

EU AI-Act

Notarisierung von Dialogen mit (High-Risk) KI-Systemen

AK Blockchain

Dr. Christian Baumann

28.5.2026

Agenda

- EU AI Act - Überblick
- Ziel und Geltung
- Risikoklassen
- In-Geltung-Treten
- Hochrisiko-KI (High-Risk)
- Protokollierung -> Notarisierung
- Proof Of Concept

EU AI Act - Überblick

- EU-weites Gesetz zur Regulierung von Künstlicher Intelligenz
- Ziel: einheitliche Regeln für Entwicklung & Einsatz von KI in der EU
- Gilt für Anbieter, Entwickler, Importeure und Betreiber von KI-Systemen
- Fokus: Risikobasierter Ansatz (von minimal bis verboten)

EU AI Act - Ziel und Geltung

- Zweck: Sicherheit, Grundrechte, Transparenz & Vertrauen in KI stärken
- Pflichten steigen mit dem Risiko (besonders bei Hochrisiko-KI)
- Betrifft grundsätzlich private Unternehmen & öffentliche Stellen
- Sanktionen bei Verstößen: bei Nicht-Einhaltung drohen hohe Geldstrafen

EU AI Act - 4 Risikoklassen von KI-Systemen

Der EU AI Act teilt KI-Systeme nach ihrem Risiko für Menschen & Grundrechte in vier Stufen ein:

- **Unannehmbares Risiko (verboten):** z.B. biometrische Kategorisierung, Social Scoring, „manipulative KI“
- **Hohes Risiko (High-Risk):** z.B. Kreditwürdigkeit, Medizin, kritische Infrastruktur → umfangreiche Liste von Pflichten
- **Begrenztes Risiko:** z.B. Chatbots, KI generierte Inhalte (auch Deepfakes) kennzeichnen → Transparenzpflichten
- **Minimales Risiko:** z.B. Spamfilter, Games → kaum zusätzliche Pflichten

In-Geltung-Treten

- 2.2.2025: Regelungen für verbotene KI-Systeme (Nutzung einstellen), Beginn nachweisliche Schulung von Mitarbeitern
- 2.8.2025: Dokumentations- & Informationspflicht für Anbieter (allgemeine Verwendungszwecke)
- 2.8.2026: Übrige Verpflichtungen: Transparenzpflichten sowie Strafbestimmungen. Übergangsfrist von 36 Monaten für High-Risk

Vgl. <https://www.wko.at/digitalisierung/ai-act-eu>

EU AI Act - Hochrisiko-KI (High-Risk)

- Hochrisiko-KI betrifft Systeme, die stark in zentrale Lebensbereiche eingreifen und dadurch **Grundrechte, Sicherheit oder Gesundheit** gefährden können.
- Beispiele: Recruiting/Personal, Bildung/Prüfungen, Kreditwürdigkeit, Medizin, kritische Infrastruktur, Justiz/Strafverfolgung.
- Pflichten: Risikomanagement, Tests, Dokumentation, Transparenz, menschliche Aufsicht.

Hochrisiko-KI (High-Risk) - Protokollierung

- Art. 12:
 - „... **internen Protokollierungsfunktionen** ... Ereignisse während ihres gesamten Betriebs automatisch zu erfassen“
- Zwecke
 - Risikoerkennung
 - Beobachtung nach dem Inverkehrbringen
 - Überwachung des Betriebs
- Rahmenbedingungen Protokollierung
 - Sichere Speicherung
 - Zugriffskontrolle
 - **Nachvollziehbarkeit, Nachweis der Integrität**

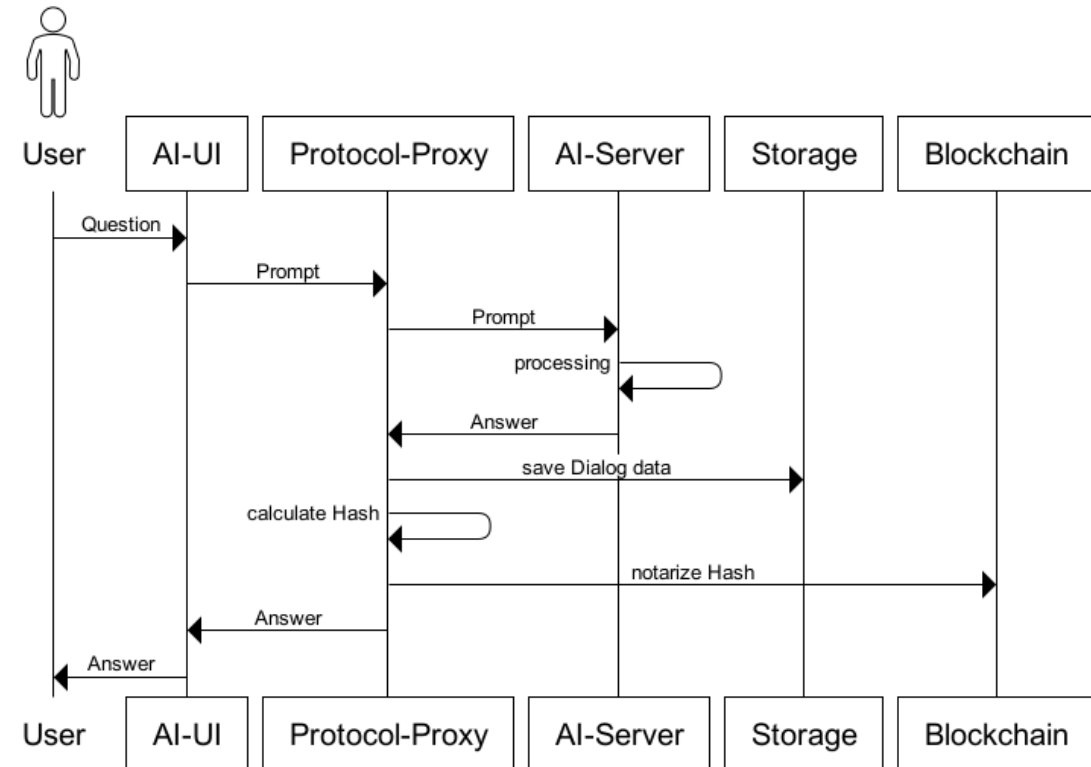
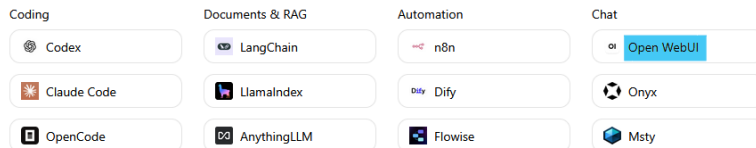
Vgl. <https://www.vde.com/topics-de/kuenstliche-intelligenz/blog/eu-ai-act--protokollierung-von-ki-systemen>

Protokollierung -> Notarisierung

- Der EU AI Act (bzw. Literatur dazu) verweist auf **integritätsgesicherte Protokollierung** mittels „Hash-Chains“.
- Teilbereich: **Userprompts** und dadurch entstandene Reasonings („Gedanken der AI“) & **Antworten der KI**
- Ansatz
 - **Vollautomatische Blockchain basierte Notarisierung**
 - Ein Service (zB. „Proxy-Server“) protokolliert die Dialoge zwischen User und KI und notarisiert die Logs automatisch
 - → Integrität & Nachvollziehbarkeit sind gesichert.

Proof Of Concept 1 - Aufbau

- Open WebUI
- Proxy in PHP/curl
- AI-Server
 - Ollama (on premise)
 - LLM zB. mistral
 - Apache License
 - <https://ollama.com/library/mistral>



Proof Of Concept 1 - UI & Protokollierung

owu.bmnn.at/c/d2bd1067-d299-47d2-bd7b-325b91ed479

mistral:7b

Write a 50-word summary of artificial intelligence

OI mistral:7b

Artificial Intelligence (AI) is a branch of computer science that aims to create intelligent machines capable of mimicking human intelligence. It involves the development of algorithms and statistical models that can learn, reason, and make decisions similar to humans. AI applications range from voice recognition and natural language processing to autonomous vehicles and advanced robotics.

Folgefragen

Can you give an example of a common AI application?

What are some challenges in the development of AI?

How does AI differ from machine learning and deep learning?

Are there any ethical concerns associated with AI?

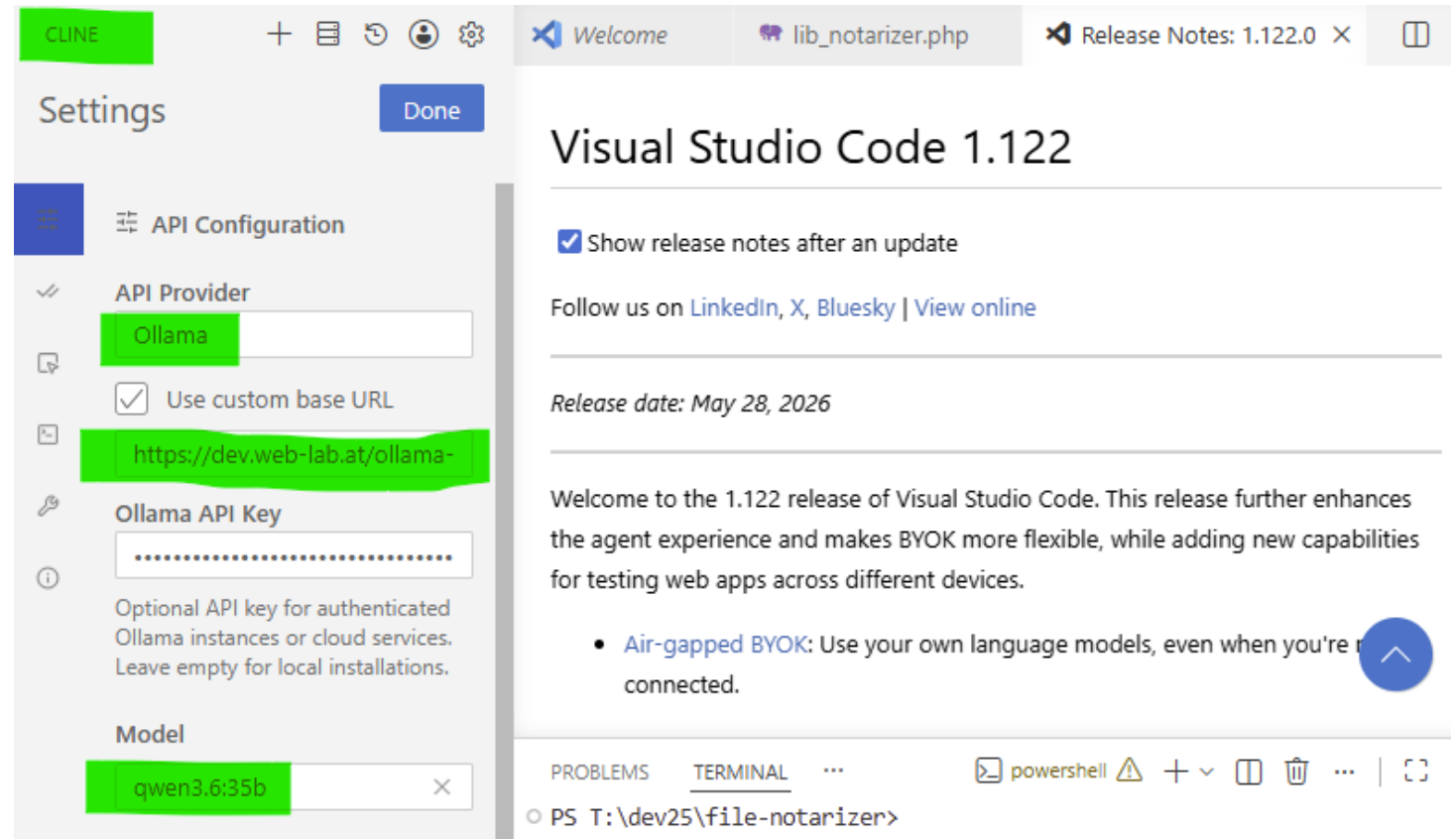
Eine Nachricht senden

```
{
  "id": "1f21d996c878a87e7d0fff9b0980fa23",
  "timestamp": "2026-01-07T12:52:55.346898Z",
  "client_ip": "194.48.239.254",
  "request": {
    ...
    "body": "{\r\n \"model\": \"mistral:7b\", \r\n \"prompt\":\r\n \"Write a 50-word summary of artificial intelligence\", \r\n \"stream\":\r\n false\r\n}"
  },
  "response": {
    ...
    "body": "{\r\n \"model\": \"mistral:7b\", \"created_at\": \"2026-01-07T12:53:00.9106148Z\", \"response\": \"Artificial Intelligence (AI) is a branch of computer science that aims to create intelligent machines capable of mimicking human intelligence. It involves the development of algorithms and statistical models that can learn, reason, and make decisions similar to humans. AI applications range from voice recognition and natural language processing to autonomous vehicles and advanced robotics.\\\"...\"
  }
}
```

Proof Of Concept 2 - („Vibe“-Coding

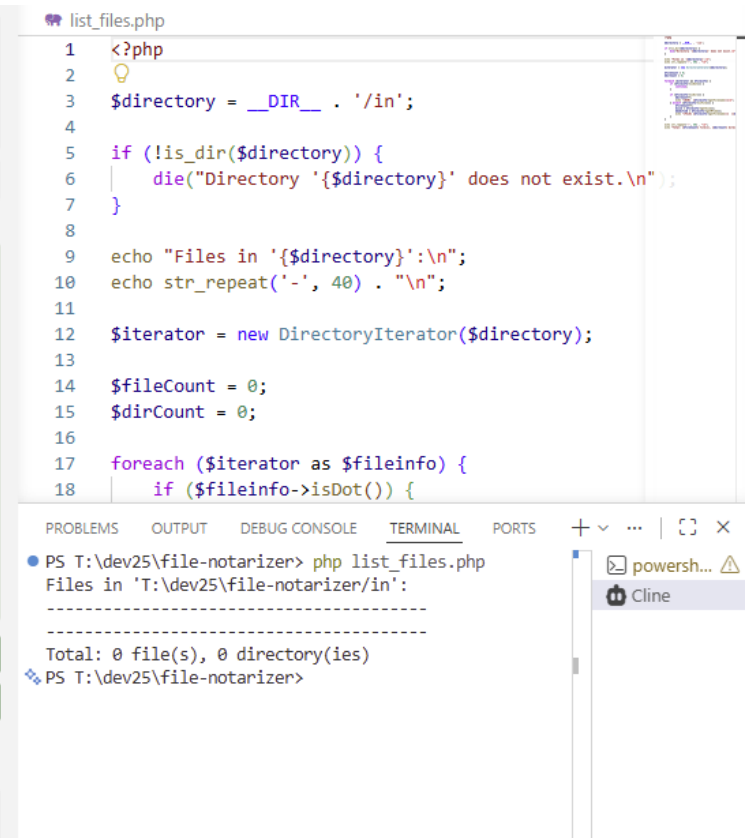
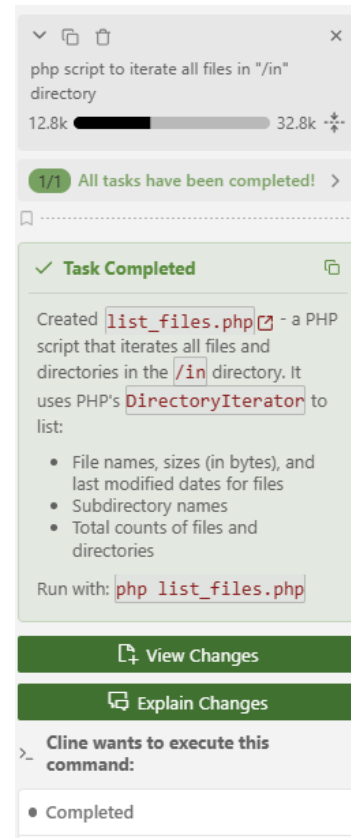
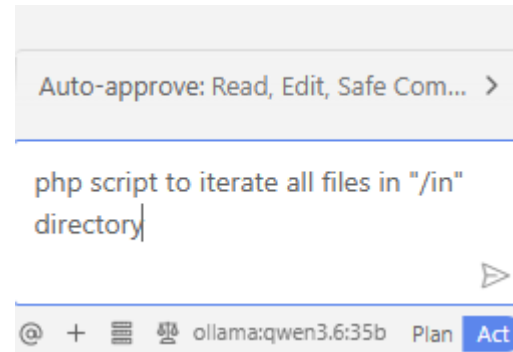
- VS-Code
- „cline“ Extension
- LLM OnPremise
- Protocol-Proxy

“Vibe coding” is not a recognized software engineering methodology. It’s colloquial slang that can be used positively (celebrating flow, creativity, and pragmatism) or humorously/critically (implying lack of structure or professionalism). Its meaning is entirely contextual to modern dev culture and AI-assisted workflows.



Proof Of Concept 2 - „Vibe-Coding“

- Prompt ...
- Thinking
- Acting
- Testing
- -> OneShot ;-)



Proof Of Concept 2 - generierter Code

list_files.php

```
1 <?php
2
3 $directory = __DIR__ . '/in';
4
5 if (!is_dir($directory)) {
6     die("Directory '{$directory}' does not exist.\n");
7 }
8
9 echo "Files in '{$directory}':\n";
10 echo str_repeat('-', 40) . "\n";
11
12 $iterator = new DirectoryIterator($directory);
13
14 $fileCount = 0;
15 $dirCount = 0;
16
17 foreach ($iterator as $fileinfo) {
18     if ($fileinfo->isDot()) {
19         continue;
20     }
21
22     if ($fileinfo->isDir()) {
23         $dirCount++;
24         echo "[DIR] {$fileinfo->getFilename()}\n";
25     } elseif ($fileinfo->isFile()) {
26         $fileCount++;
27         $size = $fileinfo->getSize();
28         $modified = $fileinfo->getMTime();
29         echo "[FILE] {$fileinfo->getFilename()} ({$size} bytes, modified: " . date('Y-m-d H:i:s', $modified) . ")\n";
30     }
31 }
32
33 echo str_repeat('-', 40) . "\n";
34 echo "Total: {$fileCount} file(s), {$dirCount} directory(ies)\n";
```

Proof Of Concept 2

- Protocol-File includes
 - Systemprompt
 - Thinking
 - Acting
 - Meta-Infos
 - Zeitstempel
 - IP-Adressen
 - Versionen

```
k
{"id": "c1faceaf32bfc3369c612caf9e81c628",
"timestamp": "2026-05-28T09:44:48.857254Z",
"client_ip": "94.136.7.8",
"request": {
  "method": "POST",
  "path": "/ollama-proxy/v1/api/chat",
  "query": "",
  "headers": {
    "Accept-Encoding": "br, gzip, deflate",
    "Sec-Fetch-Mode": "cors",
    "Accept-Language": "*",
    "Authorization": "Bearer 9c0abe51c6e66555d82de2d044d4fb194931f058c042
6c67c7285d8f5657ed64a",
    "User-Agent": "ollama-js/0.5.18 (x64 win32 Node.js/v22.22.1)",
    "Accept": "application/json",
    "Content-Type": "application/json",
    "Content-Length": "57016",
    "Connection": "close",
    "Host": "dev.web-lab.at",
    "X-Forwarded-Port": "443",
    "X-Forwarded-Proto": "https",
    "X-Real-Ip": "94.136.7.8",
    "X-Forwarded-By": "192.168.10.42"
  },
  "body": "{\n  \"model\": \"qwen3.6:35b\",\n  \"messages\": [\n    {\n      \"role\": \"system\",\n      \"content\": \"You are Cline, a highly skilled software engineer with extensive knowledge in many programming languages, frameworks, design patterns, and best practices.\\n\\nTOOL USE\\n\\nYou have access to a set of tools that are executed upon the user's approval. You can use one tool per message, and will receive the result of that tool use in the user's response. You use tools step-by-step to accomplish a given task, with each tool use informed by the result of the previous tool use.\\n\\n# Tool Use Formatting\\n\\nTool use is formatted using XML-style tags. The tool name is enclosed in opening and closing tags, and each parameter is similarly enclosed within its own set of tags. Here's the structure:\\n\\n<tool_name>\\n<parameter1_name>value1</parameter1_name>\\n<parameter2_name>value2</parameter2_name>\\n...\\n</tool_name>\\n\\nFor example:\\n\\n<read_file>\\n<path>src/main.js</path>\\n<task_progress>\\nChecklis t here (optional)\\n</task_progress>\\n</read_file>\\n\\nAlways adhere to this format for the tool use to ensure proper parsing and execution.\\n\\n# Tools\\n\\n## execute_command\\nDescription: Request to execute a CLI command on the system. Use this when you need to perform system operations or run specific commands to accomplish any step in the user's task. You must tailor your command to the user's system and provide a clear explanation of what the command does.
```


Proof Of Concept 2 - Ergebnis

- SW-Entwicklung (AI assisted coding)
- Protokollierung aller Infos
 - Aufgabenstellung (Prompt) des Entwicklers
 - Thinking-Prozess des LLMs
 - (Zwischen-) Ergebnisse = Sourcecode
 - Automatisch ausgeführte Tests & Ergebnisse
 - Meta-Informationen
- -> AI-Act Art. 12 erfüllt
 - „... internen **Protokollierungsfunktionen** ... Ereignisse während ihres gesamten Betriebs automatisch zu erfassen“
 - Sichere Speicherung
 - **Nachvollziehbarkeit, Nachweis der Integrität durch Notarisierung**

Zusammenfassung

- EU AI Act fordert
 - integritätsgesicherte Protokollierung
 - ibs. bei High-Risk Anwendungen
- Datenströme zwischen AI-Anwendung (User) und AI-Server müssen automatisiert protokolliert und gesichert gespeichert werden.
- Blockchain basierte Notarisierung (Verspeicherung von Hashwerten der Protokoll-Daten) sichert die Integrität (Unveränderbarkeit und Zeitstempel).

www.austriapro.at
austriapro@wko.at

DI Dr. Christian Baumann
c.baumann@baumann.at
+43 664 43 24 243

AUSTRIA / PRO