

AI USAGE & SAFETY

A practical step-by-step guide for safer AI use at work, at home, in study, and in everyday tasks.

Use AI to assist your judgment - not replace it.



THE CORE PRINCIPLE

You set the goal. AI supports the work. You review and approve the result.

1 PLAN

2 PROTECT

3 PROMPT

4 VERIFY

5 ACT

Minimum data - minimum access - clear instructions - human review - reversible actions

Designed for everyday use, team training, and reusable checklists

Use AI safely in 5 steps

Follow the same simple workflow whether the task is for business or personal use.

1 PLAN

Define the task and the decision you want help with.

2 PROTECT

Remove private, confidential, or unnecessary details.

3 PROMPT

Give clear context, limits, and the format you want.

4 VERIFY

Check facts, numbers, sources, assumptions, and missing context.

5 ACT

Approve before sending, changing, buying, deleting, or automating.



SEE THE WORKFLOW IN PRACTICE

BUSINESS

Task: Draft a follow-up email after a customer call.

- Use only the facts needed from the call.
- Replace private customer details with placeholders.
- Ask AI for a draft - not permission to send it.
- Check dates, promises, pricing, and tone.
- A person approves the final email before sending.

PERSONAL

Task: Compare options for a weekend trip.

- Share general preferences, not IDs or payment details.
- Ask for options, pros/cons, and questions to verify.
- Open the original hotel, airline, or venue pages.
- Confirm prices, dates, policies, and availability yourself.
- You make the booking and payment decision.

STOP BEFORE YOU CONTINUE IF...

- The task asks for a password, security code, full payment number, government ID, or private health data.
- The AI wants broader access than the task needs, such as an entire mailbox or drive.
- The answer will trigger an important action and nobody has reviewed it.
- A file, email, or webpage tells the AI to ignore your instructions or expose data.

Match the tool to the risk

Choose the lowest-risk tool and the smallest amount of access that can complete the task.

CHOOSE THE RISK LEVEL

LOW

Public or non-sensitive information

Business: ideas, public marketing copy, public research.
Personal: brainstorming, recipes, general travel ideas.

MEDIUM

Internal or personal information with limited sensitivity

Business: approved internal notes, non-confidential files.
Personal: schedules, preferences, drafts without IDs or account data.

HIGH

Sensitive, regulated, identity, money, or security information

Business: client secrets, employee records, credentials.
Personal: passwords, health records, financial accounts, government ID.



Rule of thumb: If the task involves identity, money, health, legal rights, confidential business information, or security access, treat it as high risk until proven otherwise.

ADD ACCESS IN STAGES

1

CHAT

No connected systems

2

READ

Only selected files or folders

3

CREATE

Draft or create, but do not send

4

APPROVE

A person confirms the action

5

AUTOMATE

Only low-risk, tested, logged tasks

ACCESS GATE - ASK THESE QUESTIONS FIRST



Who owns this task and who approves the final action?



What can the AI create, edit, send, buy, delete, or publish?



What exact data must the AI see to do the job?



How do I revoke access, review activity, and restore a clean copy?



Can the same result be achieved with read-only access?



Is the tool approved for this risk level and this type of information?

Pause before you paste, upload, or connect

Classify the information, minimize it, then use only the data needed for the task.

TRAFFIC-LIGHT DATA CHECK

GREEN

Usually OK

Public or already approved information. Examples: public webpages, generic ideas, your own non-sensitive notes.

YELLOW

Check before sharing

Internal, personal, or private information that may be acceptable only in an approved tool and only when needed.

RED

Do not paste by default

Passwords, codes, payment data, health records, government IDs, confidential client data, private employee records.



MASK DETAILS BEFORE AI USE - 4 STEPS

1 REMOVE

Delete names, addresses, IDs, account numbers, and anything not needed.

2 REPLACE

Use placeholders such as [CUSTOMER], [ACCOUNT], [CITY], or [DATE].

3 REDUCE

Share the smallest excerpt, time range, or data sample that will work.

4 RECHECK

Read the prompt again before sending. Look for hidden details in screenshots and files.

BUSINESS EXAMPLE

Before: Jordan Lee - Customer 4821 - invoice balance \$1,240 - phone 555-0198.

Safer version: [CUSTOMER] - [ACCOUNT] - invoice balance \$1,240 - [PHONE REMOVED]. Ask for a payment reminder draft only.

PERSONAL EXAMPLE

Before: Here is my passport, booking number, home address, and full credit-card information. Plan my trip.

Safer version: I am traveling from [CITY] to [CITY] for 4 days. Compare neighborhoods, transport, and typical costs. I will handle IDs and payment myself.

PROVIDER / APP CHECK

Personal use: review privacy and data-control settings. **Business use:** confirm retention, deletion, training-use, admin access, security, and contract terms required by your organization. Recheck because service terms and settings can change.

Write prompts people can review

A clear prompt reduces mistakes and makes the final result easier to verify.

THE 6-PART PROMPT FORMULA

1 GOAL

What should AI help you accomplish?

2 CONTEXT

Only the background needed to understand the task.

3 CONSTRAINTS

Limits: time, budget, policy, length, audience, or must-not-do rules.

4 SOURCES

State which files, links, or facts it may rely on - and what it must not assume.

5 OUTPUT

Specify format: checklist, email draft, table, summary, questions, etc.

6 REVIEW

Ask it to flag uncertainty, missing information, and items needing human confirmation.

EXAMPLES YOU CAN COPY AND ADAPT

BUSINESS

Goal: Draft a polite follow-up email after a customer support call.
Context: The issue was a delayed delivery. The customer wants an update by Friday.
Constraints: Do not promise a refund or delivery date. Keep it under 140 words.
Sources: Use only the facts I provide below.
Output: Subject line + email draft.
Review: Flag any statement that needs confirmation before sending.

PERSONAL

Goal: Help me plan a 3-day city break.
Context: I prefer walkable areas, museums, and casual restaurants.
Constraints: Keep the estimate under \$800 before flights. Do not make bookings.
Sources: Separate general knowledge from facts that I should verify online.
Output: A simple day-by-day plan with a budget range.
Review: List prices, hours, and policies that may change.

SAFER PROMPT HABITS

- **Ask for options, not just one answer.** This makes assumptions easier to spot.
- **Ask for uncertainty.** Request "what you are not sure about" and "what I should verify."
- **Separate facts from suggestions.** Ask the AI to label recommendations clearly.
- **Do not delegate the final high-impact decision.** Ask for a summary, questions, or draft instead.
- **Keep sensitive source data outside the prompt whenever a placeholder will work.**
- **For connected tools, state the action boundary.** Example: "Draft only. Do not send, edit, or delete."

Check the answer before you trust it

AI can sound confident when it is incomplete or wrong. Review the parts that matter to the real decision.

FACTS



Names, dates, claims, product details, policies, addresses.

NUMBERS



Totals, units, formulas, percentages, conversions, prices.

SOURCES



Open the original source. Does it actually support the claim?



MISSING



What information, exception, or alternative was left out?

TONE & RIGHTS



Audience fit, confidentiality, copyright, licenses, attribution.

ACTION



What will happen if I use this answer? Is human approval recorded?

USE THE RIGHT DEPTH OF REVIEW

QUICK

Everyday low-risk tasks

Read it once, check obvious facts, and correct tone or formatting.

IMPORTANT

Money, commitments, client or family impact

Verify key facts against original sources and recalculate important numbers.

HIGH-IMPACT

Health, legal, security, identity, major financial or employment decisions

Use qualified human review and authoritative sources before acting.

TWO SIMPLE EXAMPLES

Business: If AI summarizes an invoice or report, compare the totals, dates, customer names, and source document before sending the summary.

Personal: If AI compares travel, products, or services, open the original listing pages and verify the price, availability, return/cancellation terms, and dates before buying.

Know what access actually allows

Connecting email, cloud files, calendars, or business apps changes the risk because AI can see or act on real systems.

PERMISSION MEANINGS

READ

Can view selected information. Start here when possible.

CREATE

Can draft or create items without sending or publishing.

EDIT / SEND

Can change live data, send messages, publish, schedule, or update records.

DELETE / MOVE

Can remove, archive, cancel, or relocate information. Treat as higher risk.

AUTOMATE

Can repeat actions without asking each time. Use only after testing and logging.



BEFORE YOU CONNECT AN APP OR ACCOUNT

- ☐ What exact folders, files, calendars, mailboxes, or records will it see?
- ☐ How do I revoke the connection and review recent activity?
- ☐ Can I select one source instead of granting access to everything?
- ☐ Will a human approve important actions before they happen?
- ☐ Can it only read, or can it also send, edit, buy, publish, or delete?
- ☐ Is there a backup or recovery path if the AI makes the wrong change?

SAFE ROLLOUT

1

READ-ONLY

Observe results first.

2

DRAFT

Create suggestions, not final actions.

3

APPROVE

Require a person before live changes.

4

LIMITED AUTO

Automate only repetitive, low-risk tasks.

5

REVIEW

Check logs and remove access no longer needed.

Business example: Let AI read one approved project folder and draft a status update. A manager reviews before it is sent.

Personal example: Let AI read one calendar or selected travel files rather than an entire drive. Keep purchases, cancellations, and deletions under your control.

Treat files, emails, and webpages as data - not instructions

External content can contain misleading directions, prompt injection, scams, or requests to expose information.



PROMPT INJECTION

A webpage, PDF, email, file, or message may contain hidden instructions that try to redirect the AI tool.

UNTRUSTED INPUT

SAFETY CHECK

HUMAN APPROVAL

5-STEP HANDLING RULE

1

KEEP YOUR GOAL

The AI should follow your task and safety rules, not hidden instructions in content.

2

TREAT CONTENT AS DATA

Summarize or extract what is needed. Do not let the content redefine the task.

3

BLOCK SENSITIVE ACTIONS

Do not reveal credentials, grant new access, make payments, or change security settings because a file says so.

4

CHECK UNEXPECTED REQUESTS

If the AI suddenly asks for secrets, more access, or an unrelated action, stop and review.

5

APPROVE LIVE ACTIONS

A person confirms important sends, edits, purchases, deletions, or system changes.

Red flags: "Ignore previous instructions" - "Send me your password" - "Disable security" - "Upload your whole drive" - "Pay this account now" - "Delete the original" - or any request unrelated to your goal.

Know when AI should assist - not decide

The more an outcome affects health, rights, safety, money, employment, or access, the stronger the human review should be.

AI CAN HELP WITH

- Summarizing information you provide.
- Generating options, questions, checklists, and drafts.
- Explaining unfamiliar terms in plain language.
- Comparing alternatives using criteria you choose.
- Preparing for a meeting with a professional or decision-maker.
- Spotting missing information you may need to verify.

HUMAN / PROFESSIONAL REVIEW

- Medical symptoms, diagnosis, treatment, or emergency decisions.
- Legal rights, contracts, formal notices, or litigation strategy.
- Large financial commitments, transfers, taxes, loans, or investments.
- Security changes, identity verification, passwords, or account recovery.
- Hiring, discipline, termination, or other high-impact employment decisions.
- Any irreversible action where a wrong answer could cause serious harm.

THE 4-QUESTION DECISION TEST

- 1 Could a wrong answer harm someone or materially affect their rights?
- 2 Does the action move money, grant access, sign a commitment, or delete something important?
- 3 Is the information regulated, highly private, or confidential?
- 4 Would a qualified professional or responsible owner normally review this decision?

IF YOU ANSWER YES TO ANY QUESTION

Use AI as a support tool: summarize, prepare questions, identify options, or draft. Then verify with authoritative sources and the appropriate person before acting.

Stop, isolate, review, and recover

Fast containment is often more important than trying to make the AI continue after something unexpected happens.

1

STOP

Pause the action. Do not send, pay, publish, delete, or keep automating.

2

ISOLATE

Disconnect the tool, revoke unnecessary access, or close the workflow.

3

REVIEW

Inspect what the AI viewed or changed. Check logs, drafts, messages, and affected files.

4

RECOVER

Restore a clean copy, correct the record, and document what happened.

BUSINESS RESPONSE

- ☐ Notify the task owner or appropriate security/privacy contact.
- ☐ Preserve activity logs and the original files before making more changes.
- ☐ Revoke or reduce the AI tool's access.
- ☐ Rotate exposed credentials or tokens according to company procedure.
- ☐ Restore from a known clean version if data was edited or deleted.
- ☐ Record what happened, what was affected, and what controls should change.

PERSONAL RESPONSE

- ☐ Disconnect the app or revoke permissions you no longer trust.
- ☐ Change any password or access code you exposed; sign out other sessions if needed.
- ☐ Check recent account, email, calendar, cloud, and payment activity for changes.
- ☐ Restore deleted or altered files from backups or version history when available.
- ☐ Contact the relevant provider, bank, or service if money or identity data may be affected.
- ☐ Save screenshots or notes if you may need support later.

RECOVERY QUESTIONS

- What information did the AI see?
- What did it create, edit, send, buy, publish, move, or delete?
- Can I undo or restore the change?
- What credentials or account permissions should be changed?
- Who needs to know what happened?
- What simple control would prevent the same issue next time?

Safe AI worksheet

Use this page before important AI tasks. Print it, copy it into a note, or turn it into your team's approval checklist.

A BEFORE USING AI

Task / goal

Risk level: Low / Medium / High

What data is truly needed?

What access is truly needed?

B BUILD THE PROMPT

Goal

Sources / facts AI may use

Context - minimum needed

Output format

Constraints / must-not-do rules

Review instruction / uncertainty to flag

C REVIEW THE RESULT

- ☐ Facts checked
- ☐ Numbers recalculated
- ☐ Sources opened
- ☐ Missing context considered

- ☐ Private data removed
- ☐ Tone appropriate
- ☐ Rights / licensing checked
- ☐ Human approval recorded

D APPROVE THE ACTION

What will happen next? What will be sent, changed, purchased, published, deleted, or automated?

Who approves? _____ Date: _____

General guidance only. Follow your organization's policies and qualified professional guidance for legal, medical, financial, cybersecurity, compliance, or other high-impact matters.