

Before you let an AI agent act for you

Tick each box before an agent sends, buys, books or deletes anything on your behalf.

☐ **1. Approvals are on**

The agent must ask before it sends, buys, deletes or submits anything.

☐ **2. No passwords, PINs or one-time codes in the chat**

Sign in yourself, or use the app's built-in sign-in or password feature.

☐ **3. Spending is capped**

Use a low-limit or virtual card, and read the cart before you confirm.

☐ **4. It starts read-only**

"Find and list" before "book and pay". Upgrade access only once it's earned trust.

☐ **5. You're watching for prompt injection**

Web pages and emails can hide instructions meant for the AI. Odd behaviour means stop the task.

☐ **6. It has the smallest access possible**

One inbox, one folder or one calendar, not your whole account.

☐ **7. You check its work**

Click 2 of its links, and ask: "What are you least sure about?"

☐ **8. Sensitive data stays out**

Medical records, ID numbers and bank details don't go into the task.

The 4-part first prompt

Goal + limits + output + stop rule. Example:

Find 3 weekend trips from Chicago under \$400 a person in November. Show a table with prices and links. Don't book, buy, sign up or send anything. Stop and show me first.