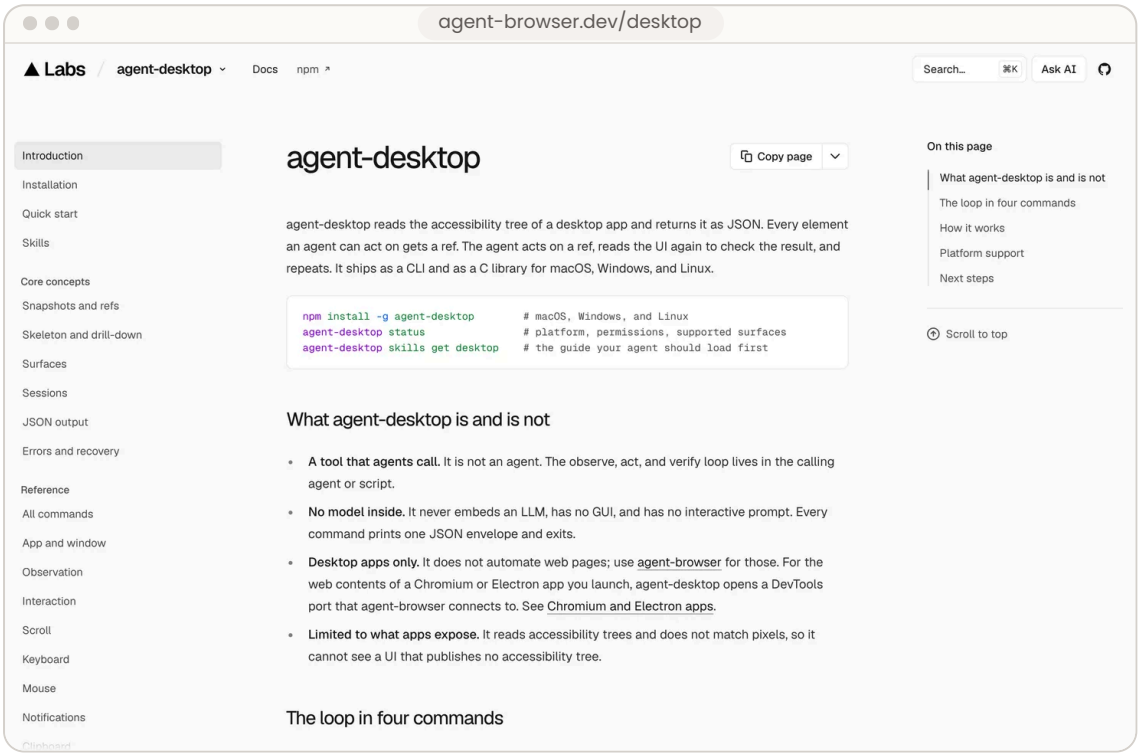
agent-desktop von Vercel

Das kostenlose Tool liest die Bedienelemente deiner Apps aus, statt Screenshots zu raten. Dein KI-Agent verbraucht dafür deutlich weniger Tokens.



Berkan Ceyhan · Stand November 2026



Bilder: Vercel Labs.

AUF EINEN BLICK

- ✓ **Keine Screenshots**: der Agent liest die Bedienelemente
- ✓ **Apps** auf Mac, Windows und Linux
- ✓ **Der Agent klickt** direkt auf das richtige Element
- ✓ **Kostenlos** und Open Source

So bedient dein Agent Apps ohne Screenshots

1 agent-desktop installieren

Du brauchst Node.js. Auf dem Mac musst du dem Tool danach den Zugriff für Bedienungshilfen erlauben, die Anleitung führt dich durch.

Befehle im Terminal

```
npm install -g agent-desktop
agent-desktop status
agent-desktop permissions --guide
```

2 Deinem Agenten die Aufgabe geben

Der Agent holt sich die Anleitung selbst und arbeitet in einem Kreislauf: ansehen, handeln, prüfen.

App mit agent-desktop bedienen

```
Bedien die App [Finder / Slack / Notes / …] mit agent-desktop und erledige diese
Aufgabe: [was du in der App erledigen willst].

So gehst du vor:
1. Lade zuerst die Anleitung für Agenten mit dem Befehl agent-desktop skills get
desktop und lies sie ganz.
2. Prüf mit agent-desktop status, ob alles läuft. Fehlt eine Berechtigung, sag mir
Bescheid und führ mich durch agent-desktop permissions --guide.
3. Arbeite in diesem Kreislauf, bis die Aufgabe erledigt ist:
   a) Mach einen Snapshot der App mit agent-desktop snapshot --app [App] -i. Bei
großen Apps mit vielen Elementen mach erst einen Überblick mit --skeleton und --
compact und geh dann in einen Bereich hinein.
   b) Such im Snapshot das Element, das du brauchst, und handle auf seiner Referenz
(zum Beispiel click oder type).
   c) Mach danach einen neuen Snapshot und prüf, ob die Aktion gewirkt hat.
4. Referenzen gelten nur für ihren Snapshot. Kommt STALE_REF oder AMBIGUOUS_TARGET,
mach einen neuen Snapshot und nimm die neuen Referenzen. Gibt es mehrere Fenster, gib
das Fenster mit --window-id an.
5. Sag mir am Ende in drei Zeilen, was du getan hast, und nenn die Schritte, bei denen
etwas nicht ging.

Regeln:
- Gib Passwörter, Codes und Schlüssel nie als Argument eines Befehls ein. Nutz --stdin
oder frag mich.
- Halt an und frag mich, bevor du etwas löschst, sendest, bezahlst oder zustimmst.
- Gib Snapshots, Zwischenablage und Screenshots nicht in öffentliche Logs oder
Uploads, sie können private Inhalte enthalten.
```

```
- Kann eine App keinen Accessibility-Baum liefern und du findest die Elemente nicht,
sag es mir, statt zu raten.
```

Du willst wissen, wo KI in deinem Betrieb Zeit spart? Buche einen Call mit mir und wir finden es gemeinsam raus.

Call buchen

Statt Pixel zu raten, liest der Agent die Bedienelemente

KURZ ERKLÄRT

Ein Test von Vercel Labs: 198.000 auf 31.000 Tokens

agent-desktop liest den Accessibility-Baum des Betriebssystems, nicht den Quellcode der App. In einem simulierten Test mit 52 Schritten sank der Verbrauch von 198.000 auf 31.000 Tokens, rund sechsmal weniger. Die Zahlen kommen von Vercel Labs selbst, einen unabhängigen Test gibt es noch nicht. Apps ohne Accessibility-Baum bedient das Tool nicht.

agent-desktop: Das Tool von Vercel Labs, Apache-2.0.
vercel-labs/agent-desktop (github.com)

Dokumentation: Schnellstart und alle Befehle.
agent-browser.dev/desktop (agent-browser.dev)



Automatisierungs-Call für Agenturen, Coaches und Dienstleister

In 30 Minuten schauen wir gemeinsam, wo KI-Agenten und automatisierte Prozesse in deinem Betrieb zuerst Zeit und Geld sparen. Kostenlos.

FÜR WEN IST DAS?

- ✓ Agenturen, Coaches und Dienstleister
- ✓ ab 10.000 € Monatsumsatz
- ✓ ein klarer Flaschenhals im Tagesgeschäft
- ✓ du willst KI im Betrieb einsetzen

Call buchen

api.leadconnectorhq.com/widget/booking/mXsszCow3Sxqm4QEnOxK

Scannen & Termin wählen