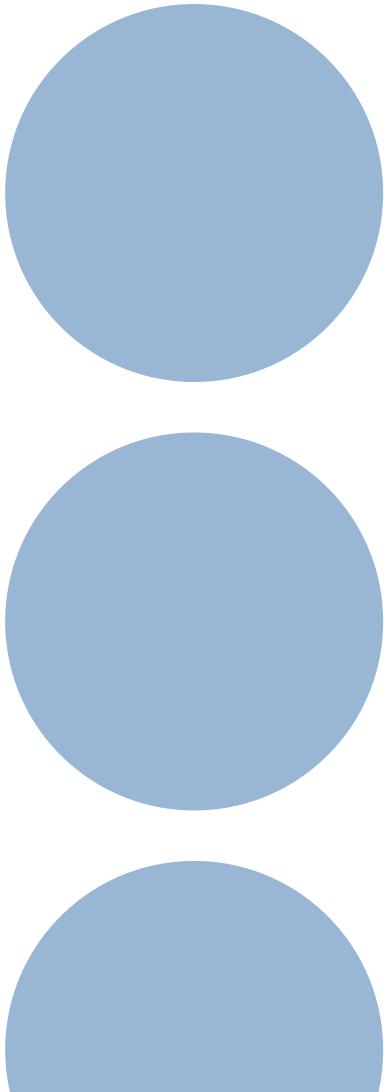
BGHM
Berufsgenossenschaft
Holz und Metall

Ihre gesetzliche Unfallversicherung

Assistenzsysteme & KI: Was Konstrukteure über den Menschen wissen müssen



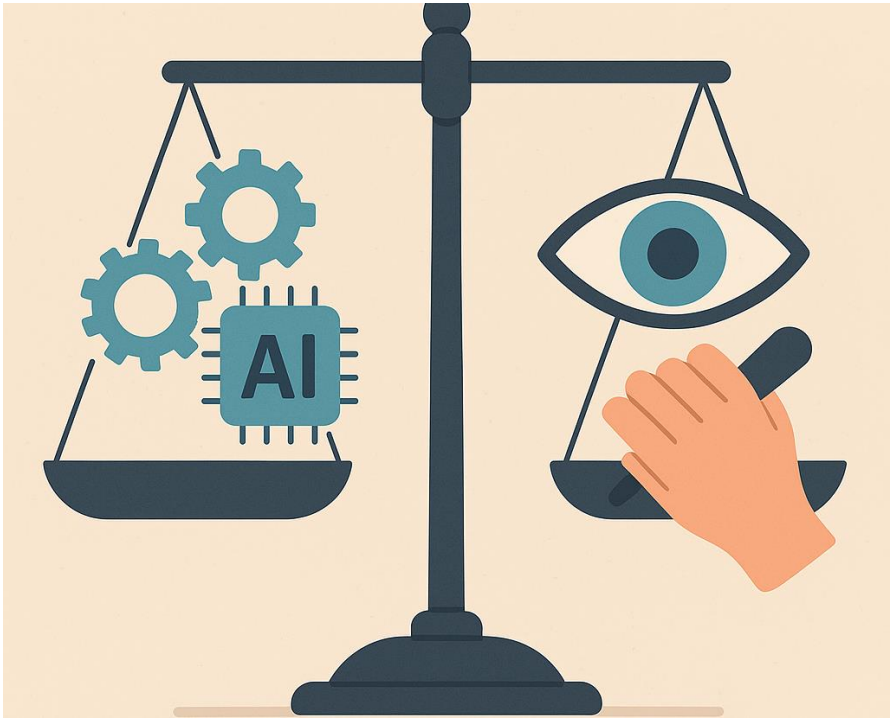
Fachveranstaltung Maschinensicherheit,
U. Zilz, 23.09.2025

Welche psychologisch relevanten Anforderungen stellen Maschinenverordnung und

KI-Verordnung?

Technik + Mensch = Sicherheit? oder

Technik - Mensch = Sicherheit?



Grafik BGHM und SORA

Ergonomischen Grundsätze zur Vermeidung psychischer Fehl-beanspruchung laut Maschinenverordnung (EU) 2023/1230 Anhang I

Vermeidung eines vorgegebenen **Arbeitsrhythmus**;

dauernde **Aufmerksamkeit** vermeiden;

Anpassung der **Schnittstelle zwischen Mensch und Maschine an die vorhersehbaren Eigenschaften der Bediener auch für Maschinen deren Verhalten oder Logik sich entwickelt**;

Anpassung von diesen Maschinen …. damit sie auf Personen in angemessener und **geeigneter Weise reagieren** und **geplante Handlungen** auf verständliche Weise **mitteilen**.

Was bedeutet „leicht absehbares menschliches Verhalten“?

„vernünftigerweise vorhersehbare Fehlanwendung“

bezeichnet die Verwendung von Maschinen oder dazugehörigen Produkten in einer laut Betriebsanleitung nicht beabsichtigten Weise, die sich jedoch aus leicht absehbarem menschlichem Verhalten ergeben kann.

MVO, ANHANG III

GRUNDLEGENDE SICHERHEITS- UND GESUNDHEITSSCHUTZ-ANFORDERUNGEN FÜR KONSTRUKTION UND BAU VON MASCHINEN ODER DAZUGEHÖRIGEN PRODUKTEN

Teil A

Technische Sicherheit kann menschliche Wachsamkeit verringern –
Risiko bleibt bestehen.

EU-Maschinenverordnung:
Sicherheitsfunktionen dürfen nicht zum Verlust der Aufmerksamkeit führen dürfen.

Quelle: Verordnung (EU) 2023/1230, Anhang III.

Grafik BGHM und SORA

Art. 14 KI-VO Menschliche Aufsicht

„.. sich einer möglichen Neigung zu einem automatischen oder übermäßigen Vertrauen in die von einem Hochrisiko-KI-System hervorgebrachte Ausgabe („Automatisierungsbias“) bewusst zu bleiben, insbesondere wenn Hochrisiko-KI-Systeme Informationen oder Empfehlungen ausgeben, auf deren Grundlage natürliche Personen Entscheidungen treffen;

Automatisierungsbias

- Die Tendenz, den Empfehlungen oder Entscheidungen eines automatisierten Systems übermäßig zu vertrauen.
- **Kognitive Entlastung:** Das System 'denkt' → Mensch spart Energie.
- **Selektive Aufmerksamkeit:** Mensch ignoriert Infos, die nicht zum System passen.
- **Reduzierte Verantwortung:** Gefühl 'Die Maschine entscheidet, nicht ich'.
- **Verhaltensadaptation:** Anpassung an Assistenzsystem, z. B. kürzere Abstände bei Fahrerassistenz.

Automatisierungsbias: Psychologischer Hintergrund:

Kognitive Sparsamkeit: Das menschliche Gehirn delegiert komplexe Aufgaben, um mentale Anstrengung zu sparen.

Bestätigungsbias: Die Ausgabe des Systems wird als Bestätigung für die eigene Entscheidung gesehen.

Error of ommission (Unterlassungsfehler): Der Bediener greift nicht ein, da das System nicht gewarnt hat

Error of commission (Begehungsfehler, Aktivitätsfehler oder Handlungsfehler) Der Bediener folgt einer falschen Anweisung des Systems.

Quelle: Mosier & Skitka, 1996

Verhaltensadaptation – Anpassung des Verhaltens an wahrgenommene Sicherheit

- bewusste oder unbewusste Anpassung von Handlungen, Routinen, Entscheidungsstrategien oder Interaktionsmustern, wenn sich die technologische Umgebung ändert.
- Kann neutral, positiv oder negativ sein.

Typen der Verhaltensadaptation

- **Strategische Anpassung:** Beschäftigte ändern Planungs- und Entscheidungsstrategien, um das Assistenzsystem optimal zu nutzen.
- **Kognitive Entlastung:** Menschen delegieren Teile ihrer Aufmerksamkeit, Erinnerung oder Problemlösung an das System.
- **Koordinative Neuausrichtung:** Neue Abläufe entstehen durch Mensch-Maschine-Synchronisation.
- **Feinanpassung durch Feedback:** Nutzer lernen über Feedback-Schleifen, wann das System zuverlässig ist, wann nicht – und adaptieren ihre Reaktionen situativ.

Lernpsychologische Mechanismen:

- **Habitualisierung:** Wiederholung führt zu neuen Routinen. Die Systemlogik wird Teil der Alltagskompetenz.
- **Situatives Lernen und erfahrungsbasiertes Lernen**: Was funktioniert in welcher Situation? Menschen entwickeln Heuristiken, wie sie das System nutzen.
- **Kognitive Rekonfiguration:** Menschen strukturieren mentale Modelle neu: „Wie funktioniert die Aufgabe jetzt?“
- **Vertrauenskalibrierung:** Anpassung des Vertrauens in das System an die Systemrealität, entscheidend, um Über- oder Untervertrauen zu vermeiden.

Abgrenzung der drei Vertrauensphasen *(nach Kraus)*

Phase	Rolle im Vertrauensprozess	Relevanz für Kalibrierung
1. Propensity to Trust	Disposition – allgemeine Vertrauensbereitschaft	Nicht kalibrierbar
2. Initial Learned Trust	Erste Einschätzung auf Basis von Oberfläche, Marke, Vorerfahrung	Nur begrenzt kalibrierbar
3. Dynamic Learned Trust	Erfahrungsbasiertes Vertrauen durch Nutzung und Rückmeldung	Kern der Kalibrierung

Einflussfaktoren Verhaltensfaktoren

Faktor	Einfluss
Systemtransparenz	Fördert aktives Lernen & Anpassung
Interaktivität	Höhere Interaktionsmöglichkeit → mehr Anpassungspotenzial
Systemzuverlässigkeit	Stabile Systeme fördern neue Routinen
Aufgabenkomplexität	Je komplexer, desto wahrscheinlicher ist strategische Anpassung
Training & Erfahrung	Fördert bewusste Adaption statt bloßer Reaktion

Positive Effekte gelungener Verhaltensadaption

Effekt	Begründung
Erhöhte Effizienz	Durch optimierte Nutzung des Systems
Weniger kognitive Last	Assistenz übernimmt Routineanteile
Bessere Mensch-Maschine-Symbiose	Höhere Interaktionsqualität
Schnelleres Fehlerlernen	Adaption über Rückmeldung
Entwicklung neuer Kompetenzen	z. B. algorithmisches Denken

Risiken nicht-gelungener Verhaltensadaption

Effekt	Begründung
Kognitive Dissoziation	System wird genutzt, ohne verstanden zu werden
Automatisierungsabbruch	System wird ignoriert oder abgeschaltet
Sicherheitslücken	Workarounds führen zu ungewollten Schwachstellen
Kompetenzverlust	Verlernen ehemals notwendiger Fähigkeiten

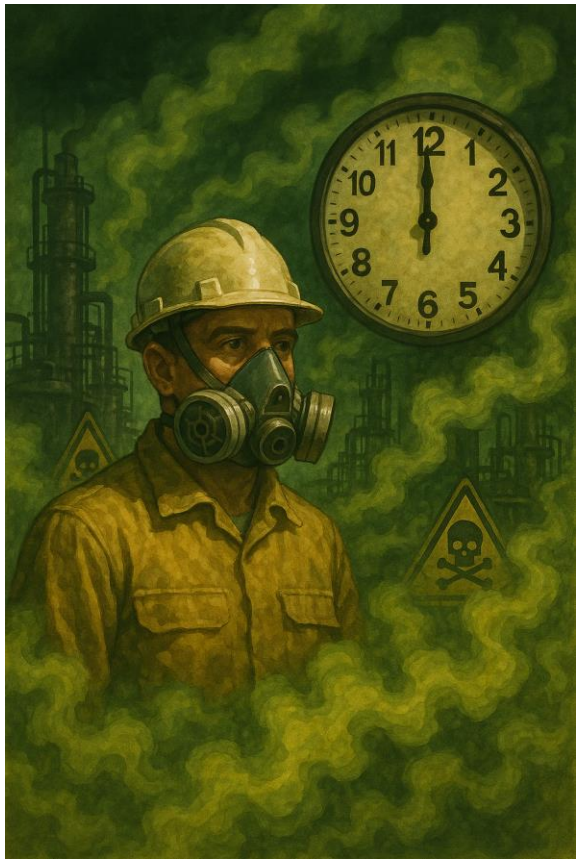
Risikokompensation (Wilde, 1994):

- **Regelkreismechanismus:** Menschliches Risikoverhalten als Thermostat. Wenn die wahrgenommene Gefahr (die "Raumtemperatur") zu niedrig ist, wird das Verhalten (die "Heizung") angepasst, um das gewünschte Risikoniveau (die "Solltemperatur") zu erreichen.
- **Wahrgenommenes vs. tatsächliches Risiko:** Die Theorie unterscheidet zwischen dem tatsächlichen, objektiven Risiko und dem subjektiv wahrgenommenen Risiko. Es ist das **wahrgenommene Risiko**, das das Verhalten steuert.
- **Verhalten als Ausgleich:** Die Gesamtsicherheit steigt nur dann, wenn die Motivation des Menschen, ein geringeres Risiko einzugehen, erhöht wird.

Verhaltensadaptation: Atemschutz & Belüftungssysteme

Arbeiter bleiben länger in belasteten Bereichen, wenn technische Maßnahmen vorhanden sind.

Trotz bekannter Restgefahr durch Chemikalien oder Staubpartikel.



Grafik BGHM und SORA

Quelle:Wilde, G. J. S. (1998). Risk homeostasis theory: an overview. Injury Prevention, 4(2), 89–91.

Verhaltensadaptation: Gabelstapler mit Sicherheitssystemen

Fahrer verlassen sich zu stark auf Personen Erkennung oder Kollisionssensoren.

Folge: schnelleres Fahren, engere Kurven, weniger Aufmerksamkeit.

Grafik BGHM und SORA

Quelle: Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. Theoretical Issues in Ergonomics Science, 1(4), 315–331.

Verhaltensadaptation: Arbeiten an Pressen

Schnellere Handbewegungen:

Die Bediener bewegten ihre Hände und Arme schneller in und dem Gefahrenbereich der Maschine. Sie zögerten weniger.

Reduzierte Taktzeit:

Die Zeit für einen vollständigen Arbeitszyklus – vom Einlegen des Materials bis zum Entnehmen des fertigen Teils – wurde deutlich verkürzt.

Weniger Blickkontrolle:

Sie verbrachten weniger Zeit damit, den Gefahrenbereich visuell zu überprüfen, bevor sie einen Arbeitsgang starteten..

Grafik BGHM und SORA

Quelle:Holmstrom, J.-O. R. S. H. (2014). Behavioral Adaptation to Machine Safety Systems: An Exploratory Study of Press Operators. Journal of Ergonomics, 4(134), 1-8.

Verhaltensadaptation und Automatisierung

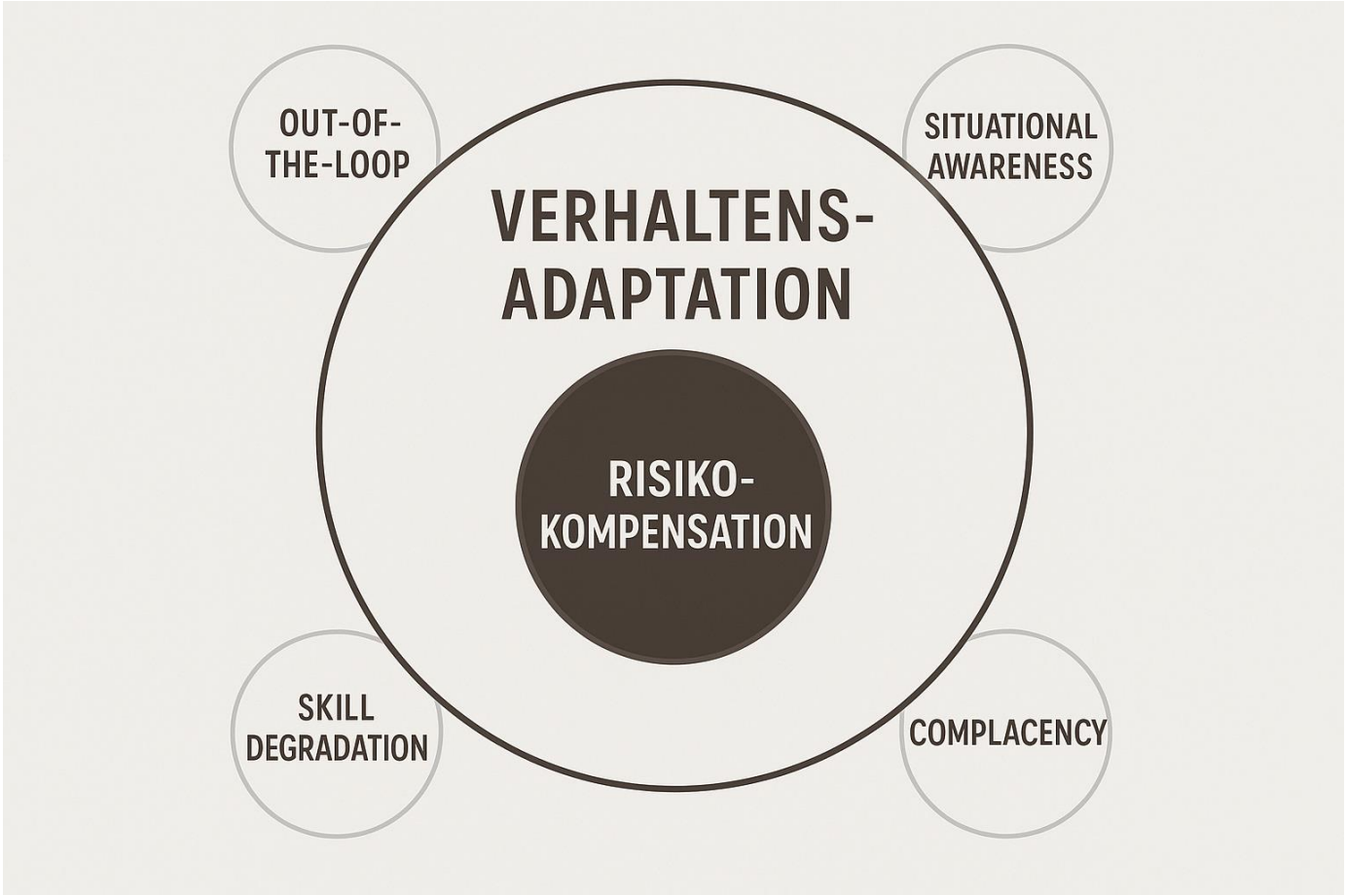
Eine Erhöhung des Automatisierungsgrades führt zu einer signifikanten **Risikokompensation.**

Höhere Automatisierung führt zu einer gleichbleibenden Anzahl von Unfällen, aber mit **neuen Unfallursachen.**

Systeme, die nicht nur warnen, sondern auch aktiv sicherheitsrelevante Steuerungsaktionen durchführen, verursachen eine **stärkere Risikokompensation** als solche, die nur Warnungen geben.

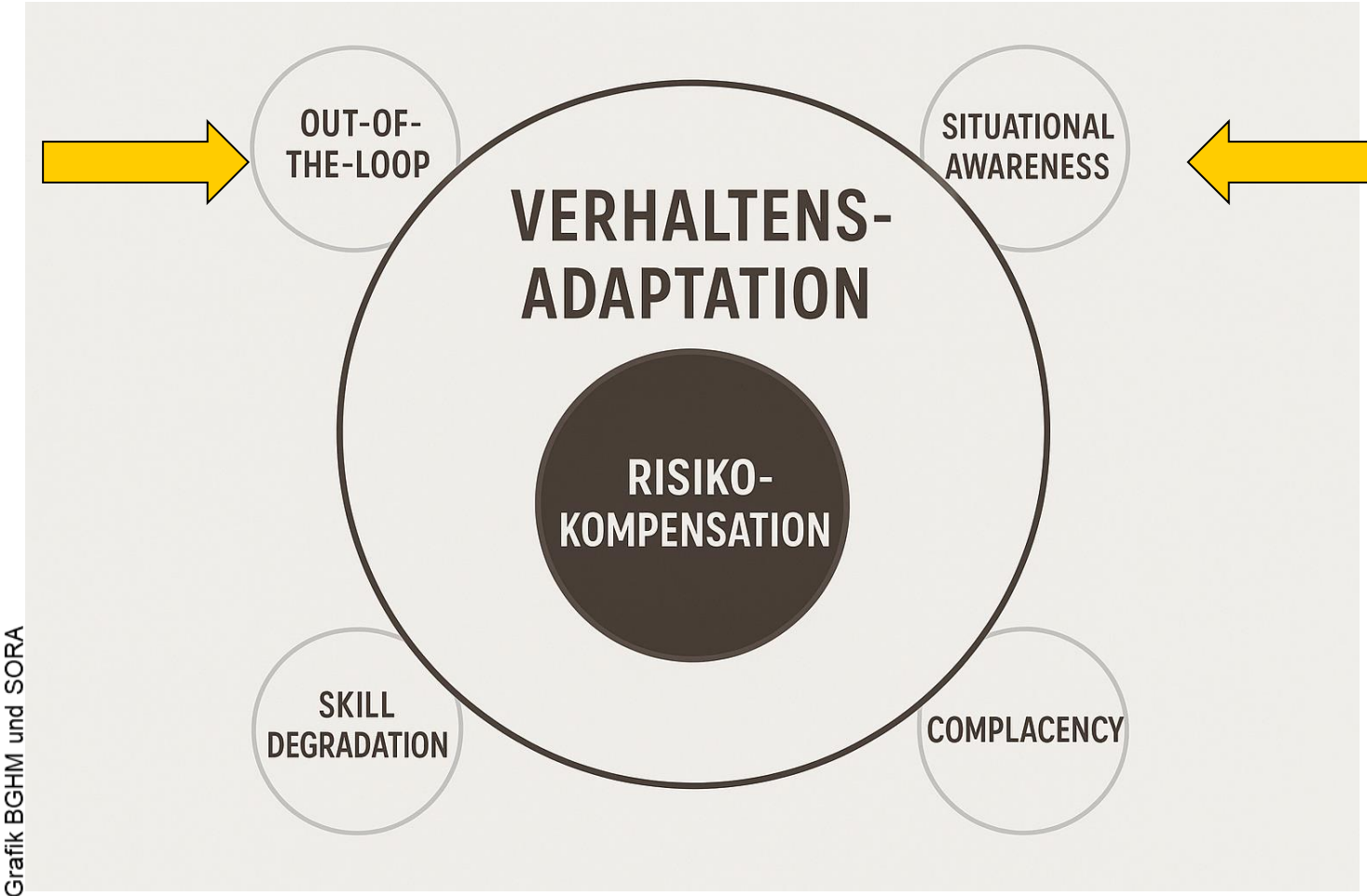
Quelle:Risk Compensation due to Human Adaptation to Automation for System Safety (2007) Makoto ITOH, Daisuke SAKAMI, Kenji TANAKA, Trans.of the Society of Instrument and Control Engineers Vol.43, No.10,926/934

Probleme bei der Verhaltensadaptation im Kontext



Grafik BGHM und SORA

Probleme bei der Verhaltensadaptation im Kontext



Grafik BGHM und SORA

Human out of the Loop

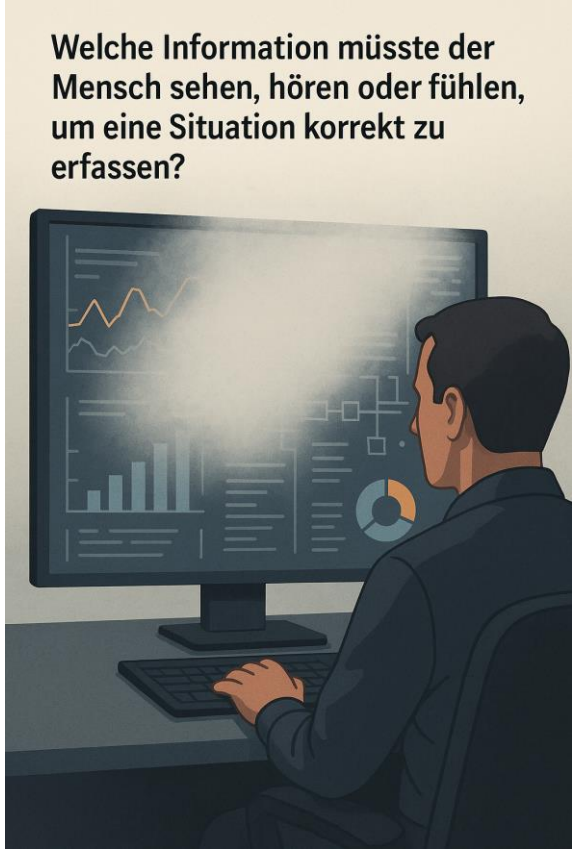
Human out of the Loop (HOOTL) beschreibt den Zustand, in dem der Mensch nicht mehr aktiv in Steuerungs- oder Entscheidungsprozesse eingebunden ist.
Stattdessen wird er zu einem **passiven Beobachter**. Dies kann dazu führen, dass wichtige Systemzustände nicht mehr rechtzeitig erkannt oder verstanden werden.

Grafik BGHM und SORA

Situational Awareness

Fähigkeit, relevante **Informationen** in der Umgebung **wahrzunehmen**, sie zu **verstehen** und zukünftige **Entwicklungen abzuleiten**.

Fehlt sie, kann der Mensch die Bedeutung technischer Rückmeldungen nicht mehr korrekt einordnen – was insbesondere in kritischen Situationen zu **gefährlichen Fehleinschätzungen** führt.



Grafik BGHM und SORA

Human out of the Loop (HOOTL) – Sicherheitsrisiken

Problem	Auswirkung
Verlust von Kontrolle	Im Notfall fehlen Handlungskompetenz und Reaktionsgeschwindigkeit
Entkopplung vom Systemverhalten	Entscheidungen des Systems werden nicht nachvollzogen
De-Skill-Effekt	Nutzer verlernen die Aufgabe selbstständig durchzuführen
Wahrnehmungsverlust	Durch Passivität sinkt Aufmerksamkeit und Problembewusstsein

Lack of Situational Awareness – Sicherheitsrisiken

Problem	Auswirkung
Sensorische Entkopplung	Mensch bekommt weniger direkte Reize (z. B. durch haptische oder akustische Signale)
Kognitive Auslagerung	Assistenzsysteme übernehmen Wahrnehmung → Nutzer erkennt kritische Signale nicht
Überforderung in Störungen	Bei plötzlich nötigem Eingreifen fehlen mentale Modelle und Reaktionsfähigkeit

Mensch im Regelkreis (Kapitel 3, Abschnitt 2, Artikel 14)

KI-Verordnung verlangt **„Human in the Loop“**. Das bedeutet, dass der Mensch:

- Das System aktiv überwacht und steuert.
- Entscheidungen trifft, die für den Fortgang des Prozesses erforderlich sind.
- Jederzeit in der Lage ist, die Kontrolle zu übernehmen oder das System zu korrigieren.

Mensch im Regelkreis (Kapitel 3, Abschnitt 2, Artikel 14)

…dass diese Systeme so entwickelt werden müssen, dass eine **„effektive menschliche Überwachung“** möglich ist.

Die beaufsichtigende Person muss die **„notwendige Kompetenz und Ausbildung“** haben, um die Funktionsweise des Systems zu verstehen.

Sie muss in der Lage sein, die „Entscheidungen des Systems zu interpretieren“, zu „korrigieren“ und es im Notfall sogar **„manuell zu übersteuern oder zu stoppen“**.

Designhinweise

Human-in-the-Loop-Design	Mensch bleibt mitverantwortlich – durch aktive Eingaben, Bestätigungen, Rückmeldeschleifen
Adaptive Automatisierung	System passt Unterstützungsgrad an den Nutzerzustand an
Multimodales Feedback	Visuelle, akustische und haptische Rückmeldungen erhöhen SA
Trainings & Simulation	Aufbau von mentalen Modellen durch regelmäßige Interaktion mit dem System
Explainable Systems	Assistenzsysteme erklären ihre Entscheidungen oder zeigen Unsicherheiten an