



CHECKLIST

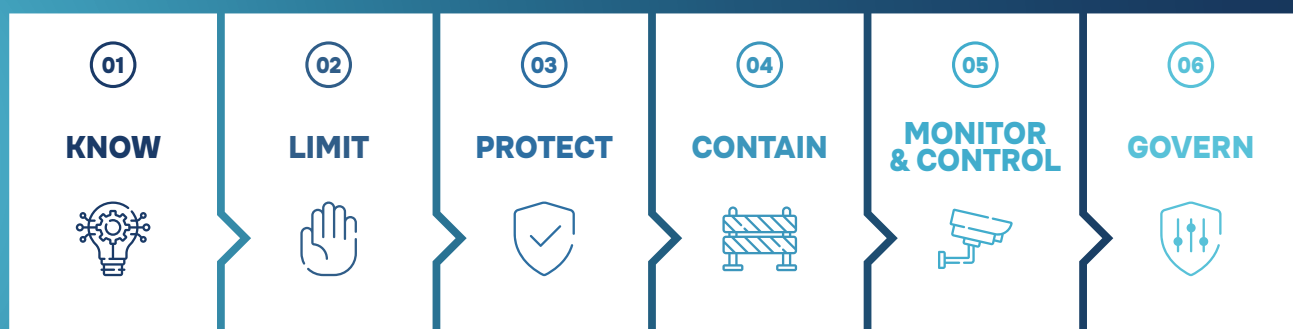


WIE SICHER IST IHRE UMGEBUNG FÜR AGENTIC AI?

WIE SICHER IST IHRE UMGEBUNG FÜR AGENTIC AI?

KI beantwortet nicht mehr nur Fragen. Agentische Systeme können über lokale Komponenten, Schnittstellen oder Nutzerkontexte näher an Dateien, Anwendungen und Workflows rücken – und dort selbst Aktionen anstoßen. Der Endpoint wird damit zu einem Ort, an dem KI nicht nur Informationen verarbeitet, sondern Entscheidungen vorbereitet und selbstständig umsetzt.

Für die IT-Sicherheit verändert sich dadurch die Kernfrage: Es reicht nicht mehr zu wissen, welche Anwendung laufen darf. Entscheidend ist, welche Fähigkeiten ein Agent besitzt, welche Ressourcen er erreichen kann und was er im Namen eines Nutzers oder Systems tatsächlich tun darf.





Welche Agenten und Fähigkeiten gibt es?

Sind KI-Agenten und ihre Fähigkeiten bekannt?

Warum wichtig?

Nur was bekannt ist, kann bewertet und abgesichert werden.

Bei Agentic AI reicht ein klassisches Software-Inventar nicht aus:

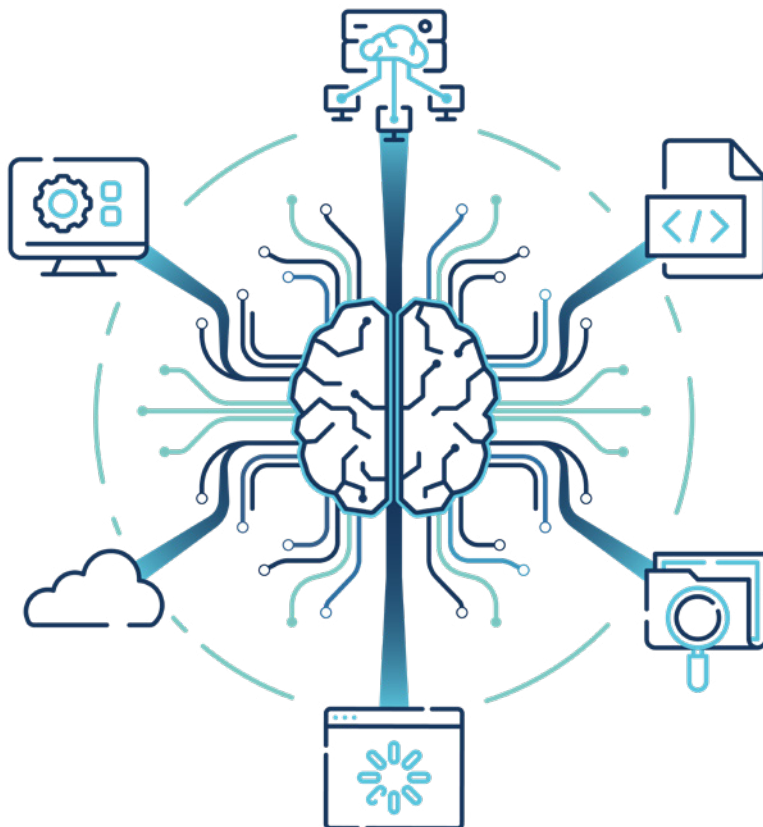
Auch Skills, Plugins, MCP-Server und angebundene Tools bestimmen den Handlungsspielraum eines Agenten.

- ☐ Eingesetzte und pilotierte KI-Agenten sind zentral erfasst.
- ☐ Verwendete Skills, Plugins, MCP-Server und Tools sind bekannt.
- ☐ Einsatzbereiche, betroffene Systeme und Verantwortlichkeiten sind dokumentiert.

Was tun?

Eine zentrale Übersicht schaffen:

Agent → Fähigkeiten → Einsatzbereich → Verantwortlichkeit.



Merksatz: Man kann nicht kontrollieren, was man nicht kennt.

LIMIT

Welche Rechte und Handlungsmöglichkeiten braucht ein Agent wirklich?

Ist die Handlungsmacht auf das notwendige Maß begrenzt?

Warum wichtig?

Klassisches Least Privilege betrachtet vor allem Berechtigungen.

Bei agentischen Systemen kommt eine zweite Dimension hinzu:

Welche Aktionen kann ein Agent mit diesen Rechten tatsächlich durchführen – insbesondere dann, wenn er selbstständig einen Weg zum Ziel auswählt?

- ☐ Der Agent kann nur die Tools und Funktionen nutzen, die für seine Aufgabe erforderlich sind.
- ☐ API-Zugriffe, Berechtigungen und Freigaben sind auf den notwendigen Umfang begrenzt.
- ☐ Nicht mehr benötigte Zugriffe und Freigaben werden regelmäßig überprüft oder entzogen.

Was tun?

Least Privilege um Least Agency ergänzen: Rechte und tatsächliche Handlungsmöglichkeiten begrenzen.

Merksatz: Je mehr Autonomie, desto enger sollten die erlaubten Handlungsräume sein.

Sie möchten mehr über Agentic AI Control Erfahren?

Dann sprechen sie mit einem unserer Experten.
Klicken Sie einfach auf den unten stehenden Button
oder nutzen sie den QR-Code um einen persönlichen
Gesprächstermin zu vereinbaren.



Jetzt Gesprächstermin Vereinbaren

Auf welche Daten, Anwendungen und Tools darf ein Agent zugreifen?

Werden externe Inhalte als potenziell manipulative Eingaben behandelt?

Warum wichtig?

Agenten reagieren nicht nur auf direkte Benutzer-Prompts. Webseiten, Dokumente, Bilder, Audio oder andere Inhalte können Entscheidungen und nachfolgende Aktionen beeinflussen. Deshalb dürfen externe Inhalte nicht automatisch als vertrauenswürdig gelten.

- ☐ Externe und nicht vertrauenswürdige Inhalte werden als potenziell manipulative Eingaben behandelt.
- ☐ Systemanweisungen, Benutzereingaben und externe Inhalte werden möglichst klar voneinander getrennt.
- ☐ Kritische Aktionen werden unabhängig davon kontrolliert, welche Anweisung ein Agent aus einem Inhalt ableitet.

Was tun?

Nicht allein darauf vertrauen, Manipulation immer erkennen zu können. Entscheidend ist, den möglichen Schaden technisch zu begrenzen.

Ist kontrolliert, was ein Agent tatsächlich tun darf?

Warum wichtig?

Der Endpoint ist dort besonders relevant, wo Rechte zu Aktionen werden: Agenten können Dateien lesen oder verändern, Prozesse starten, APIs ansprechen, Browser oder weitere Tools einbinden und externe Ziele kontaktieren. Erlaubte Software allein ist deshalb kein ausreichendes Sicherheitsmodell.

- ☐ Erlaubte Dateien, Verzeichnisse, Anwendungen, APIs und Tools sind definiert.
- ☐ Es ist festgelegt, was der Agent lesen, schreiben, verändern, starten oder kontaktieren darf.
- ☐ Besonders kritische Zugriffe sind technisch eingeschränkt oder genehmigungspflichtig.

Merksatz: Nicht nur: „Darf diese AI-Anwendung laufen?“ – sondern: „Was darf sie tatsächlich tun?“

Sind sensible Daten und ihre Zugriffswege geschützt?

Warum wichtig?

Agenten können Daten aus lokalen Dateien, Anwendungen, APIs oder Cloud-Diensten verwenden und weitergeben. Sicherheitskontrollen sollten deshalb nicht nur die Daten selbst betrachten, sondern auch, welche Quellen erreichbar sind und wohin Informationen fließen können.

- ☐ Es ist bekannt, auf welche Datenquellen ein Agent zugreifen kann.
- ☐ Besonders schützenswerte Daten und Speicherorte werden entsprechend geschützt.
- ☐ Relevante Datenflüsse zu externen Diensten, APIs oder anderen Zielen werden kontrolliert.

Was tun?

Nicht nur prüfen, wer Daten nutzen darf – sondern welche Daten einem Agenten überhaupt erreichbar gemacht werden sollen.



Wie weit kann sich eine Fehlfunktion oder Manipulation auswirken?

Können riskante Aktionen technisch begrenzt oder gestoppt werden?

Warum wichtig?

Auch ein wohlmeinender Agent kann falsche Entscheidungen treffen. Dadurch können Vertraulichkeit, Integrität oder Verfügbarkeit beeinträchtigt werden. Monitoring allein verhindert keinen Schaden; wirksame Guardrails müssen dort greifen können, wo der Agent tatsächlich handelt.

- ☐ Besonders riskante Aktionen können technisch begrenzt oder blockiert werden.
- ☐ Besonders folgenreiche Aktionen benötigen bei Bedarf eine zusätzliche Freigabe.
- ☐ Der Handlungsspielraum eines Agenten kann bei auffälligem Verhalten schnell eingeschränkt werden.

Was tun?

Abgestufte Kontrollen vorsehen: Allow, Limit, Approve, Block oder Ringfence. überhaupt erreichbar gemacht werden sollen.

Merksatz: Nicht nur Fehler verhindern – ihre möglichen Auswirkungen begrenzen.

Was passiert tatsächlich – und können Sie eingreifen?

Sind sicherheitsrelevante Agentenaktivitäten nachvollziehbar?

Warum wichtig?

Je selbstständiger ein Agent handelt, desto wichtiger wird die Nachvollziehbarkeit seiner Aktionen. Entscheidend sind insbesondere Tool-Aufrufe, Datenzugriffe, Änderungen, Kommunikationsziele und ungewöhnliche Verhaltensmuster.

- ☐ Sicherheitsrelevante Aktionen, Tool-Aufrufe und Zugriffe werden protokolliert.
- ☐ Aktivitäten lassen sich einem Agenten und seinem Ausführungskontext zuordnen.
- ☐ Ungewöhnliche oder unerwartete Verhaltensweisen können erkannt und untersucht werden.

Was tun?

Relevante Entscheidungen und Aktionen auditierbar machen – statt wahllos alles zu protokollieren.

Welche Kontrollen sind für welches Risiko angemessen?

Werden Schutzmaßnahmen an das tatsächliche Risiko angepasst?

Warum wichtig?

Nicht jeder Agent benötigt dieselbe Kontrolltiefe. Entscheidend sind unter anderem Autonomiegrad, verfügbare Tools, Datenzugriff, Sensitivität der Daten und möglicher Business Impact. Ändert sich der Einsatzbereich eines Agenten, muss auch seine Absicherung mitwachsen.

- ☐ Es gibt verbindliche Regeln für Einführung und Betrieb von Agentic AI.
- ☐ Die Kontrolltiefe berücksichtigt Autonomie, Datenzugriff und möglichen Business Impact.
- ☐ Schutzmaßnahmen werden überprüft, wenn sich Agent, Tools, Datenquellen oder Einsatzbereich verändern.

Merksatz: Mehr Autonomie + mehr Zugriffe + sensiblere Daten = stärkere Guardrails.

Schnellbewertung

20–24 Häkchen:

Gute Basis – Kontrollen gezielt weiterentwickeln.

12–19 Häkchen:

Relevante Lücken – Prioritäten definieren und schließen.

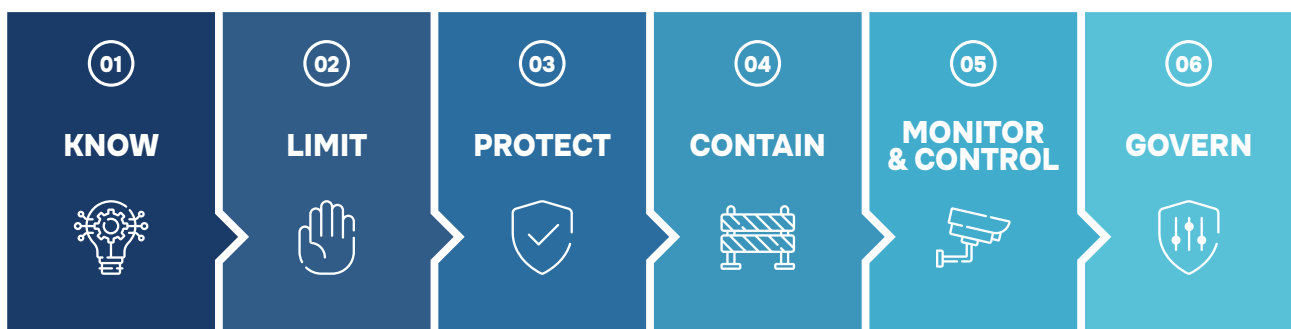
0–11 Häkchen:

Hoher Handlungsbedarf – Mit Visibility und Zugriffskontrolle beginnen.

Die Schnellbewertung dient als Orientierung und ersetzt keine formale Risikoanalyse oder Zertifizierung.

Leitplanken statt KI-Blockade

Das Ziel ist nicht, Agentic AI pauschal zu verhindern. Produktive KI-Nutzung braucht klare, wirksame und nachvollziehbare Grenzen. Bewährte Sicherheitsprinzipien bleiben dabei relevant – sie müssen auf neue Artefakte und neue Handlungsketten übertragen werden: AI-Tools, Skills, MCP-Server, Memory-Dateien, Plugins, APIs und die Aktionen, die daraus entstehen.



DriveLock SE

Landsberger Straße 396
81241 München

+49 (89) 546 36 49-0
info@drivelock.com

ÜBERZEUGEN SIE SICH

Jetzt unverbindlich
30 Tage gratis testen



Sprechen Sie mit
unseren Experten

