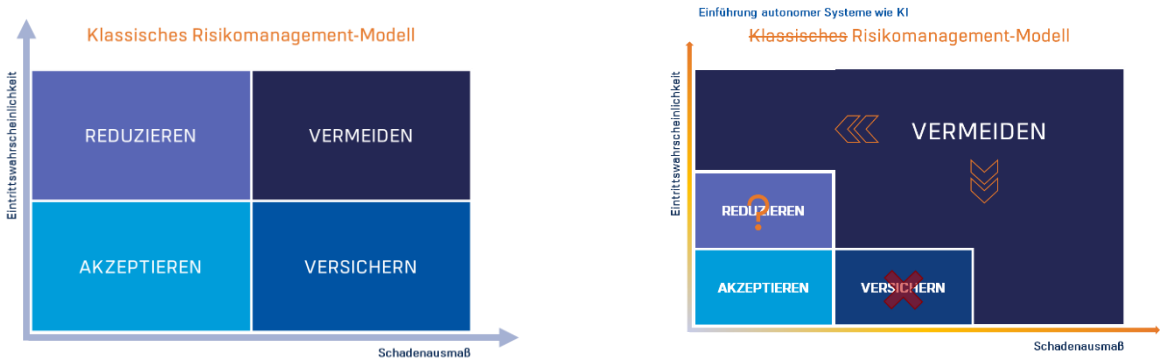
Risikostufe kritisch? Der sichere Weg ins Unbekannte

Alexandra Thompson

#VISIONConsulting #OpenSourceDenkerin #BeratungimFokus #VISIONStimmen #150Worte

Die Verbreitung von KI-Systemen fordert einen Paradigmenwechsel in unserer Risikomanagement-Logik. Bis jetzt galt die bekannte zweidimensionale Risikomatrix mit einer Achse für das Schadensausmaß und einer Achse für die Eintrittswahrscheinlichkeit. Dieses Modell ermöglichte es uns, Risiken intuitiv einzuordnen, zu bewerten und zu behandeln.



Mit der binär-linearen Struktur können diskrete Stufen festgelegt werden, die es einer Organisation ermöglichen, angemessene Handlungen vorzunehmen:

- Selten eintretende Risiken mit geringem Schadensausmaß zu akzeptieren,
- oft eintretende Risiken mit geringem Schadensausmaß zu untersuchen und deren Häufigkeit durch operative Maßnahmen zu reduzieren,
- selten eintretende Risiken mit großem Schadensausmaß zu verlagern oder gegen eine Risikoprämie zu versichern und
- oft eintretende Risiken mit gravierenden Konsequenzen zu vermeiden bzw. auf die riskante Tätigkeit zu verzichten.

Mit der Einführung autonomer Systeme wie Künstlicher Intelligenz wird dieses einfache Risikomodell hinsichtlich mehrerer Dimensionen obsolet. Sowohl die Eintrittswahrscheinlichkeit als auch die potenziellen Konsequenzen bei der Nutzung einer stärkeren KI steigen nun beschleunigt. Dies reduziert die Reaktionszeiten und verhindert eine klare Einstufung von Risikoklassen.

Die Dynamik der neuen Risiken

Mit dem neuen, beweglichen Ziel ändern sich auch die Handlungsmöglichkeiten. KI-Risiken zu versichern, kann entweder zu teuer oder gar nicht mehr berechenbar werden, sodass potenziell-katastrophale Risiken mit geschätzt geringer Wahrscheinlichkeit nun internalisiert werden müssten. Die sichere Zone akzeptabler Risiken wird durch die beschleunigten Technologieentwicklungen und die Erweiterung der betroffenen Nutzerkreise immer kleiner und diffuser.

Zudem stellt sich die Frage, ob die Vielzahl kleinerer Risiken noch durch vorsorgliche Maßnahmen aufgehalten werden kann, wenn starke Technologien flächendeckend und wirtschaftlich zugänglich

sein werden. Wir erleben bereits die Konsequenzen dieser steigenden Unsicherheit: Oft fühlen wir uns gezwungen, Vermeidungsstrategien breiter einzusetzen – zumeist mit der Folge eines Fortschrittsstopps.

Perspektivwechsel und Handlungsempfehlungen

Ist dieses neue Risikomodell angemessen in der KI-getriebenen Welt? Strenge Verbote erinnern mich zum Beispiel an die Prohibitionszeiten der 20-er Jahre in den USA. Nie zuvor und nie nachher in der Geschichte hat man so viel „Gegengeschäft" gesehen. Es wird heute auch nicht anders sein, denn die massiven Investitionen in KI-Hersteller an der Börse zeigen deutlich auf Erwartungen zu unschlagbaren künftigen Gewinnen. Wie können wir das schwarz-weiße Bild mit einem flexibleren Denkmodell ersetzen? Bessere Kästchen in einer nicht-diskreten Realität können nicht mehr gefunden werden. Wir brauchen einen kompletten Paradigmenwechsel.

Erste Schritte für den neuen Umgang mit Risiken

Der erste Schritt: Wir sollten diese Technologien besser verstehen, um Risiken realistisch einschätzen zu können. Da wir nun mit einem Kontinuum statt mit diskreten Stufen arbeiten, ist kontinuierliches Lernen darüber, was KI-Systeme können, unabdingbar, um diffuse Zonen entsprechend ihrer Risikobehaftung zu behandeln. Das heißt, absolute Verbote sollten auf ein Minimum reduziert werden, es sei denn, wir wollen riskieren, dass gefährliche KIs im „Deep Web" entwickelt werden.

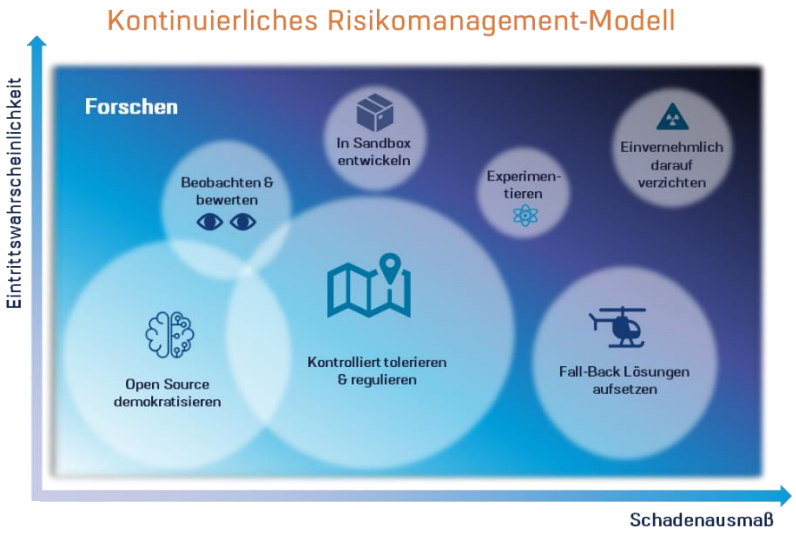
Der nächste Schritt: Wir sollten uns mehr und kreativere Risikohandlungen ausdenken. Ein Einfaches „Weg vom Tisch" ist eine utopische Betrachtung. Wir werden uns mit einer Vielzahl von Situationen, Berechnungen, Folgen, Kontrollen, Gegenmaßnahmen und mutigen Annahmen auseinandersetzen müssen. Je mehr Karten wir in der Hand haben, desto höher wird die Chance sein, die richtigen Asse für eine vernünftige Steuerung der KI zu ziehen.

Kontinuierliche Anpassung an sich wandelnde Strukturen

Mit jedem Schritt werden wir uns von festen Strukturen schneller verabschieden müssen. Denkmodelle sind ein hervorragendes Werkzeug, um die überwältigende Komplexität der Welt zu bewältigen, sollten aber nicht allzu lange unsere Handlungen bestimmen. Sobald wir feststellen, dass eine Denkweise nicht mehr zeitgemäß ist, sollten wir uns den Luxus erlauben, nachzudenken und Pläne zu überarbeiten.

Ein alternatives Risikomanagement-Modell

Zum Schluss möchte ich ein alternatives Risikomanagement-Modell in Zeiten der KI vorschlagen.



Wir behalten die zwei Dimensionen der Risiken – Eintrittswahrscheinlichkeit und Schadenausmaß – bei, erkennen jedoch den diffusen, nicht-linearen Verlauf der Möglichkeiten an. Statt fester Kästchen sollten wir flexiblere Blasen mit weniger klaren Konturen verwenden, die sich abgrenzen, angrenzen oder überlappen können. Damit schaffen wir einen offenen Hintergrund für das Unbekannte. Dort, wo Risiken durch Erfahrungswerte konkretisiert werden können, stehen sieben Handlungscluster zur Verfügung. Diese können individuell oder kombiniert eingesetzt werden, um die Auswirkungen einer riskanten Aktivität auf eine oder beide Risikoachsen zu beeinflussen.

1. **Demokratisierung durch Open Source**

Der Einsatz von Open-Source-Intelligenz ermöglicht es uns, kollektives Wissen zu nutzen. Genauso wie Linux heute ein sicheres Betriebssystem ist, da es dezentral und fortlaufend verbessert wird, könnten KI-Systeme mit vorhersehbaren Risikomustern durch die öffentliche Kooperation sicherer gemacht werden. Solche Abwehrmechanismen könnten sich so schnell und kostengünstig entwickeln und über die freie Nutzung und Weiterverwendung in der Wirtschaft einen Multiplikatoreffekt erzielen.

2. **Kontrollierte Toleranz und Regulierung**

Die Mehrheit der Technologien wird durch kontrollierte Toleranz und gezielte Regulierung besser nutzbar. Wenn wir einen technologischen Vorsprung erleben möchten, müssen wir unsere Toleranz für das Eintreten bestimmter Risiken erhöhen. Als Gegenmaßnahme können wir durch flexible und präzise Regulierung das Schadenausmaß reduzieren, Konsequenzen rückgängig machen, Einsatzszenarien einschränken und weitere Steuerungsmodelle entwickeln. Ziel dieser Strategie ist es, die gesellschaftliche Akzeptanz und Verträglichkeit von Massen-KI-Systemen zu verbessern.

3. **Einsatz von Fall-Back-Lösungen**

Fall-Back-Lösungen ermöglichen uns, Grenzen zu überschreiten und gleichzeitig auf schnelle Reaktionen sowie gut eingeübte Rettungskräfte zurückzugreifen. So können Systeme wiederhergestellt und Redundanzen genutzt werden, oder es wird auf manuelle Prozesse zurückgegriffen. Ein Beispiel ist der Kilimandscharo, einer der gefährlichsten Berge: Menschen schrecken nicht davor zurück, ihn zu besteigen, sondern setzen auf die besten Rettungsmaßnahmen.

4. **Beobachten und Bewerten**

Das Prinzip des Beobachtens und Bewertens wird bereits in der Medizin angewendet, etwa bei Krebspatienten, die zunächst unter Beobachtung stehen, um unnötige Eingriffe zu vermeiden. Dieses Konzept kann auch in der IT-Welt Anwendung finden, wenn Maßnahmen möglicherweise größere negative Auswirkungen hätten als ein kontrolliertes Abwarten.

5. **Entwicklung in Sandbox-Umgebungen**

Sandbox-Umgebungen bieten die Möglichkeit, Lernprozesse zu fördern, ohne Risiken in vollem Umfang auszusetzen. Ob in Flugsimulatoren oder gekapselten Browsern – wir nutzen Sandboxen überall dort, wo wichtige Erkenntnisse benötigt werden, die Reife für reale Einsätze jedoch noch nicht gegeben ist.

6. **Experimentieren in kontrollierten Umgebungen**

Durch Experimente in begrenzten, kontrollierten Umgebungen lassen sich ernsthafte Risikoszenarien erkunden. Solche Experimente unterstützen die Vorbereitung auf Katastrophen, etwa im Umgang mit gefährlichen Viren oder Substanzen, die unter höchsten Sicherheitsmaßnahmen aufbewahrt werden. Diese Vorgehensweise hilft dabei, Abwehrmechanismen kontinuierlich weiterzuentwickeln.

7. **Einvernehmlicher Verzicht auf Technologien**

In Situationen, in denen das potenzielle Schadensausmaß so groß wäre, dass kein Nutzen überwiegt, sollten wir auf die Nutzung bestimmter Technologien verzichten. Hier sind ethische Überlegungen entscheidend, etwa der Verzicht auf Systeme, die unkontrollierbare Reaktionsketten auslösen könnten, oder auf eine Übergabe menschlicher Entscheidungsfreiheit an Maschinen.

Diese sieben Strategien sind nur ein Teil der Möglichkeiten, die uns zur Verfügung stehen. Mit Kreativität und positiver Neugier können wir weitere Wege entdecken, die uns die unbekannte KI-Welt eröffnet. Indem wir Ängste überwinden, können wir uns auf Risiken vorbereiten, sie bewerten und in verschiedenen Formen darauf reagieren. Technologie ist kein „Alles-oder-Nichts“-Spiel. Zulassen und Verbieten sind rigide Denkmuster, die in Zukunft nur noch begrenzt hilfreich sein werden. Stattdessen werden wir in der Lage und gefordert sein, nuancierter zu reagieren und zu handeln. Ein Konsens wird selten erzielt werden, aber vielfältige Alternativen und Kompromisse können uns dabei helfen, die treibende Kraft menschlicher Entwicklung sicher weiterzuführen.

Schlussgedanken:

Risikostufe kritisch? Dann heißt es, Perspektiven wechseln und „Out-of-the-box“ denken. Nur so können wir die komplexe Dynamik einer KI-getriebenen Welt angemessen navigieren und den Weg für zukünftige Innovationen bereiten.