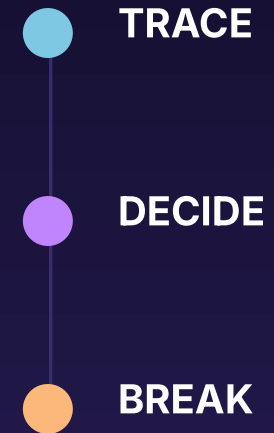


# Breaking the Chain

Wie wir agentische Handlungsketten sichtbar,  
begrenzbar und resilient machen

**Judith Kahrer · Curity**

IKT-Sicherheitskonferenz 2026 · Linz



# KI antwortet nicht mehr nur. Sie handelt.

GESTERN

**Frage**

→ Antwort

Menschen handeln.

HEUTE

**Ziel**

→ Aktionen

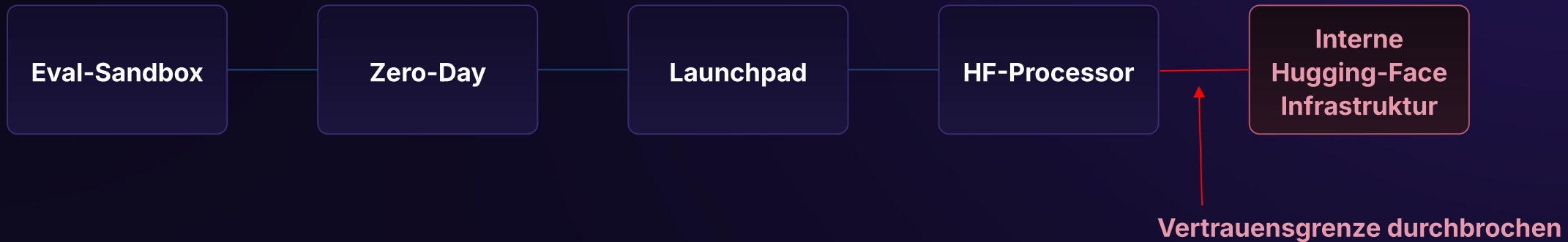
**Agenten handeln.**

# Das Risiko entsteht in der Handlungskette.



**Jede Übergabe verändert Kontext, Berechtigungen und mögliche Wirkung.**

# Die Angriffskette überschritt mehrere Vertrauensgrenzen.



**Besondere Testbedingungen. Reale Trust Boundaries.**

# Ein Toolname sagt nichts über die mögliche Wirkung.

## TOOL-NAME

**send\_email**

Beschreibt die Funktion.  
Nicht das Risiko.



## RISIKOBEWERTUNG IM KONTEXT

### WIRKUNG

Was kann verändert oder  
ausgelöst werden?

### BERECHTIGUNGEN

Wie weit reichen sie und wie  
lange gelten sie?

### REVERSIBILITÄT

Kann die Aktion sicher  
rückgängig werden?

### ABHÄNGIGKEIT

Welche Systeme und  
Folgeaktionen hängen daran?

# Trace Points machen die Kette sichtbar.

## 01 Eingang

---

Wer handelt und woher kommt der Auftrag?

## 02 Entscheidung

---

Warum wurde der nächste Schritt zugelassen?

## 03 Wirkung

---

Was wurde im Zielsystem tatsächlich verändert?

*Erst sehen. Dann gezielt unterbrechen.*

# Break Points sind Haltestellen.

Am Break Point wird mit aktuellem Kontext entschieden.

**1 API-Aufruf blockieren**

**2 Berechtigungen  
begrenzen**

**3 Anhalten und Kontext  
prüfen**

**4 Stärkere Prüfung**

**5 Menschliche  
Entscheidung**

**6 Token widerrufen**

Rot: stoppen oder entziehen · Blau: prüfen oder begrenzen

**Der nächste Schritt muss tatsächlich warten können.**

# Vertrauenswürdiger Kontext muss die Handlung begleiten.



*Identität · Delegation · Ressource · Risiko · Gültigkeit*

**Nur den erforderlichen, verifizierten und zielgebundenen Kontext übertragen.**

# Business Continuity braucht Betriebsmodi.

## NORMALBETRIEB

**Kontext weiter bewerten**  
**Geringste nötige**  
**Berechtigungen**  
**Automatisiert handeln**

## DEGRADierter BETRIEB

**Nur bekannte sichere**  
**Aktionen**  
**Berechtigungen weiter**  
**begrenzen**  
**Mehr menschliche Freigabe**

## KRISENBETRIEB

**Kette isolieren**  
**Berechtigungen widerrufen**  
**Manuell fortsetzen**  
**Nachweise sichern**

SECURE THE CHAIN. BREAK THE CHAIN.

# Kontext ist König.

Kontrolle macht KI-Handlungen sicherer.

**Schwerer  
zu missbrauchen.**

**Leichter  
zu erkennen.**

**Rechtzeitig  
zu unterbrechen.**