



Politische Empfehlungen

KI-Sicherheitsoffensive starten

Die jüngsten Vorfälle und Warnungen rund um hochleistungsfähige KI haben die Debatte über KI-Sicherheit weltweit mit neuer Dringlichkeit versehen. Der OpenAI-Hugging-Face-Vorfall hat gezeigt, dass fortgeschrittene KI-Modelle in realen technischen Umgebungen unerwartete und sicherheitsrelevante Handlungen ausführen können. Inzwischen diskutieren auch führende Frontier-Labs öffentlich über einen dauerhaften Zugang unabhängiger externer Evaluatoren und eine

stärkere externe Kontrolle. Diese neue Aufmerksamkeit für KI-Risiken ist richtig und wichtig – sie darf sich aber nicht auf schwer quantifizierbare Extremszenarien verengen.

Die Politik muss zugleich die Sicherheit im betrieblichen Alltag stärken. Denn KI-Sicherheit ist bereits heute betriebliche Realität, zeigt unsere gemeinsam mit dem BSI erstellte, KI-Sicherheitsstudie 2026: 17 Prozent der befragten Unternehmen berichten für die vergangenen

zwölf Monate von bestätigten oder vermuteten KI-gestützten Cyberangriffen beziehungsweise Betrugsversuchen. Zugleich fehlen bei 25 Prozent der KI-nutzenden oder -planenden Unternehmen Regelungen für solche Vorfälle. Um sichere KI-Nutzung zu erleichtern und Orientierung zu schaffen, sollte die Politik Unternehmen unterstützen, KI nutzbringend, sicher und rechtskonform einzusetzen.

Der TÜV-Verband sieht vor allem in vier Bereichen Handlungsbedarf:

1. Frontier-KI: Internationales Prüf- regime etablieren und AISI Deutschland einbinden

Die Bundesregierung und die EU sollten sich verstärkt für ein unabhängiges internationales Prüfregime für besonders leistungsfähige KI-Modelle (Frontier-KI) einsetzen. Die Sicherheitsbewertung darf nicht mehr allein bei KI-Anbietern liegen. Es braucht unabhängige Evaluationen vor der Markteinführung und bei wesentlichen Weiterentwicklungen. Erforderlich sind gemeinsame risikobasierte Prüfkriterien, ausreichender Zugang zu führenden Modellen sowie sicherheitsrelevante Informationen und belastbare Verfahren für KI-Vorfälle. Prüfacteure brauchen fachliche und finanzielle Unabhängigkeit; Hersteller dürfen kritische Befunde nicht unterdrücken. Prüfergebnisse müssen konkrete Konsequenzen für Risikominderung und Freigabe haben.

Das neu entstehende AI Safety and Security Institute (AISI) Deutschland sollte bei der unabhängigen Evaluation fortgeschrittener KI international (und auf Augenhöhe im International Network of AI Safety Institutes vernetzt) eine wichtige Rolle übernehmen. Zentral ist, die bereits bestehende und bewährte Qualitätsinfrastruktur aktiv in diese Prozesse einzubeziehen, um fundierte Prüf- und Bewertungsmaßstäbe global um- und durchzusetzen. Der Aufbau auf den Kompetenzen von BSI und Bundesnetzagentur sollte mit ausreichenden Ressourcen und einer strukturierten Zusammenarbeit mit unabhängigen Prüforganisationen und Forschung verbunden werden.

2. EU AI Act: Unabhängige Prüfungen stärken und hohes Schutz- niveau bewahren

Unabhängige Prüfungen müssen im AI Act gestärkt und ein hohes Schutzniveau erhalten werden. Vereinfachungen sollten Verfahren erleichtern, ohne wirksame Schutzanforderungen abzubauen. Um regulatorische Fragmentierung zu vermeiden, müssen insbesondere für sensible Bereiche wie Medizinprodukte ein harmonisierter Rechtsrahmen gewahrt bleiben und widersprüchliche Anforderungen vermieden werden.

Auch für neue System-Risiken agentischer KI-Systeme, die selbstständig Aufgaben planen oder Tools nutzen, müssen risikoadäquate Anforderungen der KI-Regulierung gelten und zukunfts fest durchgesetzt werden – etwa bei weitreichenden Zugriffsrechten oder schwer steuerbaren Risiken im Zusammenspiel mehrerer KI-Agenten. Klare Verantwortlichkeiten, wirksame menschliche Eingriffsmöglichkeiten und Tests des Gesamtsystems können und müssen diese Risiken adressieren. Dazu sollte die Politik harmonisierte Standards, praxistaugliche Prüfmethode und zugängliche Test- und Prüfangebote voranbringen. Ziel sind nachvollziehbare Nachweise, die Unternehmen bei der Beurteilung von Sicherheit, der Compliance und Anbieterauswahl unterstützen. Dabei gilt: Prüfungsumfang und Aussagekraft von KI-Zertifizierungen sollten zum Risiko und vorgesehenen Einsatz passen.

3. Umsetzung: Praxisnahe AI Act Guidance und KMU-Ansprechpartner stärken

Unternehmen benötigen klare Zuständigkeiten, abgestimmte Auslegungshilfen und verlässliche zentrale Ansprechpartner für Compliance-Fragen. Der KI-Service-Desk der Bundesnetzagentur sollte als leicht zugängliche Anlaufstelle weiter ausgebaut und mit zuständigen Fachbehörden vernetzt werden. Digitale Checklisten und branchenspezifische Leitfäden sollten Anforderungen für verschiedene Unternehmensrollen und Anwendungsfälle verständlich aufbereiten.

Gerade KMU brauchen konkrete Hilfen für Risikoeinstufung, Dokumentation, Beschaffung und den sicheren Betrieb. Praxisnahe Muster für betriebliche KI-Regeln, Verantwortlichkeiten und den Umgang mit Sicherheitsvorfällen sollten bestehende Angebote ergänzen. Die Vorbereitungszeit bis zum Greifen der AI-Act-Regeln für Hochrisiko-KI ist für die intensiviertere Fortentwicklung von KI-Standards, Prüfkapazitäten und Umsetzungshilfen zu nutzen.

4. Digitale Souveränität: Europäische Alternativen und KI-Kompetenz aus- bauen

Digitale Souveränität entsteht durch echte Wahlmöglichkeiten und KI-Kompetenz. Europäische KI- und Cloud-Angebote sollten in ihrer Leistungsfähigkeit gestärkt und für Unternehmen besser vergleichbar und verwendbar werden. Der vorgeschlagene Cloud and AI Development Act (CADA) der EU-Kommission ist ein Schritt in die richtige Richtung, um „Souveränitäts-Washing“ vorzubeugen und Marketing-Aussagen durch unabhängige Assurance auf ihren tatsächlichen Grad an digitaler Souveränität zu prüfen. Zudem sollte auch nationales Engagement den initialen Integrations- und Wechselaufwand adressieren, und offene Schnittstellen, Datenportabilität sowie interoperable Standards sind wichtige Ansatzpunkte. Ziel digitaler Souveränität ist eine echte, selbstbestimmte Anbieterwahl auf Grundlage von Sicherheit, Qualität und betrieblicher Eignung.

Flankierend müssen Bund, Länder, Wirtschaft und Bildungsanbieter anwendungsnahe KI-spezifische Schulungs- und Beratungsangebote – insbesondere auch für den Mittelstand (KMU) – ausbauen. Diese sollten unterschiedliche Erfahrungsstände berücksichtigen und Anwendungs-, Daten- und Sicherheitskompetenz miteinander verbinden. Branchenspezifische Best-Practice-Beispiele und Erprobungsmöglichkeiten können Unternehmen helfen, Nutzen und Grenzen geeigneter Anwendungen (inklusive agentischer KI-Möglichkeiten) besser zu beurteilen und die KI-Kompetenz in der Belegschaft zu festigen. Weiterbildung muss Führungskräfte und Beschäftigte kontinuierlich und wirksam erreichen und mit der technischen Entwicklung Schritt halten.

Fazit: KI-Sicherheit offensiver vorantreiben

KI ist in der Wirtschaft angekommen – die Sicherheitsstrukturen müssen jetzt nachziehen. Sichere KI braucht Eigenverantwortung, wirksame Regulierung und unabhängige Kontrolle. Die zentrale Erkenntnis unserer Studie ist deshalb: Wir müssen KI-Sicherheit auf zwei miteinander verbundenen Ebenen denken und konsequent vorantreiben. Im konkreten Betriebseinsatz geht es darum, Qualität, Cybersicherheit, Verantwortlichkeiten und Compliance von KI-Systemen beherrschbar zu machen und ihre Sicherheit dort, wo es erforderlich ist, unabhängig zu überprüfen. An der technologischen Frontier geht es zugleich darum, die Fähigkeiten und Risiken der weltweit leistungsfähigsten Modelle fortlaufend technisch zu evaluieren.

Mit dem AISI Deutschland entsteht dafür eine neue, notwendige staatliche Struktur: BSI und Bundesnetzagentur bilden zunächst gemeinsam den virtuellen Nukleus, wobei das BSI insbesondere seine Cybersecurity-Kompetenz einbringt. Das Institut muss sich aktiv in die internationale Zusammenarbeit der KI-Sicherheitsinstitute einbringen. Um hier schlagkräftig zu sein, muss der unabhängige technische und prüferische Sachverstand der bewährten deutschen Qualitätsinfrastruktur von Beginn an eingebunden werden. Mit dem TÜV AI.Lab sowie den jüngst vorgestellten und derzeit pilotierten KI-Zertifizierungen bauen die TÜV-Organisationen bereits heute konkrete, praxistaugliche Prüfmethoden für Qualität und Sicherheit von KI-Systemen auf.

Diese Prüfkompetenz gewinnt mit Blick auf den AI Act massiv an Dringlichkeit: Die Hochrisiko-Regeln für Anwendungsbereiche nach Annex III greifen ab Dezember 2027, für Hochrisiko-KI in regulierten Produkten ab August 2028. Die aktuelle globale Debatte zeigt damit vor allem eines: Frontier Safety und sichere KI-Nutzung im betrieblichen Alltag sind keine Gegensätze – beides erfordert technische Kompetenz, belastbare Evaluation und vertrauensbildende, unabhängige Prüfung.

Kontakt



TÜV-Verband e. V.
Friedrichstraße 136
10177 Berlin

www.tuev-verband.de

Dr. Patrick Gilroy
Referent KI & Bildung
+49 30 760095-360
+49 1511 20792262
patrick.gilroy@tuev-verband.de