Least privilege for AI agents in Microsoft 365.

Broad mailbox access increases the potential impact of prompt injection. Limit what an agent can read, change and send, and test those boundaries.

Toltec engineering | 23 September 2026 | Updated 28 October 2026 | 5 minute read

The short version

- Use accountable identities with named owners and permissions matched to the task.
- Choose delegated or application access for the operation; human approval is a separate control.
- Separate reading, drafting and committing changes.
- Treat retrieved content, tool responses and stored memory as possible instruction-bearing input.
- Test the boundaries, review access and verify the off switch.

Give each agent an accountable identity

Choose the identity model supported by the platform, with a named human owner and attributable actions. Microsoft Entra Agent ID provides agent identities and blueprints; a blueprint can create multiple identities, each with its own permissions. A managed identity can supply the blueprint's credentials in Azure. It is not simply a replacement for the agent identity.[1]

Existing applications may use conventional service principals. Map the actual credential and delegation path before choosing or changing it. Keep shared credentials and blueprint dependencies visible: disabling one component can affect several agents. Use managed or federated credentials where supported, and avoid borrowing a person's account or a shared administrator identity.

Choose delegated or application access deliberately

Use delegated access when an agent acts on behalf of an authenticated user and the target service supports it. Both the person's access and the application's granted scopes constrain the operation. Application access works without a signed-in user; its reach depends on the grants and resource controls, not on an automatic entitlement to the whole organisation.[2]

Use narrowly scoped application access for authorised service operations, including unattended work. Keep a person's approval of an action separate from the technical identity that performs it. A reviewer clicking approve does not, by itself, create a delegated identity context.

Scope application permissions down

Some application grants are broad: Mail.ReadWrite can cover every mailbox and Files.ReadWrite.All can cover all files available through the permission. Check the permission mode and resource scope rather than judging a grant by its name alone.[3]

For Exchange Online, RBAC for Applications can restrict the mailboxes an application serves. Check all permission paths: an overlapping tenant-wide grant can undermine a narrower RBAC assignment.[4] For SharePoint, Selected permissions require explicit resource assignments and can target supported sites, lists or files.[5]

Broad access to review	A narrower option to assess
Mail.ReadWrite across all mailboxes	Exchange application RBAC scoped to the required mailboxes, with overlapping grants reviewed.
Files.ReadWrite.All or Sites.ReadWrite.All	Selected permissions and explicit grants for the supported sites, lists or files the task needs.
Directory write permissions	Read-only access or no directory access, according to the actual task.

Separate reading from acting

An email-summary task does not require sending rights. Drafting and committing changes should have separate interfaces and permission checks. For consequential actions, show the exact destination and proposed change to the responsible reviewer. Bind approval to that action and enforce it in application code before execution.

Treat content as evidence, not authority

An email, document, image, tool response or stored memory can carry instructions aimed at the agent. OWASP's 2026 prompt-injection guidance treats the surrounding application's permissions and execution controls as essential limits on the consequences of a manipulated model.[6]

Validate tool arguments, restrict destinations and operations, and keep credentials outside the model's context. Treat memory writes and changes to tool descriptions as part of the trust boundary. Test whether hostile content can turn reading into an unauthorised write, disclosure or persistent instruction. A prompt or filter alone is insufficient evidence that the boundary holds.

Check which sign-in controls apply

Match sign-in and network controls to the selected identity type. Conditional Access for workload identities can block supported single-tenant service principals by location or risk. It requires the appropriate Workload Identities Premium licence to create or modify policies; managed identities and multitenant SaaS applications are outside that policy's coverage.[7] Verify the applicable agent-identity controls and record any remaining gaps.

Log it, review it and keep an off switch

- Retain relevant identity and action logs under an agreed access and retention policy.
- Record changes to permissions, credentials, ownership and connected tools.
- Review access on an agreed schedule and after material task or staffing changes.
- Test a platform-appropriate stop procedure: pause the runner, disable the relevant identity or access path and remove credentials or grants where appropriate. Previously issued tokens may remain usable until expiry or resource-side revocation. Check the target system to establish that the agent can no longer act.[8]

Start with the job

Write down the task, eligible users, permitted records and allowed actions before choosing permissions. Then test the denied paths as carefully as the successful one. The resulting evidence should explain what the agent can do, who authorised it and how it can be stopped.

References

1. Agent identities, service principals, and applications Microsoft Learn
2. Overview of permissions and consent Microsoft Learn
3. Microsoft Graph permissions reference Microsoft Learn
4. Role Based Access Control for Applications in Exchange Online Microsoft Learn
5. Selected permissions overview Microsoft Learn
6. OWASP Top 10 for LLM Applications 2026: Prompt Injection OWASP GenAI Security Project
7. Conditional Access for workload identities Microsoft Learn
8. Deactivate an app registration Microsoft Learn

Read the current edition: toltec.io/note-ai-agent-access.html