

NAME

cmagent - tool-using agent on a local language model

SYNOPSIS

cmagent [*OPTION*]... *MODEL* [*TASK*]

DESCRIPTION

cmagent runs a conversation in which the model may call tools: read, list and search files, query git and a language server, and, if enabled, edit files and run shell commands in a sandbox. Each tool result is appended to the conversation as a tool turn, and the model continues until it answers in prose.

MODEL is any model **cmgenerate**(1) loads. A tool call is constrained while it is sampled to valid JSON naming a known tool. *TASK* is the first task; further tasks are read at the prompt.

OPTIONS**Generation**

-c, --context *N*

Context length to allocate. Default: 32768.

--window *N*, **--sinks** *N*

Sliding-window attention; see **cmgenerate**(1).

-n, --steps *N*

Maximum tokens per answer. Default: until the end of the turn.

-t, -k, -p, -s

Temperature, top-k, top-p, seed; as in **cmgenerate**(1).

--system *TEXT*

System prompt, placed before the tool declarations.

--speculate *N*, **--no-mtp**, **--mtp-from** *PATH*

Speculative decoding; as in **cmgenerate**(1).

--effort *LEVEL*

Reasoning effort; as in **cmgenerate**(1).

Tools

--write

Enable the file writing tools: **write_file**, **edit_file**, **patch_file**, **move_file**, **delete_file**, **undo_edit**,

changes, run_check.

--run

Enable **run_command** (shell commands in the sandbox) and **check_command** (background commands).

--yes

Do not ask for approval before writing or running. Stored refusals still apply.

--no-check

Do not run the project check after an edit.

--hops *N*

Maximum tool calls per task. Default: unlimited. The third identical call that acts is refused, as is a repeated reading call whose result has not changed.

--free-calls

Do not constrain tool calls to valid JSON and known tool names.

--command-timeout *S*

Maximum seconds a command may request. Default: 3600. A command that sets no timeout and does not run in the background gets 300.

--task-timeout *S*

Maximum seconds per task. Default: 14400.

--max-output *BYTES*

Tool output kept in the context per result. Default: 16384, less if the context is small. The full output is written to */tmp/output-N.txt* in the sandbox.

Sandbox

--root *DIR*

File tools do not read or write outside *DIR*. Default: the working directory.

--allow-read *PATH*

Extra read-only path for the shell and the reading tools. Repeatable.

--allow-write *PATH*

Extra writable path for the shell. Repeatable.

--allow-network

Allow network access in shell commands.

--no-gpu-in-shell

Hide the GPU device nodes from shell commands.

--no-home-in-shell

Hide the toolchains under *\$HOME* from shell commands.

Session

--once

Run *TASK* and exit.

--require-tool

With **--once**, exit with an error unless a tool call succeeded.

--resume

Restore the saved conversation and tool results, without replaying actions.

--session-id ID

Keep a separate transcript within this project.

--no-prefix-cache

Recompute the system turn instead of loading its KV cache from *~/.cache/cmagent*.

--log PATH

Append tool calls and results to *PATH* as JSON lines.

-q, --quiet

No load report; warnings are still printed.

--vi vi key bindings at the prompt.

--tui

Full-screen interface: transcript, status line and input, with approvals answered in place. Runs the agent as a child process over **--bridge**.

--bridge

JSON Lines front-end protocol on standard input and standard error. **{"type":"cancel"}** stops the running task as Ctrl-C does.

Placement

--gpu, **--gpu-backend**, **--gpu-device**, **--gpu-as**, **--cpu-as**, **--kv-format**, **--lora**, **--lora-as**: as in **cmgenerate**(1).

TOOLS

Always offered: **list_directory** and **find_files** (discovery), **read_file** (numbered lines), **search_files** (ripgrep if installed; **regex** and **context** arguments), **outline** (symbols with line ranges, from the language server or declaration-like lines), **git** (status, diff, log, show, blame; read-only, where there is a *.git*), **remember** and **recall** (the model's notes), and **delegate** (a sub-task run in the same context window and rewound when it answers; at most 40 calls, no nested delegation).

A language server, the command in *.cmagent/lsp* or else **rust-analyzer** in a Cargo project, is started in the sandbox and offered as **find_definition**, **find_references**, **describe_symbol** and **find_symbol**.

APPROVALS

Writing and running show the edit preview or command and ask, unless **--yes** is given. An answer can be kept for the exact command or path, for every command of that program or every path of that tool, or for a pattern (**cargo test ***, **src/*.rs**), for the session or for the project (*.cmagent/allowed.json*). A command is covered only if every **&&**, **|** and **;** segment is; one containing a **\$(...)** substitution only by an exact answer.

A refusal can be stored the same way (**/never rm**, **/never git push ***, or **!** before the answer letter). A refusal takes precedence over every approval, over **--yes** and over an interactive yes; one matching segment is enough.

SANDBOX

Shell commands require Linux and bubblewrap. The project is writable, *.git* included; system directories are read-only; toolchain caches under *\$HOME* (rustup, cargo registry, npm, pip) are read-only and credential files are omitted; the GPU is visible; network access is disabled. */tmp* is *.cmagent/scratch* and persists between commands; **read_file** can read it. Agent state under *.cmagent* is not writable. All commands of a session share one network namespace, so a server started in the background is reachable at 127.0.0.1 by later commands but not from the host.

A command that needs the GPU's memory sets **gpu: true**; on Vulkan the agent's model is then moved off the device for the duration and restored before the model reads the result.

Ctrl-C terminates running commands and their background children.

PROJECT CHECK

After every edit the project check runs in the sandbox and its first diagnostics follow the edit's result: *.cmagent/check* if present, otherwise **cargo check --all-targets -q --message-format=short** for a *Cargo.toml* or **go vet ./...** for a *go.mod*.

HOOKS

Executables in *.cmagent/hooks* run at three points:

after-edit

Instead of the project check.

before-run

Receives each command line on standard input. A non-zero exit refuses the command; its output is the reason given to the model.

after-task

After a task that performed an action is answered. A failure is reported to the model once.

PROJECT TOOLS

An executable in *.cmagent/tools* is offered as a tool under its file name. Invoked with **--describe**, it prints a JSON object

```
{"about": ..., "acts": bool,
```

```
"takes": [{"name": ..., "about": ..., "kind": ..., "required": bool}]}
```

with **kind** one of string, integer, boolean. A call runs it in the sandbox with the arguments as one JSON object on standard input; its standard output is the result and a non-zero exit an error. A tool with **acts** true is offered only with **--write** or **--run** and asks for approval like the built-in ones.

INTERACTIVE COMMANDS

/help

List the commands.

/new, /system TEXT

Start the conversation over, keeping the model; with new standing instructions.

/temperature, /top-k, /top-p, /seed, /steps

Set the sampling option; **/settings** shows the current values.

/context, /tokens TEXT

Context window use; tokenisation of *TEXT*.

/undo, /undo-task

Restore the last action, or the whole task; conflicting edits are refused.

/changes, /recover

Show the task's changes; review and undo interrupted work.

/remember TEXT, /memory

Pin a decision; show the pins and the model's notes.

/compact

Archive the transcript and replace it with a summary; also automatic near the context limit. Goals and pins are kept verbatim.

/allowed, /forget N, /never PATTERN

List stored approvals and refusals, drop one, store a refusal.

/tools, /jobs

Offered tools; background commands.

/last, /verbose

Show the latest tool result in full; toggle showing every result in full.

/quit

Exit.

/edit

Write the next task in *\$EDITOR*.

!command

Run *command* in your own shell, outside the sandbox; **!!command** also keeps its output for the next task.

A line typed while the model works is inserted with the next tool result, or becomes the next task if the model has answered. Tab completes commands and paths. Ctrl-C stops the task and keeps the model loaded; Ctrl-D exits.

FILES

~/.config/cmagent/config, .cmagent/config

Options, one per line with its value after it; *#* starts a comment; a bare line names the model. Read in this order; the command line overrides both.

AGENTS.md

Project notes, included in the system turn.

.cmagent/notes.md

The model's notes; the latest 3 KiB open every task.

.cmagent/allowed.json

Stored approvals and refusals for the project.

.cmagent/history

Prompt history.

.cmagent/

Sessions, checkpoints (source files up to 64 MiB) and action evidence.

~/.cache/cmagent/

KV cache of the system turn, one file per model and project.

SEE ALSO

cmgenerate(1), **cmfinetune(1)**, **bwrap(1)**