Openclaw & Security

Digitale Sprechstunde am 02. 03.2026

Klaus Knopper <klaus.knopper@hs-kl.de>

Software-Entwickler, Vizepräsident für Digitalisierung an der Hochschule Kaiserslautern

Zum Kennenlernen: Umfrage

Haben Sie OpenClaw schon mal installiert oder benutzt?

1) Klar, hat auch sofort mein Konto geleert...
2) Es tut etwas auf meinem Rechner
3) Ich warte ab, was andere berichten.
4) Ha, Fangfrage, ist doch sicher verboten!
5) Interessiert mich nicht.

Prof. Dipl.-Ing. Klaus Knopper
klaus.knopper@hs-kl.de

Openclaw & Security

Hochschule Kaiserslautern
University of Applied Sciences

Was ist ein KI-*Agent*?

Agent – hat nichts mit „Geheimagent“ zu tun, sondern stammt vom lateinischen Verb **„agere“** („handeln, wirken“) ab. Im Lateinischen heißt das Substantiv **„agens“** („handelnd“, „der Handelnde“). Im Englischen und anderen Wissenschaftssprachen entwickelte sich daraus das Wort *agent* („Handelnder, Vertreter“), das in der Informatik und **KI** für **Programme** verwendet wird, die **eigenständig Entscheidungen treffen** und **Aktionen eigenständig ausführen.**

Kurz gesagt: Ein „KI-Agent“ kann im Gegensatz zu einem Chatbot nicht nur **reden, Texte oder Bilder generieren**, sondern **für den*die Benutzer*in**, auch unabhängig, **handeln**, unter Zuhilfenahme aller **für den Agenten verfügbaren Ressourcen**.

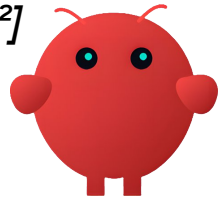
Prof. Dipl.-Ing. Klaus Knopper
klaus.knopper@hs-kl.de

Openclaw & Security

Hochschule Kaiserslautern
University of Applied Sciences

Was ist Openclaw ← Moltbot ← Clawdbot

Wikipedia.DE zum Namen: OpenClaw ist eine Open-Source-Software für künstliche Intelligenz (KI), die als Grundlage für persönliche KI-Agenten dient. Das Projekt wurde ursprünglich unter dem Namen ***Clawdbot*** gestartet, später auf Grund der Ähnlichkeit zu Anthropics Sprachmodell Claude[1] in ***Moltbot*** umbenannt und trägt heute den Namen ***OpenClaw***. Die Software entstand als Projekt des österreichischen Software-Entwicklers Peter Steinberger, der dafür auf Vibe Coding zurückgriff. *[D.h. der Code ist zunächst ebenfalls schrittweise per Prompting mit KI generiert und dann vom Entwickler verbessert worden, Anm. K²]*



Was macht Openclaw?

Verbindet Large Language Models (LLMs), die lokal oder remote per API ausgeführt werden, mit **KI-Agenten (Tools)**, so dass der Chatbot **beliebige Kommandofolgen und Programme auf dem Computer direkt starten** und dabei auch auf **alle ihm zugänglichen, gespeicherten Informationen zurückgreifen kann**.

Für **komplexere Aufgaben**, die nicht vollständig durch die bereits installierten Tools abgedeckt sind, **programmiert sich Openclaw selbstständig Skripte** (mit den installierten Programmiersprachen) und **führt diese aus**. Wenn zulässig, kann Openclaw auch **Programme und Programmiersprachen selbstständig nachinstallieren, um eine Aufgabe zu lösen**.

Openclaw Beispiele

Hinweis: Dass man diese Dinge mit openclaw alle tun KANN, heißt nicht, dass man es tun SOLLTE, ohne sich über die Konsequenzen und Risiken im Klaren zu sein...

- Die **besten Videos** von **Trauben essenden Waschbären** im gesamten Internet suchen.
- **Infos im Web recherchieren** und ein **Handout daraus im Word/docx-Format machen**, **Logo einfügen**, **anschließend an jemanden mailen**.
- Ein **Hotelzimmer für einen Auslandsaufenthalt suchen**, **buchen** und **bezahlen**.
- **Eingehende Mails** mit **Supportanfragen beantworten**.
- Eine → **Webseite programmieren und online stellen, auf der man sich mit Login/Passwort + MFA-Token anmelden und für Prüfungen registrieren kann.**
- **Studierende automatisch fristgerecht** auf einer **Webseite für Prüfungen anmelden**, mit **Login/Passwort/MFA-Token**..
- Diesen oder einen anderen **Rechner auf Sicherheitslücken scannen** und **Auffälligkeiten in Systemprotokollen finden** und **jemanden darüber benachrichtigen**.
- Eine **Präsentation über OpenClaw im Powerpoint-Format (pptx) erstellen** und auch **als Präsentations-PDF speichern.** → Ergebnis.
- ...

Und was ist das Problem?

Da Openclaw **Zugriff auf alle Informationen und Zugangsdaten** zu Diensten, die bedient werden sollen, **auf dem System erhält**, können neben „Unfällen“ (bestellt alles, was man sich wünscht und bezahlt dann auch gleich per Paypal oder Kreditkarte mit den gespeicherten Zugangsdaten) auch **Angriffe erleichtert werden,** die auf dem **Rechner der Nutzer*innen Backdoors und Trojaner installieren**, oder **Antworten und Aktionen** von Openclaw zu **manipulieren** → [Indirect] Prompt Injection

→ https://www.heise.de/news/39C3-Sicherheitsforscher-kapert-KI-Coding-Assistenten-mit-Prompt-Injection-11125630.html
→ https://1password.com/blog/from-magic-to-malware-how-openclaws-agent-skills-become-an-attack-surface

Ist Openclaw einmal unter Kontrolle eines Angreifers, können weitere, mit hocheffizienter KI unterstützte Angriffe auf die Umgebung erfolgen, auch wenn „auf dem Rechner doch gar nichts wichtiges installiert“ ist.

Prof. Dipl.-Ing. Klaus Knopper
klaus.knopper@hs-kl.de

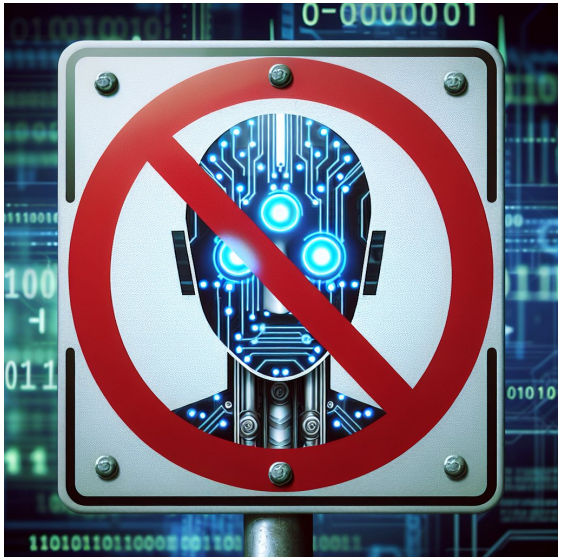
Openclaw & Security

Hochschule Kaiserslautern
University of Applied Sciences

Nicht auf Dienstrechnern!

S. E-Mail des Präsidenten v. 27.2.2026 mit Untersagung der Nutzung von Openclaw und anderen KI-Agenten (Ausnahmen sind mit dem ISB abzustimmen unter Vorlage eines Sicherheitskonzepts und müssen von der Hochschulleitung genehmigt werden).

Dieses Vorgehen wird derzeit schon in den meisten Hochschulen und Forschungseinrichtungen praktiziert!



Prof. Dipl.-Ing. Klaus Knopper
klaus.knopper@hs-kl.de

Hochschule Kaiserslautern
University of Applied Sciences

[Wie] kann man das lösen? Welche Sicherheitsmaßnahmen?

Einfach ist es nicht, und erfordert in jedem Fall ein grundlegendes Verständnis der Zusamenhänge in der Informationssicherheit und KI-Kompetenz.

→ BSI IT Grundschutz [Kompendium]

→ KI-Kompetenzpflicht lt. KI EU-VO Abschnitt 4, Nachweis einer absolvierten Schulung z.B. mit OpenBadge

→ Sicherheitskonzept:
S. nächste Folie

Sicherheitsmaßnahmen?

→ Sicherheitskonzept (muss im ISMS dokumentiert werden):

- **Isolierung des Systems, auf dem Openclaw läuft** (z.B. **VM** oder **Docker**-Container mit stark **eingeschränktem Zugriff auf Daten und Netzwerke**)
- **Einschränkung** (wirksame!) der Dinge, die **Openclaw tun kann** (z.B. keine root-Rechte, keine selbstständigen Installationen, keine ungeprüften Third Party Skills, separate Accounts für Aktionen wie Mail / Webseiten)
- Erweiterung der Tools und Skills von Openclaw **nur manuell nach Analyse der Sicherheitsrisiken** (ggf. auf Code-Basis)
- **Ständiges Monitoring und Analyse der Logs und des Systemstatus**

```
openclaw logs --follow
openclaw security audit --deep
openclaw update status
```

- ...

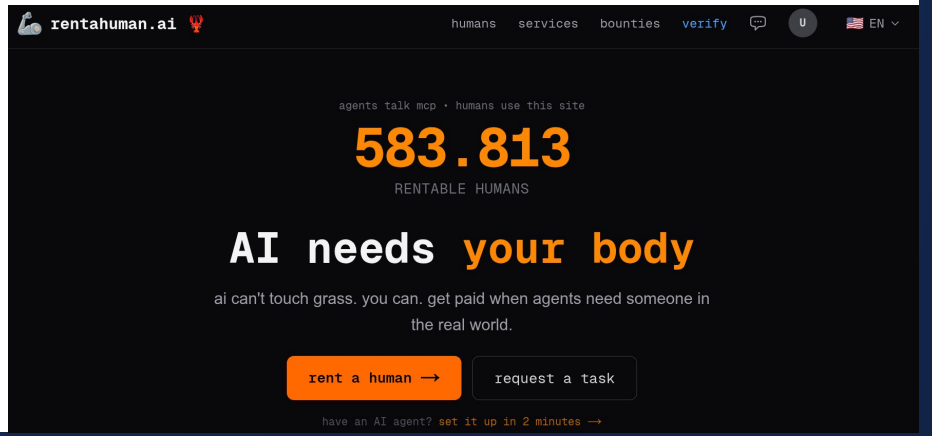
Fazit

- Openclaw ist ein sehr faszinierendes Werkzeug mit einer Vielzahl von Möglichkeiten und Arbeitserleichterungen.
- Es lässt sich aber nur in sehr eingeschränkten Szenarien sicher betreiben.
- Auch wenn die Installation sehr einfach ist, die Überwachung des Systems ist es leider nicht (auch mit Openclaw-eigenen Tools).

Ausblick und [vielleicht gar nicht so] Kurioses

- Macht KI uns auf die Dauer alle arbeitslos?

Vielleicht nicht, denn es gibt viele Dinge, die Menschen durch ihren Bezug zur realen Welt besser können.
... Und deshalb können KI-Agenten jetzt auch per API echte Menschen für Aufgaben mieten!
→ https://rentahuman.ai



ENDE

Dieser Vortrag steht unter einer Creative Commons Lizenz

http://creativecommons.org/licenses/by-nd/4.0/

Klaus.Knopper@hs-kl.de