

The 36 stratagems for AI agents

One line per stratagem: how the pattern shows up in agent output and what holds against it. Meant to lie beside the screen during a review. Nothing here goes beyond what the article documents.



01 Deceive the emperor

A long run of green results lowers the guard. Counter: a gate that does not tire.



02 Besiege Wei

Review keeps failing on the code. Counter: send the correction back to the specification.



03 Kill with a borrowed knife

Agent reviews agent with no human on record. Counter: one recorded human review.



04 Await the exhausted enemy

The check runs before the agent moves. Counter: check only after the first draft.



05 Loot a burning house

Unchecked code slips through under deadline. Counter: a server-side gate with no deadline.



06 Make noise east

Eight hundred lines of report, one real defect. Counter: a condensed finding instead of the raw report.



07 Create something from nothing

Proof lines that are not in the files. Counter: match the evidence against the files.



08 March to Chencang

Every visible test green, behaviour unclear. Counter: held-out tests.



09 Watch the fire

A single checker decides alone. Counter: two models, human only where they disagree.



10 Dagger behind a smile

Agreeable tone over thin substance. Counter: score the evidence, not the agreement.



11 The plum tree withers

One large change in a single go. Counter: small slices, each verified on its own.



12 Lead away the sheep

Extras nobody ordered, added in passing. Counter: measure shapes, not just defect counts.



13 Beat the grass

The check has never once fired. Counter: plant a known defect.



14 Borrow a corpse

Invented names instead of the common ones. Counter: use existing conventions and tool names.



15 Lure the tiger

Tests run only on the developer's machine. Counter: a freshly built machine.



16 First let it go

The rule is known before drafting. Counter: let it draft freely first.



17 Toss out a brick

A huge specification and not one question. Counter: a short opening; the questions reveal the gaps.



18 Capture the leader

Twenty single rules with nothing holding them. Counter: one load-bearing rule, the project constitution.

The 36 stratagems for AI agents, continued



19 Withdraw the firewood

The threshold is named up front. Counter: reveal the threshold only afterwards.



20 Fish in troubled waters

A raw diff instead of a finding. Counter: condensed output.



21 The cicada sheds its shell

The suite shines while behaviour diverges. Counter: the held-out tests decide.



22 Shut the door

Acceptance hangs on a number. Counter: a condition with a real truth value.



23 Ally with the distant

The checker comes from the same model family. Counter: a different family spreads the blind spots.



24 Borrow a road

The carried context grows step by step. Counter: curate the context.



25 Steal the beams

It runs but nobody can build on it. Counter: measure internal quality.



26 Scold the mulberry

The rule holds on paper only. Counter: enforce it visibly once on one example.



27 Feign stupidity

A self-report instead of a result. Counter: check the result, never the report.



28 Remove the ladder

The specification lives in a vendor's tool. Counter: a plain file in your own repository.



29 Let flowers bloom

Coverage looks complete, evidence does not hold. Counter: spot-check the evidence.



30 Turn guest into host

The agent shapes the conventions. Counter: the constitution stays with the human.



31 The beautiful woman

The verdict follows whichever answer came first. Counter: swap the order and test the instrument itself.



32 The empty city

A confident tone without substance. Counter: judge the evidence, not the tone.



33 The double agent

Foreign content read as instruction. Counter: treat reading matter as untrusted.



34 Self-injury

The checker never reports anything. Counter: plant a canary.



35 The linked chain

A single gate carries everything. Counter: a chain of mutually independent checks.



36 Retreat is best

A run gets patched without end. Counter: discard it and start again.



An agent has no intent. It produces the output that scores best for the least effort.