Using Personal AI Agents Safely and Securely: What Consumers Should Know

Advisory on AI assistants that can access your apps, websites and online accounts

Personal AI agents (such as Dots, Hermes, Muse, Grok Bot, and OpenClaw)[1] are becoming increasingly capable, helping users with everyday tasks. Unlike AI chatbots, these AI agents can take actions on your behalf. For example, they can connect to your apps and online accounts, browse websites, manage emails and make bookings. They can also store your preferences and information that you have shared.

AI agents can save you time, but they also bring new privacy and security risks. Depending on the access you give them, AI agents may read and reveal sensitive personal information, make mistakes, or take actions that you did not intend. Some AI agents can continue working in the background, even when you are not actively using them. They could be tricked by malicious actors, websites or messages into doing something harmful. People and organisations interacting with your AI agent may also not realise that they are not dealing with a person.

AI agents can be very useful, but you cannot treat them like human assistants and give them the same broad access to your information and implicit trust. Instead, grant narrowly scoped permissions, retain control of consequential actions and understand how to revoke access.

You remain responsible for how you use AI agents, including the actions these agents take on your behalf. Depending on the circumstances and applicable laws, you may also be held liable for these actions.

This advisory is intended to guide users on the best practices and precautionary steps to take when using personal AI agents, so that you can benefit from their capabilities safely and confidently.

Step 1: Limit what your AI agent can access

An AI agent may ask to connect to your email, calendar, contacts, messages, files and other online accounts and apps e.g. Gmail, WhatsApp, Facebook. Once connected, it may be able to retrieve, process and share information from these services, including personal information about other people. Only grant access for what is necessary.

- **Only connect the apps and accounts the AI agent needs to perform its tasks**. Where you can, further limit the information that the agent can access within each app and account.

- **Start small to understand how the agent works.** Use accounts, apps and/or content that are low-risk and less sensitive (e.g. secondary email accounts).

- **Avoid granting access to sensitive accounts** such as banking, government or healthcare accounts. For work accounts, check if your employer allows it.

[1] For more specific information on OpenClaw, see CSA's Advisory on Cybersecurity Risks of OpenClaw and IMDA's Case Study on the Responsible Deployment of OpenClaw.

- **Do not disclose your usernames, passwords, passcodes, OTPs or other authentication credentials to an AI agent**. This includes any documents, such as files, emails, or messages, where you may store such credentials.
    - **If the digital service uses Singpass,** opt for passkeys login as default, complete the Singpass authentication yourself, and do not approve unexpected authentication requests. **For other digital services**, opt for authentication options that require human action.
    - **If you connect an email account that receives OTPs or other authentication credentials, your AI agent may be able to access it.** For sensitive accounts, consider setting up a separate email account to receive OTPs that cannot be accessed by the AI agent.
- **Where feasible, limit the agent's permissions to specific folders only** and avoid installing it on devices where sensitive data is stored.
- **Avoid sharing credit card information with an AI agent.**
- **Keep track of all access** that has been given to the AI agent.

Step 2: Stay in control of important actions

Depending on the permissions you give it, some AI agents may send messages, make purchases, or carry out other actions on your behalf.

- **Start with limited or read-only permissions where possible.**
- **Set the agent to seek your permission before it performs actions that may have significant impact**, such as sending messages, deleting or sharing files, or posting public content. Choose one-time approval over "always-allow".
    - **Be extra careful when using AI agents for high-risk transactions** and check the details before approving any actions.
- **Use app-level settings to manage permissions.** Do not rely solely on chat instructions to limit your agent's actions, as these may not always be followed consistently. Use app settings to set permissions instead, e.g. in the "Settings" of your agent's app and the app that your agent is accessing, such as Gmail.
- **Review permissions and credentials regularly and revoke those you do not use or need.**
- **Regularly check in on agent activity.** AI agents may act inconsistently even for routine or similar tasks. Periodically review its activity history to make sure it is acting as intended.

Step 3: Find out how your information is used and stored

Information you share with an AI agent, including information from connected accounts, may be processed and stored by the agent provider and other companies it works with.

- **Find out:**
  - What data the agent collects, who it shares this with, and how long it is kept.
  - What happens when you disconnect an account or stop using the service.
  - What you may be responsible for, including unexpected actions by your AI agent.

- **Check training and ads settings**. Find out whether your information (e.g. chat history, connected accounts) is being used to train AI models and to personalise ads, and adjust these settings based on what is right for you.

- **Understand how the agent is learning about you and others around you**. Many personal AI agents have persistent memory, and they can create personal profiles of you and your contacts based on your files and interactions. Delete this information if it is inaccurate or if you do not want it to be stored or used by the agent.

Step 4: Protect yourself from scammers or other malicious actors

Scammers may target your AI agent instead of you. For example, your AI agent could be tricked by instructions hidden in what they read (e.g. web pages, emails, calendar invitations, messages, documents or images), to manipulate your agent into doing something you did not ask for, such as buying something or transferring funds, or revealing your personal data. **As some agents operate in the background, you may not immediately realise that something has gone wrong. The best protection is to limit what your agent can see or do.** In addition:

- **Pay attention when your agent is browsing on your behalf**. Ensure it is not logged in to any important accounts and stop it if it attempts to visit unintended websites or take actions you did not request.

- **Make sure only you can give it instructions**. Use settings to limit who can send it instructions and do not add your agent to group chats with people you don't trust.

- **Use strong authentication for the agents**. Use strong, unique passwords or a passkey, as anyone who gains access to your agent account may be able to access connected accounts.

- **Download AI agents only from official sources and use only trusted third-party add-ons**. Malicious add-ons could steal your information or manipulate your agents to take harmful actions. Watch out for look-alike apps or fake installers.

If your agent behaves unexpectedly or you suspect your information has been exposed:

- **Stop the agent.** Stop the current task in the app. For cloud-based agents, you may need to pause or disable it in its settings. If the agent runs on your computer or device, switch off the device it runs on.

- **Disconnect affected accounts.** Do this both in the agent's settings and the affected accounts (e.g. your calendar or email account) themselves. Change affected passwords and sign out of all sessions.

- **Look out for changes you did not make:** messages, payments, deleted items, changed settings, email forwarding rules, new devices, new payees, and changes to your recovery email or phone number.

- **Clean up the agent** by deleting unrecognised memories, add-ons, saved logins, and tasks. Revoke all credentials, e.g. passwords, API keys, and use the provider's reset or uninstall options if needed. Save activity records before deleting.

- **Report and notify suspicious behaviours.** If your agent is compromised, it could be used to send malicious messages in your name. Inform anyone who may be affected and report to the relevant authorities.

Where to get help

Suspected scams or unauthorised financial transactions

- Stop the agent and revoke its access to affected accounts.
- Money at risk: Activate the kill switch in your banking app to block online access to your bank account or lock your card to prevent further usage. Contact your bank's official hotline for assistance.
- Not sure if something is a scam? Call the 24/7 ScamShield Helpline at 1799, or check in the ScamShield app.
- Lost money, or think a crime has been committed: make a police report through the Singapore Police Force's e-Report service, or at any Neighbourhood Police Centre. In an emergency, call 999.

Compromised accounts or cybersecurity incidents

- Singpass: Contact the Singpass Helpdesk at 6335 3533. You may also follow the steps at https://go.gov.sg/protect-singpass-account to secure your account.
- Hacked device or accounts: report the cyber incident to SingCERT through CSA's Cyber Aid form. If you suspect a crime, make a Police report.

Step 5: Think about the impact on others

When an AI agent acts for you, the people or services it deals with may not know they are dealing with an AI agent.

- **Respect other people's information.** Your emails, messages and files contain information about family, friends, and colleagues. Think about whether they would be comfortable with an AI agent reading or sharing it. Avoid sharing sensitive information about others unless you need to. Review what the agent keeps about other people, and delete what is not needed.

- **Indicate that the email or message was sent by an AI agent.**

- **Do not ask the agent to circumvent or breach technical controls or access restrictions** (such as paywalls or login requirements) on other websites, systems, or apps.

Annex A: Where to find controls in your AI agent

Menus and terms differ between agents. This table shows common terms you can look for to manage the agent's settings.

Step	What to look for	Common names	What to do
1	Connected apps and accounts	Connectors, Apps, Integrations, Linked accounts	Keep track of what is connected. Disconnect what you do not need.
2	Permissions for each app	Permissions, Manage permissions	Provide read-only or limited permissions by default, and grant ability to perform actions only based on what is required to perform assigned tasks.
	Approvals	Allowed websites, Always allow, Take over	Require approval for important actions. Remove approvals you no longer need.
	Activity monitoring	Activity log, History, Upcoming, Scheduled, Automations	Monitor for suspicious activities performed by the agent
3	Terms of use, privacy policy	Terms of use, Data controls, Privacy policy, Model training	Read the terms of use and privacy policy/statement to understand what data is collected, how it is used, and whether you can limit this
	Memory	Memory, Saved info, Personalisation	Review what the agent has saved about you and how it uses it
4	Stop and pause	Stop, Pause, Take over	Find these before you start a task.
	Download and reset	Download your data, Reset, Delete account	Download your records first. A reset usually cannot be undone, and may not undo changes in your connected accounts.
	Account security	Account, Security, Password and security	Turn on 2FA, or use a passkey.
	Add-ons	Skills, Plugins, Extensions, Custom connectors	Only use trusted third-party add-ons

Annex B: Illustration of 5 steps with Meta's Muse

This Annex uses Muse, Meta's AI agent, to show how the 5 steps can be applied. The menus and settings here are accurate as of 5 Nov 2026 for "Muse from Meta" on iOS (v14.1) and Android (v14.0) in Singapore. They may change, so refer to Meta's Muse Help Centre for the latest. Other AI agents may have similar controls.

These steps are illustrative and non-exhaustive. It is not an endorsement of Muse or Meta, and is not an assessment of Muse's security or privacy.

1. Limit what Muse can access	**Connectors:** Muse connects to apps such as Gmail and Google Calendar. Meta calls these "Connectors". You can enable or disable them in **Settings** (⚙ in the sidebar) **> Connectors**. Before you connect an app, read what information it will share with Muse and what Muse can do in it. **Using your Meta account:** If you use Muse with an account in the same Meta Accounts Centre as your Facebook, Instagram or Threads accounts, those accounts are automatically connected to Muse. Review this in Accounts Centre. If you prefer Muse not to use your other Meta accounts, consider using Muse with a Meta account that is not connected to your other Meta accounts. **Custom connectors:** Be careful about asking Muse to create a "custom connector" for a service that is not in its list. Meta says that it does not review custom connectors.
2. Stay in control of important actions	**Permission:** When you connect an app, depending on the app's features, you can choose whether Muse can only **read** (for example, read your emails) or **write and delete** (for example, send emails). Choose read-only if you only need Muse to find information. **Web access: go to Settings > Permissions > Web access defaults.** • If "Always ask" is selected: Muse asks before visiting any website. • If "Ask for some actions" is selected: Muse decides when to ask, for example when your information may be shared or a website that is unfamiliar. • If Muse can reach important accounts, choose "Always ask". **To review or remove permissions:** Go to **Settings > Permissions > Connectors**, and then select the Connector/check each Connector. **Review approvals carefully:** When Muse requests permission to take an action, users may be given options such as "Allow once", "Allow for this task", "Allow for this site" or "Always allow". Depending on the option selected, Muse may be able to carry out similar actions for the same task, website or Connector without seeking approval each time. • Users should check the scope of each permission before approving it.

	• Tap See task details on any request to see more before you decide. If you are unsure, choose “Deny”. **Monitoring:** Muse has features to facilitate users' monitoring of its actions and provide transparency into what Muse is doing: • **What agent is currently doing:** The assistant icon at the top of the screen shows what Muse is doing (for example "Is working", "Making something", "Is updating memory"). • **What permissions are being requested:** Tap **See task details** on any permission request to see additional task details before deciding what permission to grant. • **What agent has done:** Tap the assistant icon to open the **Activity log** (⁝☰), a record of what Muse has done and the permissions you gave. If what Muse says in the chat does not match the Activity log, rely on the Activity log. • **What agent will do:** Tap the assistant icon, then **Upcoming** (🕘) to see scheduled tasks and reminders. To cancel one, tap on the task > **Delete** or ask Muse in the chat to delete it, then check **Upcoming** again to confirm it is gone.
3. Find out how your information is used and stored	**Use of user data for training:** By default, Meta may use users' interactions with Muse to train and improve its AI models. Users can turn off this setting at any time: 1. Go to **Settings**, then tap **Data controls**. 2. Toggle off **Help improve our AI models**. 3. Select **Turn off** to confirm. If you allow your interactions to be used, Meta still removes certain key categories of personally identifiable information like names, email addresses, phone numbers, and government ID numbers. They also disassociate the interactions from your Meta account. **Review what Muse knows and remembers about you:** Muse learns about the user over time and stores information (e.g. the conversation history, artefacts and other files that it uses to provide the user's experience). Users should periodically review their stored information and remove anything that is incorrect, unfamiliar or no longer needed. Users can use the controls in Muse to manage this information (e.g. to access, edit or delete it). For example, to manage files created when interacting with Muse: 1. Navigate to the **Library** tab in the sidebar > **More** > **System files**. 2. Select the file that you want to review. In particular, the /memory folder and MEMORY.md file can hold facts and preferences stored about you. For more information, see How to manage your Muse data.

4. Protect yourself from scammers and malicious actors	**Account security:** Protect your Meta account with two-factor authentication: **Accounts Centre > Password and security > Two-factor authentication**. **Warning: Resetting Muse permanently deletes your chat history, files, and active tasks. It cannot be undone. Try the first four steps first.** 1. **Stop the task.** If Muse is using its browser, select Stop the task to end it, or Take control of the browser to pause it. You can also ask Muse to stop. 2. **Disconnect affected Connectors in Settings > Connectors.** For Facebook, Instagram or Threads, use Meta Accounts Centre. Information Muse has already used may stay in its memory. 3. **Check what happened.** Look at the Activity log and "Upcoming" and remove standing permissions you do not recognise (see Step 2). 4. **Save a copy of your records:** Settings > Data controls > Download your agent data. **If issues persist**, users can consider resetting Muse using the controls provided in the app: go to **Settings** > **Data controls** > **Reset Muse**, follow the on-screen instructions, then tap **Reset**. Resetting does not undo actions already taken or information already shared. Before doing so, check what information and tasks will be deleted. Users should also review connected accounts and revoke affected permissions where necessary.
5. Think about the impact on others	**External communications:** In external communications, ask Muse to indicate that the communication was sent by an AI agent. For example, when Muse writes emails or messages for you, ask it to say they were prepared or sent by an AI agent, and check what is actually sent. **Memory:** Muse's memory may include information about people you deal with. Review it regularly, and delete what is not needed.