Using Generative AI Safely: Singaporeâ??s New Checklist

Description

Treat generative AI as an untrusted assistant: minimise what you enter, verify important outputs, use official tools and updates, and give agents the smallest possible permission. Never let fluent text or a familiar interface substitute for checking a source, recipient or transaction.

A Singapore individual using a public generative-AI chatbot, assistant or agent faces a narrower question than the headline suggests: decide what data and authority to give the tool and verify output before it affects money, accounts, work or safety. The table separates the branches that change the answer before the practical checks and worked example.

Choose the branch before acting

Situation	Practical next step
Prompt contains confidential or personal data	Remove, mask or use an organisation-approved tool and policy
Output affects health, law, money or safety	Open controlling sources and obtain qualified help where needed
AI asks to install an extension or sign in	Verify publisher, permissions and destination independently
Agent can send, buy, delete or publish	Use narrow scope, previews, limits and human confirmation before action

Decide whether the prompt is safe to send

CSA and IMDA issued their joint advisory on 28 July 2026 to guide individuals using generative-AI tools safely and securely. The central practical idea is to manage both information risk and action risk; a useful answer can still contain errors or expose data. CSA and IMDA joint advisory of 28 July 2026.

Do not enter passwords, one-time codes, identity numbers, medical records, client files, unpublished work or another personâ??s personal data into a public service without an approved basis. â??The model needs contextâ?• is not permission to disclose information you do not control. CSA and IMDA joint advisory of 28 July 2026.

Verify output at the claim level

Before sending, reduce the prompt to the minimum facts needed. Replace names with roles, amounts with ranges and documents with small relevant excerpts where policy allows. For workplace use, follow the employerâ??s approved tools, retention settings and data-classification rules. CSA and IMDA joint advisory of 28 July 2026.

Generative systems can produce plausible but false citations, calculations and instructions. For consequential claims, open the cited source and verify the exact number, date, eligibility and caveat. A link to a relevant page is insufficient if the page does not support the sentence. CSA and IMDA joint advisory of 28 July 2026.

Control accounts, files and extensions

Use official app stores and vendor domains, update software promptly and enable 2FA with strong unique passphrases. Be cautious with browser extensions, plug-ins and â??AI wrappersâ?• that request access to all pages, mail, drive files or clipboard history. CSA Stop and Check campaign.

Scammers can use AI-generated voices, images and messages to create urgency or familiarity. Pause and verify requests through a known channel, especially when money, credentials or secrecy are involved. Do not use contact details supplied inside the suspicious message itself. CSA Stop and Check campaign.

Give agents bounded authority

Agentic tools can act rather than only answer. Limit folders, accounts, spending, recipients and duration; require a preview before sending or deleting; and test with reversible low-risk tasks. Revoke access after the task and review logs for unexpected activity. CSA Stop and Check campaign.

Keep human accountability. Record the source, tool, version or date and checks used when AI materially contributes to a work product. Do not represent generated observation, interviews, prices or testing as first-hand evidence. The human publisher remains responsible for the result. CSA Stop and Check campaign.

Put the numbers or sequence to work

A travel assistant asks for a passport scan and permission to buy any flight under S$2,000. The user instead provides non-identifying travel constraints, searches official entry rules separately and limits the agent to draft itineraries. Payment and booking remain behind a final preview and manual confirmation.

The example is a planning model, not a quoted price, official calculator result, medical instruction or promised outcome. Replace its assumptions with the issued notice, live service, signed contract, current timetable or professional advice that controls the real decision.

Before you commit

1. Classify data before prompting
2. Minimise and mask context
3. Open sources for important claims
4. Verify app and extension publishers
5. Use 2FA and updates
6. Bound agent permissions and spending
7. Review output and logs before relying

A useful working note combines **a prompt-data minimisation ladder from full records to masked facts** with **an agent-authority matrix covering folders, money, recipients, duration and confirmation**. Enter only details that can be tied to a current document or live record.

Missteps that change the answer

- Uploading an entire confidential document
- Trusting invented citations
- Installing a broad-permission extension
- Approving unlimited agent action
- Passing generated material off as observed fact

If one of these conditions appears, pause before payment, submission, travel or implementation and reconcile it through the relevant official service. Save the issued result or acknowledgement; a search snippet or forwarded screenshot cannot establish a current entitlement.

The decision to carry forward

Use the current official record to resolve decide what data and authority to give the tool and verify output before it affects money, accounts, work or safety. Save the dated result and revisit it when the underlying rule, timetable, account or personal facts change.

Related next steps

Once this decision is settled, you may need to check a deviceâ??s cybersecurity label. The next adjacent check is to verify a suspicious digital message.

Common questions

Can I enter personal data into a public chatbot?

Only when there is a proper, approved basis; minimise data and follow applicable policy. CSA and IMDA joint advisory of 28 July 2026.

How should I verify AI citations?

Open the source and confirm it contains the specific claim and caveat. CSA Stop and Check campaign.

Are AI agents riskier than chatbots?

They can be because they may act; use narrow permissions and human confirmation. CSA and IMDA joint advisory of 28 July 2026.

Rules, service details and schedules can change. Reopen the linked official page before acting when the date, eligibility, payment destination, safety instruction or live availability is decisive.

Date Created
29/07/2026
Author
vanessakoh