SAFEGUARD

Kingfisher Advisory

WhiteCloud Accounting

SafeGuard Integration Assessment

Independent review of an existing AI-enabled client intake workflow

November 2026

Prepared for WhiteCloud Accounting

ILLUSTRATIVE SCENARIO | NOT A CLIENT ENGAGEMENT OR VERIFIED TEST

Risk at a Glance

The key findings and the decision they support

6	3	3	0
FINDINGS	HIGH	MEDIUM	LOW

EXECUTIVE DECISION

HOLD automatic client emails

Keep staff in control while the three High findings are corrected and verified.

What the colors mean

HIGH	Fix before any automatic client emails resume.
MEDIUM	Address before expanding the system.
LOW	Schedule routine improvements. None identified.

The six findings

HIGH	SG-01	The system can access unrelated client files
HIGH	SG-02	Extra client notes are sent to the AI service
HIGH	SG-03	The system can email clients without staff approval
MEDIUM	SG-04	No clear owner or tested emergency stop
MEDIUM	SG-05	Client notes can change what the AI writes
MEDIUM	SG-06	Errors and duplicate reminders are hard to track

Severity shows how urgently action is needed. It is not a compliance score.

EXISTING AI INTEGRATION

Executive decision

RECOMMENDATION

Do not allow the system to email clients automatically until three serious issues are fixed.

Staff should continue handling client intake and reviewing all outgoing messages until the fixes are tested and approved by management.

The situation

WhiteCloud Accounting is a 24-person accounting practice. An outside provider has connected its Microsoft 365 tools with an AI service to help prepare client intake reminders. The system is being used by a small internal team. WhiteCloud wants to know whether client information is properly protected and whether staff remain in control of outgoing messages.

What leadership needs to decide

Keep automatic emails off

Owner: Managing partner and IT provider. **Proof:** Demonstrate that no email can go out without recorded staff approval.

Limit the client information used

Owner: Operations lead and IT provider. **Proof:** Show that the system can access and share only the information needed for reminders.

Approve a controlled restart

Owner: Managing partner. **Proof:** Review successful test results, confirm who oversees the system, and approve the restart in writing.

REVIEW BOUNDARIES

What we reviewed

The purpose is to examine this connected workflow, not every system used by WhiteCloud or the AI provider.

Included

- The application identity and which SharePoint files it can reach.
- Information passed through Power Automate to the AI service.
- Staff approval before sending emails.
- How errors, duplicate events, and shutdown are handled.

Not included

- A full Microsoft 365 security audit.
- Testing the AI provider's entire platform.
- Legal advice about tax or privacy obligations.
- Changing systems, operating the workflow, or guaranteeing that it is secure.

What counts as evidence

Evidence records (E-series)
Example permissions, configuration exports, process records, and message data that an assessor would request.

Test records (T-series)
Controlled checks that show what worked, what failed, and what could not be verified.

Assessment limitation
The conclusions apply only to the agreed systems, available information, and checks carried out.

NIST-informed means we use established guidance to structure the review. It is not certification or a guarantee.

HOW THE SYSTEMS CONNECT

Where the risks arise

The provider built a workflow that reads intake records, prepares reminders with AI, and sends messages through Outlook. The critical controls sit between those systems.

Connected step	What should happen	What we found
SharePoint access	Read only approved intake records.	The connection can also reach unrelated client files. **SG-01**
Information sent to AI	Send only the fields required to draft reminders.	Unnecessary client notes are included. **SG-02**
Outgoing email	A staff member approves the exact draft and recipient.	A path can send without a recorded approval. **SG-03**
Ownership and input	Staff can stop the workflow, and client notes cannot become instructions.	Shutdown is untested; notes can change drafts. **SG-04, SG-05**
Errors and retries	Identify repeated requests and assign failed ones to a person.	Duplicate drafts and unassigned errors can occur. **SG-06**

Immediate priority: Keep automatic client emails off while the issues involving file access, information sent to AI, and staff approval are corrected.

The provider should confirm the actual configuration and current software behavior before making changes.

HIGH

SG-01

The system can access unrelated client files

What we found

The intake automation can read SharePoint files outside the workspace it needs to prepare reminders. Its Microsoft Graph permission is broader than the task requires.

Why it matters

A mistake, misuse, or compromised connection could expose unrelated client records. That could trigger an investigation, client notifications, and damage to trust.

RECOMMENDED ACTION

Keep automatic emails paused. Ask the IT provider to restrict access to the approved intake workspace and confirm that unrelated files are no longer reachable.

How we will know it is resolved

Attempts to open unrelated sites and files are denied and recorded. The normal intake-status lookup still works.

Evidence and test results

E-01: A constructed export for **wc-intake-reminders** shows Microsoft Graph **Sites.Read.All**. The ClientArchive test site is outside the approved workspace.

T-01: A test read of ClientArchive returned **HTTP 200** and a synthetic file. The permitted intake read also succeeded. The access-boundary test failed.

Technical notes for the provider

Use appropriately scoped site permissions, check effective application and delegated access, remove broad grants, and repeat denied-access tests.

Framework references: NIST CSF 2.0 PR.AA-05; AI RMF MAP and MANAGE.

HIGH

SG-02

Extra client notes are sent to the AI service

What we found

The reminder workflow sends free-text client notes to the AI service, even though it needs only intake status and an approved reminder category.

Why it matters

Private information unrelated to the reminder can leave the approved workflow. That creates unnecessary exposure without improving the reminder.

RECOMMENDED ACTION

Agree on exactly which client details are permitted. Ask the provider to prevent all other fields from being sent to the AI service.

How we will know it is resolved

Captured requests, including retries and error cases, contain only approved fields. Unapproved notes and test markers do not reach the AI service.

Evidence and test results

E-02: An example request for case **WC-1042** includes **case_ref**, **status**, and a **client_notes** field referring to a bank account update.

T-02: Marker **TEST-DO-NOT-SEND-42** appeared in a captured AI request. The data-minimization check failed.

Technical notes for the provider

Enforce an allowed-field list before invoking the model. Remove unnecessary notes and review logging, error, and retry paths.

Framework references: AI RMF MAP 1.1 and MANAGE; NIST AI 600-1 privacy considerations.

HIGH

SG-03

Emails can be sent without staff approval

What we found

One Power Automate path can send a drafted reminder through Outlook without a recorded employee approval tied to the final message and recipient.

Why it matters

An incorrect or misdirected message could go to a client under WhiteCloud's name without anyone authorizing it.

RECOMMENDED ACTION

Turn off automatic sending. Require a named staff member to approve the exact message and recipient before Outlook can send it.

How we will know it is resolved

Messages with missing, expired, repeated, or mismatched approvals cannot send. Approved synthetic messages send once and record the reviewer.

Evidence and test results

E-03: The constructed flow connects the AI response to **Send_an_email_(V2)**. The approval action is optional on that path.

T-03: A synthetic email reached an internal test mailbox without a reviewer identity or approval record. External client delivery was not tested. The approval check failed.

Technical notes for the provider

Separate draft and send privileges. Bind an auditable, expiring approval to the message content, case, and recipient.

Framework references: AI RMF GOVERN and MANAGE; NIST CSF 2.0 PR.AA-05.

MEDIUM SG-04

There is no clear owner or tested emergency stop

What we found

The team has not named who owns the integration or practiced how to stop and safely restart it.

Why it matters

If the system behaves unexpectedly, staff may lose time deciding who can shut it down and how to recover.

RECOMMENDED ACTION

Assign a named owner and rehearse stopping the integration before expanding its use.

How we will know it is resolved

The designated person can stop processing and sending in a controlled test and restore operation only with documented approval.

Evidence and test results

E-04: The constructed service inventory has no values for **service_owner**, **rotation_date**, or **disable_procedure**.

T-04: In a simulated shutdown discussion, staff could not identify who could revoke the connection. A technical stop test was **not** carried out; this is an ownership gap, not a proven technical failure.

Technical notes for the provider

Document credentials, dependencies, revocation steps, recovery conditions, and notification responsibilities. Practice the stop procedure.

Framework references: NIST CSF 2.0 GV.RR-02; AI RMF GOVERN and MANAGE.

MEDIUM SG-05

Client notes can change what the AI writes

What we found

Free-text notes supplied by clients are placed near the workflow's instructions, without a reliable check that the resulting reminder still serves its intended purpose.

Why it matters

A manipulated note could cause the system to prepare a misleading message or inappropriate request. Staff would need to catch the problem before sending it.

RECOMMENDED ACTION

Keep staff approval in place. Ask the provider to limit unnecessary free text and verify that client input cannot redirect the reminder.

How we will know it is resolved

Tests using misleading client instructions cannot change the recipient, reminder purpose, approved fields, or requirement for staff approval.

Evidence and test results

E-05: An example prompt builder appends **client_notes** after the reminder instructions without structured separation or reliable output checks.

T-05: A test note asked recipients to reply with payment information. The generated draft followed that instruction. **No email was sent.** The content-integrity check failed.

Technical notes for the provider

Treat client content as untrusted data. Use constrained templates and output checks. Retest with deliberately misleading inputs.

Framework references: AI RMF MEASURE and MANAGE; NIST AI 600-1 information-integrity risks.

MEDIUM SG-06

Errors and duplicate reminders are hard to track

What we found

The workflow does not reliably recognize when two runs belong to the same client request, and some failures have no assigned person to resolve them.

Why it matters

Clients could receive duplicate reminders or no reminder at all. Employees may have to reconstruct what happened by hand.

RECOMMENDED ACTION

Assign an operations owner. Ask the provider to prevent duplicate actions and make failures visible to the right staff.

How we will know it is resolved

Repeating the same request has only one approved result. Failed requests are clearly logged, assigned to an owner, and recoverable.

Evidence and test results

E-06: Two test runs for case **WC-1091** relate to a single intake event but have different platform run IDs.

T-06: Replaying the event created two draft reminders without a duplicate warning. A simulated connector error remained unassigned. The exception-handling checks failed.

Technical notes for the provider

Use durable event IDs, processing that does not repeat an action, clear state tracking, actionable alerts, and an assigned exception queue.

Framework references: NIST CSF 2.0 IDENTIFY and RESPOND; AI RMF MANAGE.

WHO DOES WHAT

Turning findings into action

Kingfisher provides the independent assessment and the tests required for acceptance. WhiteCloud appoints its provider to make changes and keeps approval authority.

What needs fixing	Who owns the change	What must be demonstrated
Limit file access **SG-01**	IT provider	Unrelated SharePoint reads fail; approved reads still work.
Limit information sent to AI **SG-02**	Integration provider	Captured requests contain only approved fields, including retries.
Require staff approval **SG-03**	Workflow provider and operations	No client-facing email can go out without recorded approval.
Define ownership and handle exceptions **SG-04 to SG-06**	IT provider and operations	Shutdown test, misleading-input tests, duplicate prevention, and assigned error handling pass.

Recommended sequence

Now	Stop automatic emails. Continue staff-led intake and confirm which client files the system can access.
Before a limited restart	Verify the three High findings and test manipulated client inputs. Keep a person approving every outgoing message.
Before expanding use	Test shutdown and recovery, prevent duplicate work, assign owners, and review remaining risks with management.

REVIEW FRAMEWORK

How established guidance informs the review

These are the NIST questions most relevant to this workflow. They do not imply that every control has been audited or certified.

Reference	The question for WhiteCloud	Supporting evidence
AI RMF MAP 1.1	Is the business purpose clear, and do we know which client data and recipients are permitted?	Approved use and allowed information list.
CSF 2.0 PR.AA-05	Can the system access only the information and actions it needs?	Effective permissions and blocked access tests.
CSF 2.0 GV.RR-02	Is there a named person who can respond when the system fails?	Owners, emergency-stop test and escalation process.
AI RMF MEASURE and MANAGE	Were likely failures tested and reviewed before wider use?	Negative tests, misleading-input tests and management decision.
NIST AI 600-1	Could sensitive information be exposed or the AI draft be misleading?	Request captures, manipulated-input tests and draft review rules.

What this means: The findings are informed by NIST AI RMF 1.0, NIST Cybersecurity Framework 2.0, and the NIST Generative AI Profile. This review is not a certification or a guarantee that the system is secure.

RELEASE DECISION

When can automatic emails resume?

RESTART REQUIREMENT

The managing partner should approve a limited restart only after tests confirm that the serious access, data-sharing, and email-approval issues are fixed. Staff approval remains required for every outgoing message.

Evidence to keep with the decision

Record	What the evidence must show	Who confirms it
E-01, T-01	Access permissions and proof that unrelated client files are blocked.	IT owner and managing partner
E-02, T-02	Approved data fields and captured requests showing nothing extra is sent.	Operations and privacy owner
E-03, T-03	Proof that only a staff-approved message and recipient can be sent.	Operations lead
E-04, T-04	Named owner, emergency-stop test and written restart procedure.	IT service owner
E-05, T-05	Results from tests using misleading client instructions.	IT provider and operations
E-06, T-06	Duplicate-handling test results and a process for assigning failures.	Operations owner

Final responsibility: The provider implements changes. WhiteCloud management decides whether the remaining risk is acceptable.