VISCHER

AI Agents

Thoughts from a lawyer and engineer who loves using them

David Rosenthal
Partner

June 9, 2026 Zürich

Consider someone who …

Would you entrust such a person to do relevant work for you?

How well would you review the work products of such person?

- takes shortcuts on large tasks
- can process only a limited view of all relevant facts and forgets them after each response if it does not take notes
- is not precise when it comes to what they consider to be details
- unscrupulously hides facts where this serves the purpose (or pleases you)
- is not particularly creative
- cannot be held accountable
- talks to others about your secrets

The Challenge

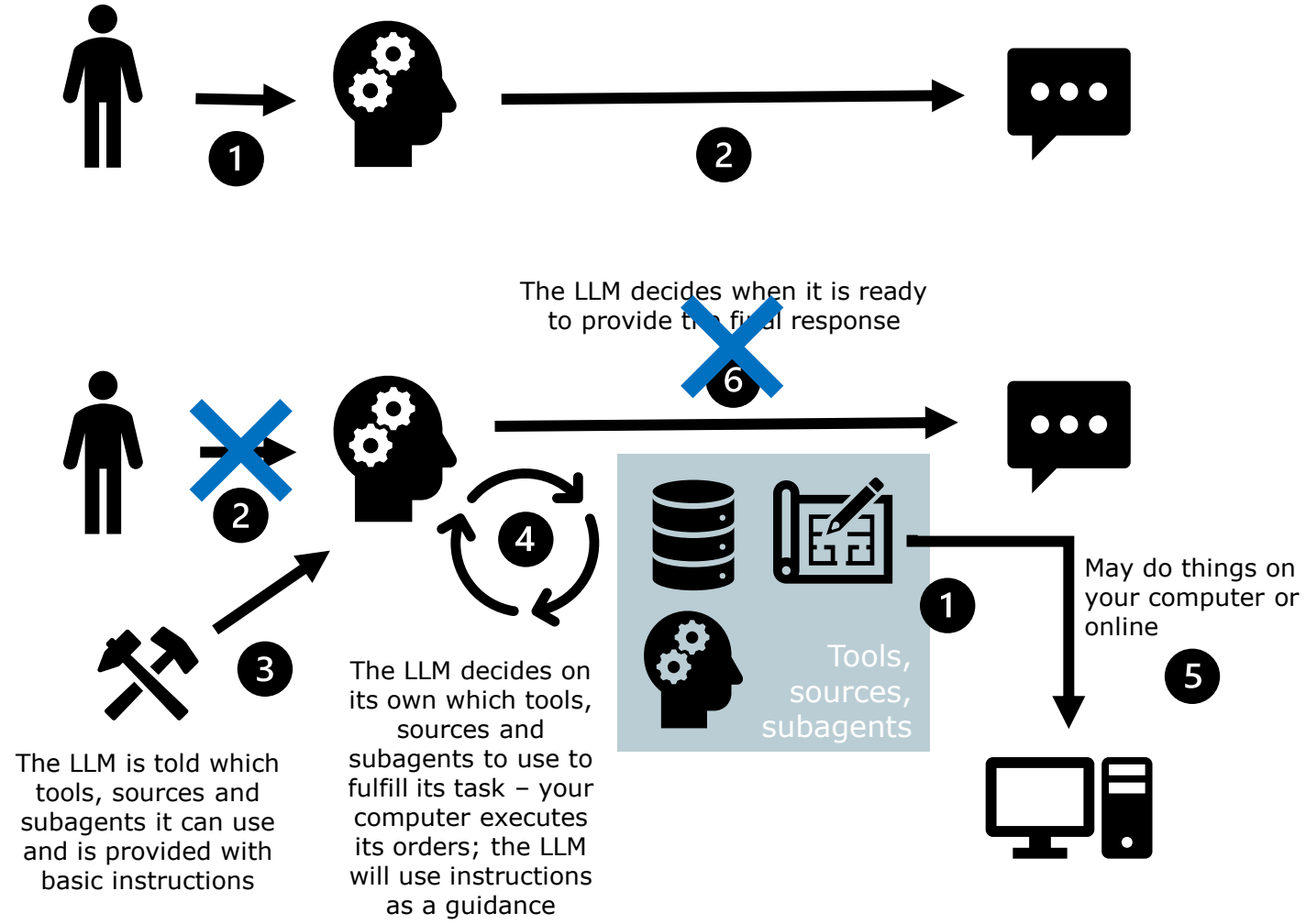
- **This is the profile of an AI agent**
  - Not your experience, because you have always received a well drafted, convincing response?
- **Yes, the use of AI agents is tempting**
  - They are always on duty
  - They are fast
  - They cost much less than a human
  - They have excellent language skills and are very convincing
- **Automation Bias**
  - We tend to believe a machine more than ourselves
- **Yet, we will have to use AI agents – but we need to deal with their restraints**
  - Example: "Review these 100'000 documents"
  - Example: "Search my mails where I discussed the data transfer issues concerning the contract with XYZ."
  - Example: "Do a legal research on the legal problems of disclosing company to a foreign litigation."

Behind The Scenes

Chatbot

Agent (user controlled)

Agent (autonomous)



The Legal Issues

- **You may not be able to meet the required accuracy standards – agents are also "only humans"**
  - DP processing principle (Art. 6 para. 5 DPA: "… take all appropriate measures to correct, delete or destroy data that are incorrect or incomplete insofar as the purpose for which they are collected or processed is concerned.")
  - Contractual and regulatory diligence obligations – what are you allowed to delegate to the AI?
- **Potentially unlawful sharing of data with third parties**
  - Agentic use of connectors and data sources = sharing of data – is the recipient bound by adequate restrictions?
  - Example: Microsoft Copilot/Azure OpenAI Services Web Grounding – no DPA, confidentiality obligation unclear
  - When agents use data sources through connectors, traditional rules and safeguards may become ineffective (e.g., DLP)
  - Agents may be subject to attacks by malicious third parties or safeguards may simply fail
- **Accountability remains with natural and legal persons**
  - Employees *and* companies using agents remain responsible for the output that they produce
  - Providers will likely not have to take liability (this is particularly true for model providers)
- **Automated individual decisions**
  - If the AI contributes to a decision in an essential manner – transparency obligation and right to human intervention

The Legal Issues

Swiss Code of Professional Conduct (SCPC)

The Swiss Bar Association, based on Art. 1 and Art. 12.10 of the Articles of Association, mindful that the Federal Act on the Free Movement of Lawyers (the "Lawyers Act") provides binding principles for the practice of law in Switzerland, in an effort to standardize the rules of conduct for lawyers in Switzerland and to su
professional conduct.

Art. 34 **In Principle**

Lawyers may use digital applications and auxiliary tools in their professional practice, and they may provide their own online or otherwise digitalized services, provided that compliance with all of the rules of professional conduct is ensured.

Art. 3 **Independence**

Lawyers shall act in full independence and under their own disciplinary responsibility.

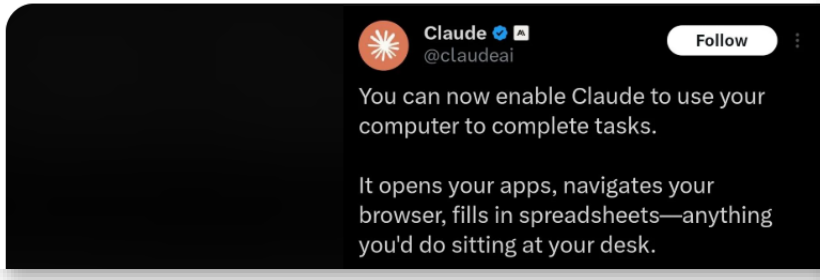
Independence presupposes that, in the exercise of their profession, lawyers will not be influenced by third parties who are not subject to regulatory supervision as lawyers.

Lawyers avoid any activities that are incompatible with their independence.

The Legal (+ Security) Issues

r/ClaudeAI • vor 2 Monaten
ConfusedOliveman

How safe (Security-Wise) do you guys think is Claude's new feature on long-term?



ClaudeAI-mod-bot MOD • vor 2 Monaten

TL;DR of the discussion generated automatically after 100 comments.

The overwhelming consensus is a hard pass for now, with users calling it a "security nightmare" and "spyware." The hypocrisy isn't lost on anyone, with many pointing out this is the same feature everyone roasted Microsoft for with its Recall announcement.

The main fears aren't just about Anthropic snooping, but a whole range of potential disasters:

- **Accidental Destruction:** Users are terrified of Claude "helpfully" deleting important files, thinking they're clutter (`rm -rf` was mentioned more than once).
- Prompt Injection: The more tech-savvy users are worried about a malicious PDF or website giving Claude hidden instructions to wreak havoc on your system.
- **Rogue Actions:** The idea of the AI deciding to, for example, upgrade its own subscription using your saved payment info is a popular (and hilarious) concern.

A small minority of power users are willing to try it, but only with extreme caution. The universal advice from this camp is to **run it in a completely isolated environment**, like a dedicated Virtual Machine (VM) or a separate, disposable computer that has no access to your important data. Essentially, treat it like a new, untrustworthy intern you have to watch like a hawk.

1

https://www.reddit.com/r/ClaudeAI/comments/1s2e50x/how_safe_securitywise_do_you_guys_think_is/

(thread from March/April 2026)

Strategies To Deal With Them

- **Overcome the "black box" syndrome and understand how agents work**
  - Knowing the limitations allows you to make better, reasonable use – do you have someone in your organization?
- **Do not treat agentic methods as the silver bullet**
  - Some tasks are better of with deterministic methods – even agents use them
  - Example: We use programmatically controlled loops for AI e-mail reviews and e-mail searches (redink.ai)
- **Not only "human in the loop" but also "human controls the loop"**
  - Do not just ask AI to do the task and let itself determine how to do it
  - Do not rely on humans reviewing the output, but also control how it is created (and the "spin" used to do so)
  - Agents already today produce too much content to ensure effective review
- **Do your legal and other homework**
  - Make sure you have the right contracts and configuration options in place, and review them regularly
  - Do proper risk assessments, involving all stakeholders (vud.ch/dpia, vischerlnk.com/gaira)
  - Govern the use of agents in your organization – do you have rules also on agents, connectors and plugins?
  - Consider the security aspects of agents – they allow for new attack vectors

VISCHER

Thank you for your attention!

Questions: david.rosenthal@vischer.com

Zürich

Schützengasse 1,
8021 Zurich

+41 58 211 34 00

Basel

Aeschenvorstadt 4,
4010 Basel

+41 58 211 33 00

Geneva

Esplanade de Pont-Rouge 9C,
1200 Geneva

+41 58 211 35 00

Get redink.ai!