted matters and ethical walls.

Red flag: Multiple customers share a single retrieval index, shared memory, or customer-trained model without demonstrable isolation.

22. What happens if the vendor receives a subpoena, warrant, or other demand?

A strong answer: The vendor promptly notifies the firm unless legally prohibited, provides the demand, allows the firm to seek protection, challenges overbroad requests where appropriate, and discloses only what is legally required.

Formal Opinion 512 identifies notice concerning service of process for client information as an important vendor safeguard.

Red flag: The vendor may disclose information without notice whenever it believes disclosure is appropriate.

V. Security Controls and Evidence

A certification is evidence, not a verdict. The scope, exceptions, testing period, and relevance to the actual product all matter.

23. How is information encrypted, and who controls the keys?

A strong answer: Information is protected in transit using modern transport encryption and at rest using strong encryption, commonly TLS 1.2 or later and AES-256 or an equivalent control. The answer covers primary databases, logs, backups, indexes, and temporary storage. Keys are separated, access-controlled, rotated, and monitored.

For especially sensitive matters, the firm may also ask about customer-managed keys or dedicated encryption boundaries.

Red flag: "Everything is encrypted," with no information about protocols, backups, logs, or key access.

24. What independent assurance documentation is available?

A strong answer: The vendor provides, under appropriate confidentiality restrictions, a current SOC 2 Type II report and any relevant ISO/IEC 27001 certification. The scope clearly covers the service being purchased. Material exceptions, control failures, and remediation are disclosed.

Red flag: A certification logo on the website, an outdated report, or a report covering only the parent company's corporate network rather than the AI product.

25. How is the product tested and maintained?

A strong answer: The vendor has a secure development lifecycle, code review, dependency scanning, secret scanning, vulnerability management, patch deadlines based on severity, independent penetration testing, and a coordinated vulnerability-disclosure process.

For an AI agent, testing also addresses prompt injection, cross-tenant retrieval, unauthorized tool use, and data exfiltration.

Red flag: No independent penetration test, no remediation summary, or a claim that the underlying model provider handles all security.

26. What audit logs are available to the law firm?

A strong answer: Logs identify the mailbox, message or record identifier, timestamp, application identity, human identity where applicable, action taken, permission changes, support access, model or workflow version, and outcome. Logs can be exported and retained for the firm's required period.

Logs should permit reconstruction of events without creating unnecessary additional copies of complete message content.

Red flag: The only available records are user sign-in logs or monthly usage totals.

27. How does the system detect unusual or unauthorized behavior?

A strong answer: The service monitors unexpected access volume, new locations, unusual mailbox activity, repeated failures, abnormal exports, unauthorized tool calls, and suspicious data movement. The firm can receive alerts and disable the connector through an emergency kill switch.

Red flag: The vendor relies solely on post-incident customer reports and offers no immediate method to stop access.

ABA Formal Opinion 483 emphasizes reasonable efforts to monitor external vendors and connected technology rather than leaving breach discovery to chance.

VI. AI Agent Safety and Email-Specific Controls

An AI agent is not merely producing text. It may be authorized to search, classify, move, delete, send, or disclose information. That makes its tool permissions as important as the model itself.

28. How does the system defend against indirect prompt injection?

A strong answer: The system treats email bodies and attachments as untrusted content, not as instructions. System instructions are separated from retrieved information. Tool calls are independently authorized. High-risk actions require

approval. The vendor regularly tests attacks involving malicious content embedded in retrieved documents.

NIST warns that indirect prompt injection can cause an AI-enabled application to behave in unintended ways when it retrieves adversarial content.

Red flag: "The model is intelligent enough to recognize malicious instructions," without technical controls or testing.

29. Can the agent browse external links, open attachments, or call other systems?

A strong answer: External network access is disabled by default or limited to an allowlist. Attachments are scanned and processed in a controlled environment. Retrieved content cannot expose system credentials, invoke arbitrary tools, or upload information to unapproved destinations.

Red flag: The agent can follow any link in an email, connect to arbitrary websites, or invoke third-party tools without separate authorization.

30. Which actions require human approval?

A strong answer: Sending, forwarding, exporting, permanently deleting, changing mailbox rules, creating forwarding rules, sharing attachments, and altering retention settings require explicit approval by an authorized person.

Red flag: The system is marketed as "fully autonomous" and cannot place human approval gates around destructive or disclosure-related actions.

31. Does the cleanup process support a dry run, quarantine, and rollback?

A strong answer: The system first produces a proposed action report. Messages are moved into a review or quarantine folder rather than permanently deleted. Batch limits, confidence thresholds, approval screens, and a defined restoration period are available.

Red flag: The agent permanently deletes messages in real time with no preview or reliable reversal process.

32. How has classification and summarization accuracy been validated?

A strong answer: The vendor provides testing results relevant to the intended task, including false-positive and false-negative rates where meaningful. Results link back to source messages. AI summaries supplement rather than replace the original record. The firm can test a representative sample before broader deployment.

Red flag: General claims such as "95 percent accurate" without defining the test population, task, error type, or consequences.

Model Rule 1.1 and Formal Opinion 512 require lawyers to understand a tool's limitations and apply appropriate review rather than relying uncritically on AI output.

VII. Incident Response, Continuity, and Contract Terms

33. How does the contract define a security incident?

A strong answer: The definition includes unauthorized access, disclosure, loss, improper retention, cross-customer leakage, use for training contrary to contract, credential compromise, prompt-injection exploitation, unauthorized agent actions, and incidents involving subprocessors.

Red flag: An incident exists only after the vendor confirms that a network intruder successfully extracted data.

34. When will the firm be notified?

A strong answer: Notice is required without undue delay, preferably within 24 hours after discovery of a suspected or confirmed incident affecting firm information. The vendor does not wait for a final forensic report before providing initial notice and regular updates.

The ABA does not prescribe a particular vendor deadline. A short contractual deadline gives the law firm time to investigate its own duties under professional-conduct rules, privacy laws, contracts, and court orders.

Red flag: Notice is provided only "when required by law," only after the incident is fully confirmed, or weeks after discovery.

35. What information and cooperation will the vendor provide?

A strong answer: The vendor preserves evidence and supplies known dates, affected systems, data categories, affected customers, indicators of compromise, containment steps, forensic findings, root-cause analysis, remediation, and continuing updates. It cooperates with the firm's investigators and reasonable client or regulatory response.

Formal Opinion 483 explains that lawyers must make reasonable efforts to determine what occurred and advise affected clients of the known or reasonably ascertainable extent of the compromise.

Red flag: A generic notification stating that the vendor "takes security seriously," with no records allowing the firm to determine which client information was affected.

36. What happens if the service fails or becomes unavailable?

A strong answer: The vendor maintains and tests business-continuity, backup, recovery, and fail-safe procedures. It identifies recovery-time and recovery-point objectives. An interrupted agent stops safely and does not blindly repeat queued actions when service resumes.

Red flag: No tested recovery plan, or destructive actions may replay after an outage.

37. Does the contract create meaningful accountability?

A strong answer: The agreement includes confidentiality obligations, a data-processing and security addendum, audit or assurance rights, appropriate cyber and technology-errors-and-omissions insurance, incident cooperation, and enforceable remedies.

The firm should review indemnification, liability caps, exclusions, venue, governing law, and carveouts for confidentiality, security, intellectual-property, and willful misconduct with qualified counsel.

Red flag: Liability for a major confidentiality incident is capped at one month of subscription fees, with no meaningful carveout, while the vendor disclaims responsibility for all subprocessors.

38. Can the vendor materially change the service without the firm's approval?

A strong answer: The vendor provides advance notice before changing the model provider, processing region, retention period, data use, subprocessor list, requested permissions, or material security controls. The firm may object, suspend processing, or terminate and require deletion.

NIST's Generative AI Profile recommends ongoing vendor assessment, subprocessor inventories, contract provisions addressing third-party processes, incident-response commitments, and review of secondary data uses and liability terms.

Red flag: The vendor may change all terms and providers immediately, with continued use treated as automatic acceptance.

Minimum Acceptable Design for an Email-Cleanup Agent

For the email-cleanup scenario that prompted this article, a defensible pilot should ordinarily include these controls:

1. Access is limited to the single approved mailbox, not every mailbox in the tenant.
2. The initial phase is read-only.
3. The agent cannot send, forward, create rules, export content, or permanently delete messages.
4. Proposed actions are presented for human approval.
5. Messages are moved to a quarantine or review folder with a reliable rollback period.
6. Litigation holds, retention requirements, journaling, and restricted matters remain effective.
7. The exact downstream model and service tier are disclosed and contractually prohibited from training on firm information.

8. Original and derived data are retained only for a documented, minimal period.
9. Email and attachment content are treated as untrusted input for prompt-injection purposes.
10. The firm receives actionable audit logs and can revoke access immediately.
11. All subprocessors are identified and contractually bound.
12. The agreement provides prompt incident notice, forensic cooperation, and verified deletion at the end of the project.

A request for tenant-wide read, write, send, forwarding, and permanent-deletion rights should pause the project unless the vendor can demonstrate why each permission is indispensable and why a narrower architecture cannot work.

Automatic No-Go Indicators

Regardless of a vendor's overall score, the following conditions deserve a stop decision until corrected:

- The vendor cannot identify the underlying AI provider or exact product tier.
- A single-mailbox project requires unrestricted access to the entire tenant.
- Client information may be used for training, evaluation, or product improvement without a separate opt-in.
- Retention periods are undefined or deletion cannot be verified.
- The vendor will not identify subprocessors.
- The system permits autonomous sending, forwarding, exporting, or permanent deletion.
- The vendor has no meaningful prompt-injection controls.
- The firm cannot obtain access and action logs.
- Incident notice is discretionary or materially delayed.
- The vendor can materially change data practices without notice or termination rights.

Questions the Firm Must Answer Internally

Vendor diligence is only one part of the analysis. Before deployment, the firm should also determine:

- Whether state ethics rules or opinions impose additional requirements

- Whether client engagement terms or outside counsel guidelines restrict AI use
- Whether the affected mailbox contains information subject to protective orders, legal holds, or special contractual restrictions
- Whether informed client consent is required for the proposed tool and data flow
- Whether restricted matters and ethical walls can be maintained
- Who within the firm may approve, monitor, suspend, and reauthorize the system
- How frequently the vendor, permissions, model, and terms will be reassessed

Formal Opinion 512 states that informed consent may be required before entering client information into certain self-learning AI tools that create a risk of later disclosure. When consent is required, generic engagement-letter language is not enough. The client must receive meaningful information about the proposed use, benefits, information involved, and risks.

Suggested Instructions to Send With the Questionnaire

Please answer each question in writing. Identify the policy, contract provision, configuration, report, or other evidence supporting the response. Where a requested control is not available, explain the limitation, the reason for it, and any compensating control. Please identify which answers are contractually binding and which are subject to change during the agreement.

Final Takeaway

AI can help a law firm bring order to years of email. It should not receive more authority than the job requires.

A vendor's sales presentation is an opening statement. The permission manifest, data-flow diagram, model-provider terms, security evidence, audit logs, and contract are the exhibits.

Before an AI agent enters the firm's inbox, the firm should know exactly what the agent can see, exactly what it can do, exactly where the information will go, and exactly who will answer when something goes wrong.