TeleTrusT
Pioneers in IT security.
Bundesverband IT-Sicherheit e.V.

IT-Sicherheitsrechtstag 2026

Berlin, 24.09.2026

Vom Prompt zur Web-App: Vibe Coding trifft IT-Sicherheit

Janos Rumpel, DCB DevConsult Berlin

CVSS 3.1 · CRITICAL

9.3

Unauthentifizierter Lese- und Schreibzugriff
auf die Datenbanken generierter Seiten.

SCHWACHSTELLE	CWE-863 Incorrect Authorization
OWASP 2025	A01 Broken Access Control
BETROFFEN	Lovable bis 15.04.2025

CVE-2025-48757

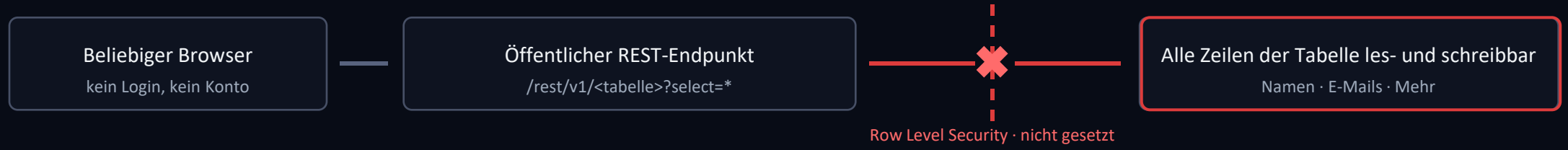
$ cat befund.txt

// Plattform für Apps ohne Entwickler
Eine fehlende Zeile Datenbankkonfiguration.
Row Level Security nicht gesetzt.

// Wortlaut NVD
“…allows remote unauthenticated attackers
to read or write to arbitrary database
tables of generated sites.”

// Reaktion des Herstellers
Der Einstufung wird widersprochen.

WAS GENAU GEFEHLT HAT



NVD, CVE-2025-48757 · CVSS 3.1 Base Score 9.3 · CWE-863 · veröffentlicht 29.05.2025 · Hersteller widerspricht · Beleg [1]

01 · DAS PHÄNOMEN

Vibe Coding, ehrlich definiert

definition.md

Software in natürlicher Sprache beschreiben,

statt sie zu schreiben.

und das Ergebnis akzeptieren,

ohne jede Zeile zu prüfen.

Über den ersten Teil müssen wir nicht reden. Über den zweiten schon.

Tools der Kategorie: Lovable · Bolt · v0 · Claude Artifacts · Codex

01 · DAS PHÄNOMEN

Nicht die IT baut. Die Fachabteilung.

KANZLEI

baut das Mandantenportal für den Dokumenten-Upload, weil der Versand per E-Mail zu unsicher war.

VERTRIEB

baut den Angebotsrechner, weil das Excel niemand mehr versteht.

Ohne Ticket, ohne Budget, ohne Freigabe. Und der Grund heißt jedes Mal gleich:

Sie mussten liefern, und der offizielle Weg war zu langsam.

02 · DER BEFUND

Wie oft baut ein Modell eine bekannte Schwachstelle ein, wenn im Prompt nichts über Sicherheit steht?

44%

dieser 80 Testaufgaben erzeugten Code mit einer bekannten Schwachstelle. Die Syntax stimmte in 99,9 % der Fälle.

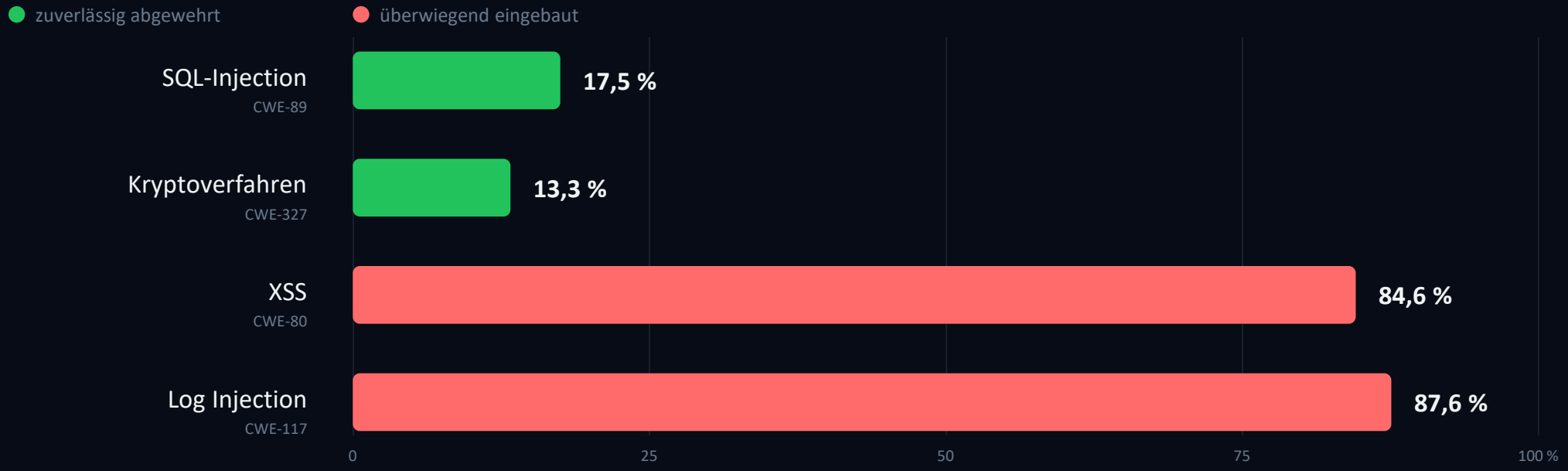
testaufbau

```
$ veracode --report=2026

11 Modelle im Sommer-2026-Datensatz
100+ Modelle über die gesamte Serie
80 Coding-Aufgaben, 4 Sprachen, 4 CWE

// Auswertung per SAST gegen MITRE CWE
// Prompts bewusst ohne Security-Hinweis
```

02 · DER BEFUND

Die berühmten Lücken sind beherrschbar

Anteil der Aufgaben mit eingebauter Schwachstelle, nach Schwachstellentyp



SQL-Injection sitzt seit 25 Jahren in jedem Lehrbuch. Log Injection nicht.

Errechnet als 100 – Pass-Rate (82,51 · 86,71 · 15,36 · 12,43 %) · Veracode 2026 · Beleg [2]

02 · DER BEFUND

Syntax gelöst. Sicherheit nicht.

Über 100 Modelle nach Erscheinungsdatum · Sicherheit in vier Messungen: ≈55 % → 56,3 %

GRÖSSER HILFT NICHT

über 100 Mrd. Parameter 53,3 %

unter 20 Mrd. 51,1 %

SPEZIALISIERT HILFT NICHT

Coding-Modelle 51,3 %

Allzweck-Modelle 52,4 %

DER NACHFOLGER WAR SCHLECHTER

Claude Opus 4.7 61 %

Claude Opus 4.8 51 %

„Warten wir das nächste Modell ab“ ist keine Strategie. **Oder doch ein Mythos?**

Trendlinien nachgezeichnet aus Veracode 2026, Kap. 2 „Security and Syntax Pass Rates vs LLM Release Date“ · Messreihe: 3× ≈55 % → 56,3 % · Beleg [2]

02 · DER BEFUND · AUSBLICK

„Die neue KI findet doch jede Lücke.“

200 echte Schwachstellen. Der erzeugte Patch wird zweimal geprüft.

PRÜFUNG 1 · FUNCPASS

Laufen die Tests des Projekts danach noch durch?

59,8%

der Patches: ja

PRÜFUNG 2 · SECPASS

Scheitert der Angriff, der die Lücke vorher ausgenutzt hat?

19,0%

der Patches: ja

und selbst die Treffer …

```
// 38 der 200 Lösungen kamen nicht aus
// Können, sondern aus Erinnerung:
33× Patch aus den Trainingsdaten
4× Lösung im Projekt gefunden
1× Antwort aus der Git-Historie
// einer davon Zeichen für Zeichen
identisch mit dem Original.
```

Vier von fünf Patches sehen aus wie ein Fix, laufen wie ein Fix und schließen die Lücke nicht.

Lücken aufspüren kann dieselbe Modellgeneration sehr gut: **73 %** der Expertenaufgaben gelöst. Finden ja. Schließen nein.

Endor Labs, 200 reale Fixes mit Claude Code + Fable 5: FuncPass 59,8 % · SecPass 19,0 % · Beleg [3] · AISI, Mythos Preview, 13.04.2026 · Beleg [4]

03 · DIE FALLEN

Vier Fallen, immer dieselben

01 · SECRETS IM CLIENT

Der Schlüssel liegt im Browser.

A02:2025 Security Misconfiguration

02 · FEHLENDE AUTORISIERUNG

Geprüft wird, wer Sie sind. Nicht, was Sie dürfen.

A01:2025 Broken Access Control · Platz 1

03 · PAKETE AUS DEM NICHTS

Abhängigkeiten, die niemand ausgewählt hat.

A03:2025 Supply Chain Failures · neu 2025

04 · KEIN LOGGING

Unsicher. Und im Zweifel nicht nachweisbar.

A09:2025 Logging & Alerting Failures

Vier Fallen, vier Kategorien der OWASP Top 10:2025. Keine hat die KI erfunden.

OWASP Top 10:2025 · Datenbasis 2,8 Mio. Anwendungen, 13 Organisationen · nicht zu verwechseln mit den OWASP Top 10 for LLM Applications · Beleg [5]

04 · DER BEWEIS

Ich habe es selbst gebaut.

Ein Prompt. Vier Minuten. Eine öffentliche App mit echter Datenbank. Ohne Ticket, ohne Freigabe.

prompt.txt

Baue eine Web-App „Feedback-Kiosk“
für eine Konferenz.
Besucher geben **Name, E-Mail** und einen
Kommentar ein.
Alle Einträge erscheinen darunter in
einer Liste mit Zeitstempel.
Speichere die Daten in Supabase.
Schlichtes, modernes Design.

// kein Wort über Sicherheit.
// genau wie im echten Leben

SICHERHEITS-SCAN · CRITICAL

Feedback submitter emails publicly exposed

„The feedback_entries table allows anyone (anon role) to read all rows, exposing submitter names and email addresses to any unauthenticated visitor.“

A01:2025 Broken Access Control · dieselbe CWE-863 wie Folie 02

Ich habe die Lücke nicht gesucht. Eine Maschine hat sie mir in 30 Sekunden gezeigt.

Eigenes Demo-Projekt, gebaut mit Lovable + Supabase · Befund im Wortlaut des Scanners

04 · DER BEWEIS

Kein Einzelfall.

380.000

öffentlich erreichbare Anwendungen, gebaut mit Lovable, Base44, Replit und Netlify

5.000

davon mit sensiblen Unternehmensdaten

WAS TATSÄCHLICH OFFEN LAG

- Eine Reederei: welche Schiffe in welchen Häfen erwartet werden
- Kliniken: Zusammenfassungen von Arzt-Patienten-Gesprächen
- Ein Küchenhersteller: unredigierte Kundenservice-Verläufe
- Schulen: Unterrichtsaufnahmen und Schülerdaten

Keine dieser Firmen wurde angegriffen. Es hat nur nie jemand nachgesehen.

RedAccess, Mai 2026, von Axios unabhängig verifiziert · zusätzlich Phishing-Seiten, die Bank of America, FedEx und McDonald's imitieren · Beleg [8]

04 · DER BEWEIS

Es wurde nicht besser. Nur größer.

Dieselbe Plattform, ein Jahr Abstand: Lovable.

2025 · ERSTER VORFALL

Fehlende Row Level Security in generierten Apps. CVSS 9.3.

Im Nachgang: 303 Endpunkte in 170 Projekten ohne wirksame Regel.

Der Hersteller-Scanner prüfte, ob eine Regel existiert. Nicht, ob sie etwas verhindert.

2026 · ZWEITER VORFALL

Quellcode, Datenbank-Zugangsdaten und KI-Chatverläufe fremder Projekte über die Schnittstelle lesbar.

Offen vom 3. Februar bis 20. April 2026, nach Angaben des Herstellers.

Meldungen wurden geschlossen, weil die Prüfstelle das Verhalten für beabsichtigt hielt.

GEMEINSAM IST BEIDEN DIE ERSTE REAKTION DES HERSTELLERS

Reaktion 2025: „Der Einstufung wird widersprochen.“

Reaktion 2026, nachträglich: „Our first public response was dismissive.“

NVD/Palmer 2025 · Lovable, „Our response to the April 2026 incident“, eigene Angaben · The Register, 21.04.2026 · Belege [1] [9]

ZUM ABFOTOGRAFIEREN

Fünf Sätze, die den Unterschied machen

security-prompt-template.txt

Härte diese App gegen die häufigsten Sicherheitsfehler:

1. Zugriff standardmäßig verweigern, nur explizit Benötigtes freigeben. Bei Datenbanken: Autorisierung pro Tabelle und Operation. Anonyme Schreibzugriffe nur, wenn fachlich nötig, dann mit Missbrauchsschutz.
2. Eingaben serverseitig validieren, Datenbankabfragen parametrisieren, Ausgaben kontextgerecht encodieren. Für Passwörter und Verschlüsselung aktuelle Verfahren verwenden, keine abgekündigten.
3. Keine Secrets in Client, Repository, Logs oder Prompts. Personenbezogene Daten nur speichern und ausgeben, wenn sie wirklich gebraucht werden.
4. Nur existierende, gepflegte Abhängigkeiten verwenden. Liste alle neu hinzugefügten Pakete mit Version auf.
5. Nenne zum Schluss: umgesetzte Maßnahmen, getroffene Annahmen und offene Risiken. Sag ausdrücklich, was ein Mensch prüfen muss.

Punkt 5 macht die App nicht sicher. Er macht sie prüfbar.

05 · DIE REGELN

Wo die Guardrails greifen

Regel 01 + 02

Template und Secrets, vor der Generierung.

MODELL

Black Box. Hier greifen Sie nicht ein.

Regel 03

Ein Mensch mit Namen sieht drauf.

Regel 04

Scanner laufen bei jedem Commit.

NUTZER

Das Ziel. Hier darf keine Lücke ankommen.

Die zwei Kästen ohne Regel sind kein Zufall: Modell und Nutzer können Sie nicht kontrollieren. Nur alles dazwischen.

05 · DIE REGELN

Fünf Regeln für die Praxis

01 **Der Prompt ist die Spezifikation.**

Was nicht drinsteht, ist nicht bestellt.

---

02 **Secrets nie in den Prompt, nie in den Client.**

Und einen Rotationsplan, bevor Sie ihn brauchen.

---

03 **Ein Mensch mit Namen pro App.**

Nicht „das Team“. Eine Person, die auf den Code schaut.

---

04 **Maschinencode wird von Maschinen geprüft.** gleich im Detail →

Manuelles Review allein ist bei diesem Tempo chancenlos.

---

05 **Ein KI-Anwendungsregister statt eines Verbots.** gleich im Detail →

Ein Verzeichnis aller Apps, die mit KI gebaut wurden.

Regel 4 und 5 brauchen mehr als einen Satz. Die beiden schauen wir uns gleich genauer an.

REGEL 04

Maschinencode prüfen Maschinen

🔍

CODE

Semgrep

🔑

SECRETS

gitleaks

📦

ABHÄNGIGKEITEN

osv-scanner

der ganze Aufwand

```
$ semgrep --config auto . && gitleaks detect && osv-scanner .
// drei Zeilen in der Pipeline · Lizenzkosten: keine
```

Eine Stufe davor: **nicht alles muss das Modell schreiben.** Projektgerüst, Abhängigkeiten und wiederkehrende Bausteine liefern die etablierten Werkzeuge zuverlässiger. Was nie erzeugt wurde, muss auch niemand prüfen.

REGEL 05

Das KI-Anwendungsregister

42 % wissen oder vermuten KI-Nutzung ohne Freigabe. **23 %** haben Regeln dafür.

Eine Tabelle reicht für den Anfang.

ANWENDUNG	URL / HOSTER	REPO	VERANTWORTLICH	GEPRÜFT
Mandantenportal	portal.lovable.app	github.com/…/portal	M. Weber	08/2026
Angebotsrechner	rechner.intern.de	gitlab.intern/vertrieb	T. Kern	07/2026
Urlaubstool	urlaub.vercel.app	keines	offen	nie

ZWEI PFLICHTEN, EIN DOKUMENT

Zugleich Asset-Management nach NIS2 und Vorstufe zur SBOM, die der CRA ab 2027 verlangt.

DIE UNBEQUEME ZEILE

Zeile 3: kein Repository, kein Verantwortlicher, nie geprüft.

Bitkom Research, repräsentativ, 604 Unternehmen ab 20 Beschäftigten, 21.10.2025 · Beleg [10]

Das Tempo ist der richtige Weg.

Die Prüfung muss nur mitwachsen.

DREI TIPPS ZUM SCHLUSS

01

Fragen Sie nach, welche Apps schon existieren.

02

Schreiben Sie ein Prompt-Template ins Intranet.

03

Schalten Sie einen Scanner in eine Pipeline.

DANKE

Janos Rumpel

Geschäftsführer · DevConsult Berlin

📍 Uhlandstraße 32 · 10719 Berlin

🌐 devconsult.berlin

Auf LinkedIn vernetzen →

Kein Quishing. 🙃

janos-rumpel

linkedin.com/in/janos-rumpel

ANHANG · A · BELEGE 1 BIS 6

Quellen und Belege

[1]	NVD, CVE-2025-48757 · 29.05.2025	„…allows remote unauthenticated attackers to read or write to arbitrary database tables of generated sites.“ (Lovable) CVSS 3.1: 9.3 · CWE-863 → OWASP A01:2025. Hersteller widerspricht.
[2]	Veracode, 2026 GenAI Code Security Report · 28.07.2026	80 Testaufgaben, 4 Sprachen, 4 CWE, 11 Modelle (100+ über die Serie), Prompts ohne Sicherheitshinweis. Ø Pass-Rate 56,3 %, Syntax 99,9 %. CWE-327 86,7 · CWE-89 82,5 · CWE-80 15,4 · CWE-117 12,4 %.
[3]	Endor Labs, Fable-5-Benchmark · 08/2026	Harness Claude Code + Modell Fable 5 auf 200 realen Fixes: FuncPass 59,8 %, SecPass 19,0 %; 38/200 reproduzierten Trainingsdaten.
[4]	AI Security Institute (UK), Claude Mythos Preview · 13.04.2026	73 % der Experten-CTF-Aufgaben gelöst; Ø 22 von 32 Schritten (Opus 4.6: 16). Fazit: „investment now in cyber defence is vital“. Andere Prüfung als [3], kein direkter Vergleich.
[5]	OWASP Foundation, OWASP Top 10:2025	Datenbasis 2,8 Mio. Anwendungen, 13 Organisationen. Die vier Fallen = A02 Security Misconfiguration · A01 Broken Access Control · A03 Supply Chain Failures (neu) · A09 Logging & Alerting. Nicht die LLM-Liste.
[6]	NIS2UmsuCG · verkündet 06.12.2025	Gilt seit Dez 2025, ohne Übergangsfrist. 29.500 Unternehmen, Schwelle 50 MA / 10 Mio €. BSI-Frist verlängert auf 31.07.2026. § 30 BSIG: Lieferkettensicherheit.

Belege 7 bis 12 auf der nächsten Folie · Volltext: QUELLEN.md

ANHANG · B · BELEGE 7 BIS 12

Quellen und Belege

[7]	Verordnung (EU) 2024/2847 (CRA)	Meldepflichten ab 11.09.2026 (24 h / 72 h / 14 Tage) über die ENISA-Plattform. Vollständig anwendbar ab 11.12.2027, dann mit SBOM.
[8]	RedAccess, via Axios · 07.05.2026	380.000 öffentlich erreichbare Assets aus Lovable, Base44, Replit, Netlify · davon ~5.000 mit sensiblen Firmendaten. Axios verifizierte unabhängig. Ursache u. a.: Sichtbarkeit stand standardmäßig auf öffentlich.
[9]	Lovable, „Our response to the April 2026 incident“ · The Register 21.04.2026	Quellcode, DB-Zugangsdaten und KI-Chatverläufe fremder Projekte lesbar, 03.02.–20.04.2026 (eigene Angabe). Erst: „We did not suffer a data breach.“ Dann: „Our first public response was dismissive.“
[10]	Bitkom Research · 21.10.2025	604 Unternehmen ab 20 Beschäftigten, repräsentativ. Private KI-Nutzung ohne Freigabe: 8 % weit verbreitet (2024: 4 %), 17 % Einzelfälle, 17 % vermuten es. Regeln für KI-Tools: 23 %.
[11]	IBM · Cost of a Data Breach 2026	Anteil der Sicherheitsvorfälle mit Schatten-KI mehr als verdoppelt, auf 43 %. Über zwei Drittel der Organisationen ohne Governance-Prozess.
[12]	Werkzeuge aus Regel 4	Semgrep (SAST) · gitleaks (Secrets, auch in der Git-History) · osv-scanner (Abhängigkeiten gegen die OSV-Datenbank). Grundfunktion jeweils kostenfrei.

Volltext aller Belege: QUELLEN.md im Vortragspaket