# KI in militärischen Frühwarn- und Entscheidungssystemen

Ingo J. Timm, Jörg Siekmann, Karl Hans Bläsius

https://www.uni-trier.de/index.php?id=35358 , http://siekmann.dfki.de/de/home/ , https://www.hochschule-trier.de/informatik/blaesius/

Trier, Saarbrücken, 11.6.2020, www.fwes.info/fwes-ki-20-1.pdf

Siehe auch www.akav.de

## Zusammenfassung

Die Sicherung der atomaren Zweitschlagfähigkeit ist die Grundlage der Abschreckungsstrategie, die bis heute jeden potentiellen Angreifer abgehalten hat, einen atomaren Angriff zu starten: „Wer als erster schießt, stirbt als zweiter." Um auch bei einer Gefährdung der Zweitschlagfähigkeit reagieren zu können, haben die Atommächte computergestützte Frühwarn- und Entscheidungssysteme entwickelt und installiert, mit dem Ziel einen Angriff rechtzeitig zu erkennen, um die eigenen atomaren Trägerraketen vor dem vernichtenden Einschlag aktivieren zu können. Eine solche Strategie wird als „launch-on-warning"-Strategie bezeichnet. Obwohl die Zeitspanne für Entscheidungen bei einer Angriffsmeldung in den letzten Jahren auf wenige Minuten gesunken ist, bleibt jedoch bisher die letzte Entscheidung – nicht zuletzt wegen der Fehleranfälligkeit solcher Systeme – Menschen überlassen.

Das Ende des INF-Vertrages (INF steht für Intermediate Range Nuclear Forces) hat zu einem neuen Wettrüsten auch mit Hyperschallraketen geführt, die diese Zeitspanne nun noch weiter verkürzen. Für eine Analyse und Bewertung dieser Alarmmeldungen bleibt für Menschen daher so wenig Zeit, dass hierfür seit neuerem verstärkt Systeme der Künstlichen Intelligenz (KI) eingesetzt werden sollen. Aber auch KI-Systeme können bei solchen Anwendungen keine sicheren Ergebnisse liefern, denn die zugrundeliegenden Daten sind *unsicher*, *vage* und *unvollständig*. Automatische Erkennungsergebnisse gelten deshalb nur mit einer gewissen Wahrscheinlichkeit und können falsch sein.

Aufgrund der unsicheren Datengrundlage stützen Menschen ihre Entscheidungen auch auf Kontextwissen über die politische Lage und die Einschätzung des "Gegners", die durch das potentielle Ende des "Open Skies"-Vertrags zusätzlich erschwert wird. Beispielsweise hat in einem Alarmfall die Bedienmannschaft des amerikanischen Frühwarnsystems entschieden, dass es sich um einen Fehlalarm handeln muss, da zu dieser Zeit der sowjetische Staatschef auf Staatsbesuch in Washington war. Ein ernster Fehlalarm im sowjetischen Frühwarnsystem geschah 1983, nur ein mutiger Eingriff des Kommandeurs Stanislav Petrow hat die atomare Katastrophe verhindert.

Auch bei maschinellen Entscheidungen muss zur Bewertung von Alarmmeldungen Kontextwissen zur weltpolitischen Lage einbezogen werden und dieses Wissen ist ebenfalls unsicher, vage und unvollständig. Das Ergebnis der Analyse durch ein KI-System ist daher immer nur im Rahmen einer statistischen Wahrscheinlichkeit korrekt. Die Gefahr eines Atomkrieges aus Versehen kann deshalb nicht durch immer stärkeren Einsatz von Methoden der Künstlichen Intelligenz in Frühwarnsystemen reduziert werden. Aufgrund der unsicheren und unvollständigen Datengrundlage können weder Menschen noch Maschinen eingehende Alarmmeldungen in so kurzer Zeit zuverlässig bewerten.

# 1 Militärische Frühwarn- und Entscheidungssysteme

Frühwarnsysteme dienen der Erkennung eines Angriffs durch Atomraketen auf der Basis von Sensordaten. Eine möglichst frühe Erkennung eines gegnerischen Angriffs soll Gegenmaßnahmen vor einem vernichtenden Einschlag ermöglichen. Bei einer Alarmmeldung stehen in der Regel nur wenige Minuten zur Verfügung, um diese zu überprüfen und die Situation zu bewerten. Dabei hängt die Situationsbewertung auch von der weltpolitischen Lage ab. Zum Beispiel kann es in einer Krise mit gegenseitigen Drohungen und dem zufälligen Zusammentreffen mit weiteren Ereignissen (z.B. Cyberangriffen) zu einer falschen Bewertung kommen, das heißt, die Meldungen könnten als gegnerischer Angriff gedeutet werden, die einen eigenen Angriff scheinbar rechtfertigen. So könnte es zu einem Atomkrieg aus Versehen kommen.

Durch die steigende Anzahl verfügbarer Sensoren und Überwachungssysteme, auch im Weltraum steigt die verfügbare Daten- und Informationslage in einer konkreten Entscheidungssituation überproportional an. So sind für die Klassifikation von Sensordaten und die Bewertung einer Alarmsituation immer mehr computergestützte Verfahren insbesondere der Künstlichen Intelligenz (KI) erforderlich, um für gewisse Teilaufgaben Entscheidungen automatisch zu treffen, bzw. menschliche Entscheidungen vorzubereiten.

Das Ende des INF-Vertrages hat bereits jetzt zu einem neuen Wettrüsten geführt, bei dem vor allem Hyperschallraketen eine hohe Priorität haben. Mit diesen neuen Waffen werden die Vorwarnzeiten weiter sinken. Es gibt bereits Forderungen autonome KI-Systeme für die Bewertung und Verarbeitung von Alarmmeldungen zu realisieren, da eventuell keine Zeit für menschliche Entscheidungen bleiben wird.

Entscheidungsträger wie Politiker und Militärs haben die Erwartung, dass KI-Systeme auch in Frühwarnsystemen zu besseren Entscheidungen fähig sind als Menschen, ähnlich wie dies auch für das autonome Fahren erwartet wird.

Dieser Artikel behandelt die Fragestellung, ob KI-Entscheidungen in solchen Frühwarnsystemen sinnvoll sein können und ob diese Systeme im Hinblick auf mögliche Fehlalarme durch KI sicherer gemacht werden können.

# 2 KI-Entscheidungen

## 2.1 Klassifikation

Bei vielen KI-Anwendungen geht es um Klassifikation. Hierbei besteht die Aufgabe darin, eine gegebene Situation oder ein Objekt auf sinnvolle Weise einer oder mehreren möglichen Klassen zuzuordnen. Eine gegebene Situation oder ein Objekt kann durch eine Reihe von Merkmalen (Symptomen) beschrieben werden und aus einer größeren Menge von gegebenen Klassen (Diagnosen) ist dann eine oder es sind mehrere auszuwählen, zu denen das Objekt, bzw. die Situation passt.

Die Bewertung von Sensorsignalen in Frühwarnsystemen ist auch eine Klassifikationsaufgabe: Aufgrund der Signale der Sensoren ist zu entscheiden, ob diese auf einen möglichen Angriff hinweisen. Zu den Teilaufgaben gehören Entscheidungen über den Typ angreifender Flugobjekte und die Art des Angriffs.

Die Ergebnisse von solchen Erkennungsaufgaben gelten immer nur mit einer gewissen Wahrscheinlichkeit, das heißt sie können falsch sein. In vielen Fällen kann mit den Erkennungsergebnissen auch ein Sicherheitsmaß angegeben werden, das ausdrückt, wie sicher das Ergebnis durch die automatische Erkennung eingeschätzt wird. Allerdings können Ergebnisse auch dann falsch sein, wenn die automatische Erkennung diese als sehr sicher einstuft.

Entsprechende Erkenntnisse gibt es auch in vielen anderen Anwendungen automatischer Klassifikation, wie bei OCR-Ergebnissen (Zeichenerkennung aus Bildern) oder bei automatischer Rechnungserkennung. Bei solchen Anwendungen werden Rechnungen automatisch gebucht und bezahlt, ohne dass ein Mensch dies prüft, falls das automatische System ein Erkennungsergebnis mit einem gewissen Sicherheitsmaß ausgibt. Aber auch in diesen Fällen kommen Fehlbuchungen und falsche Zahlungsvorgänge vor.

Das Gleiche gilt auch für Frühwarnsysteme. Alle Ergebnisse automatischer Erkennung gelten immer nur mit einer gewissen Wahrscheinlichkeit und können falsch sein. Für die üblichen Anwendungen mag dieses Risiko tragbar sein, für einen irreversiblen atomaren Einsatz mit Millionen von Toten und unabsehbaren gesundheitlichen, ökologischen und ökonomischen Folgen für die Menschheit, ist dies jedoch anders zu bewerten.

## 2.2 Datengrundlage

Entscheidungen in Frühwarnsystemen beruhen auf umfangreichen Daten, die von den Sensoren geliefert werden, sowie auf Kontext-Informationen, wie Erfahrungen aus

Szenariensimulationen, Gefährdungsanalysen oder Analysen der weltpolitischen Lage. Sowohl im Falle von menschlichen Entscheidern als auch bei Entscheidungen durch KI-Systeme werden solche Daten und Informationen benötigt und müssen entscheidungsrelevant zusammengefasst werden. Diese Datengrundlage ist allerdings unsicher, vage und unvollständig. Verarbeitungsaspekte von Vagheit, Unsicherheit und Unvollständigkeit in KI-Systemen werden nachfolgend kurz beschrieben.

### 2.2.1 Unsicherheit, Vagheit, Unvollständigkeit

**Sicheres Wissen**

Beispiel: Wenn x Kind von y ist und y Kind von z ist, dann ist x Enkel von z

Eine solche Regel kann als gültig angenommen werden. Schlüsse, die darauf basieren, führen wieder zu gültigen, korrekten Ergebnissen, sofern die Prämissen korrekt waren. Eine wichtige Eigenschaft ist hierbei, dass das Wissen interpretationsfrei - also frei von individueller Bewertung ist. Die Aspekte Unsicherheit, Vagheit, Unvollständigkeit treffen hier nicht zu.

**Unsicherheit**

Beispiel: Wenn x ein Auto ist und y ist der Besitzer von x, dann ist y derzeitiger Nutzer von x.

Eine solche Regel gilt nicht immer, es kann Ausnahmen geben. Der Nutzer eines Autos könnte ein Kind des Besitzers sein. Auch bei Firmen können Besitzer und Nutzer unterschiedlich sein. Eine solche Regel ist also unsicher, sie gilt nicht immer.

Weiteres Beispiel: Wenn eine Person x Fieber hat und x hat Husten
und x hatte in den letzten 10 Tagen Kontakt mit y
und y hat nachweislich eine Corona-Infektion,
dann hat auch x eine Corona-Infektion."

Dieser Zusammenhang muss nicht gelten, sondern gilt mit einer gewissen Wahrscheinlichkeit w. Eine Unsicherheit kann sich aber auch bereits in einer einfachen Aussage, wie “Person x hat eine Corona-Infektion” verbergen: Die Diagnose ist immer auch abhängig von Testverfahren, die üblicherweise nur mit einer gewissen Wahrscheinlichkeit ein korrektes Ergebnis liefern.

**Vagheit**

Beispiel: wenn x ein schweres Auto ist, dann benötigt x viel Kraftstoff.

Die Frage ist hier: was bedeutet „schwer“, was bedeutet „viel“.

In diesem Beispiel können die Aussagen „x ist ein schweres Auto“ und „x benötigt viel Kraftstoff“ nicht einfach mit den Wahrheitswerten wahr und falsch belegt werden. Diese Eigenschaften sind vage und der Wahrheitswert könnte hier als ein beliebiger Wert (reelle Zahl) zwischen 0 und 1 dargestellt werden, wobei 0 für falsch und 1 für wahr steht.

**Unvollständigkeit**

Die benötigten Informationen als Grundlage für Entscheidungen sind oft unvollständig. Da es häufig nicht möglich ist, vollständige Informationen zu erlangen, müssen Annahmen getroffen werden, die in typischer Weise gelten oder zu erwarten sind. Auf dieser Basis können dann Schlüsse gezogen und Entscheidungen getroffen werden.

Beispiel: Wenn x ein Vogel ist, dann kann x fliegen.

Dies ist zwar typisch, gilt im Normalfall, aber es gibt Ausnahmen. Ein Strauß ist ein Vogel, kann aber nicht fliegen. Ebenso kann das eingangs als "sicheres Wissen" eingeführte Beispiel auch als unvollständig betrachtet werden: Geht es hierbei um die biologische Verwandtschaft bzw. Abstammung oder um eine gesetzliche Regelung des Familienbegriffes?

Die Künstliche Intelligenz Forschung hat Verfahren entwickelt, um mit den unterschiedlichen Arten von Wissen und deren Grad an Glaubwürdigkeit umzugehen, aber die Schlussfolgerungen gelten dann ebenfalls nicht absolut, sondern nur wahrscheinlich, wie nachfolgend beschrieben.

### 2.2.2 Automatisches Schließen bei unsicherer Datengrundlage

In der Praxis gibt es viele Zusammenhänge, die unsicher sind, also nicht uneingeschränkt gelten. Unser normales Alltagswissen ist vage, unsicher und unvollständig. Trotzdem sind auch in solchen Situationen (z.B. Straßenverkehr) Schlussfolgerungen möglich und eventuell notwendig. In der KI sind verschiedene Methoden zur Behandlung von Unsicherheiten entwickelt worden.

Besonders wichtig sind Methoden des probabilistischen Schließens. Hierbei werden numerische Werte für die Gültigkeit von Formeln verwendet, die dann beim Schlussfolgern miteinander verrechnet werden. Verschiedene Wahrscheinlichkeitsmodelle unterscheiden sich darin, wie Formeln verknüpft werden können und wie die Wahrscheinlichkeitswerte dann verrechnet werden.

Auch zur Darstellung und Verarbeitung von vagen Werten werden meist numerische Werte verwendet.

In vielen Fällen kann Unsicherheit oder Unvollständigkeit so behandelt werden, dass zunächst eine „normale", „typische" Regelanwendung erfolgt. Typisch ist, dass Vögel fliegen können und dass der Besitzer eines Autos auch ein Nutzer dieses Autos ist. Solange nichts Gegenteiliges bekannt ist und kein Widerspruch entsteht, kann ein entsprechender Schluss gezogen werden. Im Falle eines Konfliktes müssen dann geeignete Maßnahmen zur Auflösung des Konfliktes getroffen werden. Auch für diese Art von Schlussfolgerungen gibt es unterschiedliche Methoden in der KI , insbesondere logische Verfahren.

Unabhängig vom gewählten Verfahren ist die Behandlung von Vagheit und Unsicherheit recht komplex und die Schlussfolgerungen sind ebenfalls unsicher, das heißt diese können auch falsch sein. Falsche Annahmen und falsche Schlussfolgerungen führen häufig zu Inkonsistenzen. In diesen Fällen können Korrekturmaßnahmen vorgenommen werden.

Solange keine Inkonsistenzen auftreten, kann automatisch nicht festgestellt werden, dass ein Schluss falsch ist.

### 2.2.3 Unsicherheit in Frühwarnsystemen

Auch die Datengrundlage für Entscheidungen in Frühwarnsystemen ist vage, unsicher und unvollständig. Dies gilt sowohl für Menschen als auch für Maschinen. Fehler in Frühwarnsystemen sind z.B. durch spezielle Lichteffekte von Mond oder Sonne oder durch die Erfassung von Vogel-Schwärmen durch Radaranlagen entstanden.

Mit neuen technischen Möglichkeiten wird die Vielfalt an Sensordaten zur Erkennung eines Raketenangriffs wachsen. Auch die Vielfalt der Objekttypen, die zu erkennen sind, wird wachsen, z.B. durch eine zunehmende Anzahl an Objekten im Luftraum (Drohnen) und im Weltraum (Satelliten, Weltraumwaffen, Abwehrsystem). Zusätzlich können Kollisionen mit Weltraumschrott und ein Verglühen in der Erdatmosphäre Sensorsignale verursachen, die von den Frühwarnsystemen erfasst werden und schwer interpretierbar sind. Die Unsicherheit der Daten in Frühwarnsystemen wird also eher noch wachsen.

Bei der Bewertung von Sensorsignalen spielen auch vage Werte wie Helligkeit und Größe eine Rolle. Signale werden auch nicht immer auftreten, können also unvollständig sein. Dies kann insbesondere für neue lenkbare Raketensysteme gelten, die einer Erfassung ausweichen können. Des Weiteren sind für die elektronische Kampfführung Systeme wie „Kalaetron Attack“ entwickelt worden, die es ermöglichen sollen, eine Erkennung durch die gegnerische Flugabwehr abzuwehren.[1] Im Falle einer Angriffsmeldung kann also nicht sichergestellt werden, dass die Daten auf Basis mehrerer unabhängiger Signalquellen überprüft werden können.

Welche Auswirkungen fehlende Informationen und falsche Annahmen haben können, zeigt ein Vorfall während der Kuba-Krise 1962. Ein russisches U-Boot, das sich vor Kuba in internationalen Gewässern befand, wurde von der amerikanischen Marine eingekesselt und attackiert. Die Amerikaner wollten es zum Auftauchen zwingen und hatten Moskau darüber informiert. Was die Amerikaner nicht wussten:

- Die Akkus des U-Boots waren fast leer, die Klimaanlage war ausgefallen und die Temperatur an Bord lag über 45 Grad.
- Viele Besatzungsmitglieder waren am Rande einer Kohlendioxidvergiftung und ohnmächtig.
- Das U-Boot hatte seit Tagen keinen Kontakt mehr mit Moskau.
- Das U-Boot hatte eine Atomwaffe an Bord, die unter bestimmten Bedingungen, ohne weitere Freigabe von Moskau, eingesetzt werden durfte.

Aufgrund der Attacken glaubte die russische Besatzung, der Krieg sei bereits ausgebrochen, und musste über den Einsatz der Atomwaffe an Bord entscheiden. Der Kapitän des U-Boots hielt die Situation des U-Boots und der Besatzung für aussichtslos und entschied den

[1] Behördenspiegel, Mai 2020, Seite 45, https://issuu.com/behoerden_spiegel/docs/2020_mai

nuklearen Torpedo abzuschießen. Der Torpedo-Offizier stimmte dem Abschuss zu. Für die Entscheidung über den Atomwaffeneinsatz waren auf diesem Boot drei Offiziere zuständig, da hier auch der Flottenkommandant anwesend war. Nur wenn alle drei zustimmten, war ein Einsatz zulässig. Der dritte Offizier, Wassili Archipow, verweigerte die Zustimmung für den Abschuss und verhinderte damit möglicherweise einen atomaren Krieg.

Auch andere dokumentierte Fehlalarme zeigen, dass die Daten, die in Frühwarnsystemen angezeigt werden, unsicher sind, also falsch sein können. Bei einem Alarm müssen die vorliegenden Informationen bewertet werden. Die vorliegenden Informationen sind in der Regel aber keine vollständige Beschreibung einer gegebenen Situation. Wichtige Informationen können fehlen, d.h. für die Bewertung einer Bedrohungssituation müssen Annahmen getroffen werden, die auch falsch sein können. Da nicht davon ausgegangen werden kann, dass die Gegenseite zufällig handelt, müssen nicht nur die sichtbaren Aktionen und Reaktionen der Gegenseite richtig erkannt werden, sondern auch die zu Grunde liegende Absicht oder Strategie abgeleitet werden. Dazu müssen Hypothesen über den Datenstand der Gegenseite und deren Interpretation der Daten genauso eingeschlossen werden wie die Informationsasymmetrie durch fehlende Transparenz.

### 2.2.4 Sensornetze und Netzwerk

Eine wichtige Strategie im Umgang mit unsicheren Daten ist die Verknüpfung unterschiedlicher Sensoren und Kontextinformationen mit dem Ziel, dass die Mehrheit der Sensoren “richtig” liegen wird und eine kritisch falsche Entscheidung vermieden werden kann. In der Medizin, beispielsweise, werden häufig Diagnosestrategien entwickelt, die Tests mit unterschiedlichen Unsicherheiten verknüpfen. Soll eine Epidemie verhindert werden, ist es wichtig mit einem eher sensitiven Test zu starten, der darauf optimiert ist, alle Infizierten zu erfassen. Je sensitiver ein Test ist, desto eher geraten auch Gesunde in das Raster und werden als “infiziert” klassifiziert. Daher wird im Anschluss ein spezifischer Test angeschlossen, um auszuschließen, dass Gesunde als “infiziert” erfasst werden.

In Frühwarnsystemen werden zahlreiche und z.T. unterschiedlichste Sensoren miteinander verknüpft. Redundante Sensoren sollen dabei die Robustheit gegen Ausfälle und Messfehler erhöhen, unterschiedliche Sensoren hingegen die Unsicherheit und Fehlerrate der Daten durch systematische Fehlerkennung eines spezifischen Sensortyps verringern.

Daten die automatischen Entscheidungen zu Grunde liegen, entstehen aber auch an unterschiedlichen Orten (z.B. Sensoren) und werden über Netzwerke übermittelt. Dabei können Daten eines Sensors von Daten eines anderen Sensors abhängen, beispielsweise die Positionsschätzung eines per Kamera erfassten Objektes kann mit der eigenen Positionsbestimmung über GPS verknüpft werden. Ist diese fehlerhaft so wird auch die relative Schätzung der Position des Objects vor der Kamera fehlerbehaftet sein. Häufig wird aber bei Konstruktion eines Sensors nicht über dessen exakte Nutzung in dem Frühwarnsystem entschieden, sondern die Daten werden der übrigen Datenverarbeitung im Frühwarnsystem zur Verfügung gestellt. Darauf aufbauend können neue Sensoren abgeleitet werden. So entstehen interdependente Netzwerke von Sensoren, in denen sich richtige,

unsichere und falsche Informationen beinahe in Lichtgeschwindigkeit durch das Netzwerk propagieren können.

Zusätzlich stellen die Netzwerke ein Risiko dar: Daten können durch Fehler oder Angriffe möglicherweise zerstört oder verändert werden. Im Extremfall kann auch das Netzwerk und seine grundsätzliche Funktion selbst Ziel eines Angriffs oder "Opfer" eines Naturereignisses sein. Die Frühwarnsysteme müssen dann bei Ausfall eines Teils des Netzwerkes entscheiden, ob dies bereits Anzeichen eines Angriffs sind oder durch ein natürliches Ereignis hervorgerufen wurde.

## 2.3 Maschinelles Lernen und Simulation

Der aktuelle Erfolg und die große Wahrnehmung der KI als Problemlösungsansatz in zahllosen Anwendungsbereichen basiert auf dem Maschinellen Lernen. Mittels Maschinellem Lernen ist ein KI-System in der Lage die zu Grunde liegenden Daten zu erweitern und so das Entscheidungsverhalten den in den Daten repräsentierten Erfahrungen anzupassen. Für erfolgreiche lernende Systeme werden große Datenmengen benötigt, die vorwiegend als Realdaten in "echten" Krisensituationen gewonnen werden können. Sind keine Daten verfügbar, z.B., wenn die Entscheidungssituation nur sehr selten oder noch gar nicht eingetreten ist und somit keine Erfahrungsdaten vorliegen, so können Computersimulationen genutzt werden, um in plausiblen Szenarien Daten für die unterschiedlichen Sensoren und ggf. das Frühwarn- und Entscheidungssystem zu generieren. Mit diesen künstlichen Daten können dann die Lernansätze trainiert werden. Zu beachten ist dabei, dass Computersimulationen auf Modellen basieren, die die Zusammenhänge in der Welt verkürzt darstellen und nicht alle Einflussfaktoren berücksichtigen können.

Z.B. hat die Entwicklung von automatischen Unterstützungsfunktion für die Flugzeuge 737 Max 8 von Boing zu zwei Flugzeugabstürzen geführt (siehe auch Kap. 4). Ursache waren fehlerhafte Messdaten von falsch eingestellten Sensoren. Aufgrund von Fehlern in Konzeption und Programmierung konnten die Piloten sich nicht den Entscheidungen der Maschine widersetzen. Nicht alle Varianten, die in der Realität vorkommen können sind vorhersehbar und können mit Computersimulation überprüft werden, so hatten in diesem Fall Simulationen mit den fehlerhaften Messdaten gefehlt.

## 2.4 Kontextwissen

In Friedenszeiten und Phasen politischer Entspannung sind die Risiken relativ gering, dass die Bewertung einer Alarmmeldung zu einem atomaren Angriff führt. In solchen Situationen werden von menschlichen Entscheidern im Zweifelsfall Fehlalarme angenommen. Die

Situation kann sich jedoch drastisch ändern, wenn politische Krisensituationen vorliegen, eventuell mit gegenseitigen Drohungen oder wenn in zeitlichem Zusammenhang mit einem Fehlalarm weitere Ereignisse eintreten. Hierfür werden bei einer Bewertung Ursachen gesucht, d.h. es wird versucht kausale Zusammenhänge zu finden. Wenn solche kausalen Zusammenhänge gefunden werden und logisch plausibel sind, besteht die große Gefahr, dass diese als gültig angenommen werden, d.h. dass die Alarmmeldung als gültig angenommen wird, auch wenn es um zufälliges zeitliches Zusammentreffen von unabhängigen Ereignissen geht.

Wenn die weltpolitische Lage und sonstige Kontextinformationen von automatischen Entscheidungskomponenten eines Frühwarnsystems nicht verwendet werden, dann sind Fehlalarme immer gefährlich, auch in Friedenszeiten.

Wenn die KI-Systeme von Frühwarnsystemen auch solches Kontextwissen für ihre Entscheidungen verwenden sollen, dann gilt auch hier, dass die Datengrundlage hochgradig vage, unsicher und unvollständig ist.

Die Bewertung der weltpolitischen Lage ist Gegenstand eines Projektes mit dem Namen „Preview“, das die Bundeswehr im März 2018 gestartet hat, mit dem Ziel, auf der Basis von Methoden der Künstlichen Intelligenz Krisen und Kriege vorherzusagen. Dazu sollen große Datenmengen automatisch analysiert werden. Ausgewertet werden hierbei Internet-Quellen sowie militärische und wirtschaftliche Datenbanken und auch Geheimdienstinformationen. Die Art der verwendeten Daten umfasst ein großes Spektrum, wozu auch Handelsdaten, Marktpreise, demographische Entwicklungen, Kriminalitätsraten, Meinungen in sozialen Netzwerken oder Daten über politische Gewalt gehören. Die KI-Plattform Watson soll dazu u.a. eingesetzt werden. Auch in anderen Staaten (z.B. Schweden, USA) gibt es solche KI-basierten Systeme zur Vorhersage von Krisen und Kriegen.[2]

Auch wenn solche Vorhaben wie das Projekt Preview sinnvoll zur frühzeitigen Erkennung von potentiellen Krisen, z.B. in Afrika sein können und es derzeit keine Hinweise auf einen Zusammenhang mit Frühwarn- und Entscheidungssystemen gibt, kann ein solcher Zusammenhang eintreten: Wenn ein Frühwarnsystem einen Raketenangriff meldet und diese Situation über mehrere Alarmstufen hinweg in den entsprechenden Krisensitzungen bewertet wird, ist es durchaus möglich, dass Kommissionsmitglieder auch Zugriff auf ein solches System zur Kriegsvorhersage haben. Wenn dieses KI-System in einer solchen Situation einen Krieg vorhersagt, kann dies erheblichen Einfluss auf die Bewertung der Alarmmeldung durch die Kommissionsmitglieder haben.

[2] Süddeutsche Zeitung, 23.7.2018, Seite 5 und 9.10.2018, Seite 16

# 3 Entscheidungsvorschläge

In der deutschen Öffentlichkeit wird mit großer Übereinstimmung gefordert, dass Entscheidungen zur Tötung von Menschen nicht automatisch erfolgen dürfen, sondern dass eine solche Entscheidung nur von einem Menschen getroffen werden darf. Solche Forderungen betreffen vorwiegend autonome Waffensysteme, aber sie müssen gleichermaßen auch für Gegenreaktionen auf Alarmmeldungen in Frühwarnsystemen gelten.

Auch wenn eine solche Forderung eingehalten wird, haben Menschen in der Regel wegen der kurzen Zeitspanne keine echte Entscheidungsmöglichkeit. Die einem maschinellen Entscheidungsvorschlag zu Grunde liegenden Informationen sind zu komplex, um diese in der kurzen verfügbaren Zeit (nur wenige Minuten) überprüfen zu können.

Eine fachliche Beurteilung der von einem KI-basierten System getroffenen Entscheidungen durch Menschen ist in der kurzen verfügbaren Zeit praktisch unmöglich. Dies gilt schon deshalb, weil die automatische Erkennung oft auf Hunderten von Merkmalen basiert. Die KI-Systeme können in der Regel keine einfachen nachvollziehbaren Begründungen liefern und selbst wenn Erkennungsmerkmale von einem KI-System ausgegeben werden, könnten diese nicht in der verfügbaren Zeit überprüft werden.

Dem Menschen bleibt deshalb meist nur zu glauben, was ein KI-System liefert. Die zunehmende Verbreitung von KI-Systemen in unserer Alltagswelt fördert zudem das Vertrauen in die Entscheidungskompetenz von technischen Systemen und es ist zu erwarten, dass automatisierte Entscheidungsvorlagen oder Lagebeurteilungen von den entsprechenden Menschen als nur schwer zu ignorierende Faktoren zu bewerten sind. Möglichkeiten, wie sie beim autonomen Fahren durch Austesten der Grenzen von Assistenzsystemen bestehen, sind in solch komplexen Frühwarn- und Entscheidungssystemen nicht oder nur eingeschränkt zu realisieren.

# 4 “Segen” oder Kontrollverlust bei Teilautomatik

Einhergehend mit neuer Technik entstehen zumeist auch neue Risiken und Ängste. So wurde aus Teilen der Bevölkerung häufig eine irrationale Gefahr beschrieben, die real gar nicht bestand. Beispielsweise wurde die Gefahr schneller Reisen im Zug im 19. Jahrhundert als gefährlich eingestuft und aktuell die Gefahr von autonomen Fahrzeugen oder Robotern. Dabei hat objektiv die Sicherheit durch den Einsatz neuer Technologien zugenommen und beim autonomen Fahren gehen viele Experten von einer deutlich reduzierten Unfallgefahr aus. Wichtig ist bei der Entwicklung von Innovationen zu zuverlässigen Technologien immer der wiederholte Kreislauf von “Trial-and-Error”: Die Fehlklassifikation einer LKW-Plane als “freie Straße” im früheren Einsatz des Tesla, die zu einem schweren Unfall führte, wurde behoben und führte zu besseren und robusteren Klassifikationsverfahren. Jedoch sind in solchen Fällen die Risiken überschaubar. Bei den Frühwarn- und Entscheidungssystemen ist ein Ausprobieren auf Grund der katastrophalen Folgen jedoch nicht möglich. Hier hat bereits der erste Fehler das Potential, der Gesellschaft und seiner Ökologie irreparable Schäden in einer Dimension zuzufügen, die einen Fortbestand unserer modernen Gesellschaft gefährdet.

Die deutsche Industrie- und Wirtschaftsgeschichte ist durch konsequente Nutzung von Automatisierungspotential und Substitution menschlicher Arbeit geprägt. Die 2012 ausgerufene Hightech-Initiative “Industrie 4.0” sieht eine zunehmende Vernetzung sowie die intensive Nutzung von Daten, KI und Prognose-Systemen in der industriellen Produktion und Logistik vor und wird in die unterschiedlichsten Domänen übertragen. Damit soll eine höhere Flexibilität und Robustheit bei gleichzeitiger Steigerung der Effizienz erreicht werden. KI-Systeme sind für die breite Bevölkerung durch intelligente Sprachassistenten der großen Technikkonzerne sowie durch zahlreiche Assistenzsysteme in modernen Kraftfahrzeugen zunehmend erfahrbar. Das heißt, bei vielen Anwendungen können KI-Systeme zu besseren Entscheidungen fähig sein als Menschen. Dies wird z.B. auch für das autonome Fahren erwartet; hier wird sogar darüber spekuliert, ob autonome Fahrzeuge dazu führen könnten, dass keine Verkehrsunfälle mehr vorkommen. Voraussetzung hierfür sind umfangreiche Lerndaten auf Basis vieler Tests, auch unter realen Bedingungen. Auch wenn die Anzahl der Unfälle pro gefahrene Strecke deutlich geringer ist als bei menschengeführten Fahrzeugen, passieren dennoch Unfälle.

Für KI-Entscheidungen in Frühwarnsystemen gelten jedoch andere Bedingungen. Für die zuletzt so erfolgreichen „deep-learning“-Ansätze besteht das Problem, dass „Lerndaten“ für die Erkennungsaufgaben in Frühwarnsystemen nur sehr eingeschränkt verfügbar sind. Ein Testen in realen Situationen ist kaum möglich. Auch auf Basis weniger Beispiele können KI-basierte Erkennungen realisiert werden, aber es ist nicht möglich, alle Varianten und Ausnahmensituationen vorherzusehen, die vorkommen können. Deshalb kann es zu falschen Klassifikationsergebnissen kommen.

Ein Unfall auf Grund einer falschen Entscheidung eines KI-Systems beispielsweise beim autonomen Fahren kann auch einzelne Menschenleben fordern oder im industriellen Kontext zu Produktionsstop oder Einnahmeverlusten führen. Die Folgen sind jedoch begrenzt und lassen sich durch kurzfristige Änderungen an den Programmen in ihren Folgen eingrenzen. In einem Frühwarnsystem mit der Folge eines Atomkriegs aus Versehen würde es jedoch zu nicht umkehrbaren Folgen führen, bei der im Extremfall Millionen von Menschen getötet und

durch den nuklearen Winter im Anschluss noch Milliarden Menschen die Lebensgrundlage genommen werden würde.

In der Vergangenheit gab es immer wieder Drohungen, bei einer Angriffsmeldung die eigenen Raketen vollautomatisch durch Computer zu starten, ohne Eingriffsmöglichkeiten durch Menschen. Dies lässt vermuten, dass entsprechende Software-Komponenten entwickelt wurden, auch wenn nicht die Absicht bestand, diese in vollautomatischer Variante einzusetzen.

Die beiden Abstürze der Boeing 737 Max 8 Flugzeuge vom 29.10.2018 und 10.3.2019 zeigen, dass es durch eine falsche und ungünstige Programmierung passieren kann, dass Menschen sich den Entscheidungen der Maschine nicht widersetzen können. Obwohl sich die beiden Piloten korrekt verhalten hatten, konnten Sie die Abstürze und den Tod aller Insassen nicht verhindern. Sie waren nicht in der Lage, die falschen Entscheidungen der Maschine zu korrigieren.

Wenn eine oder mehrere der Nuklearstreitmächte in Zusammenhang mit Frühwarnsystemen Software-Komponenten im Einsatz haben, die bei einer Angriffsmeldung eine Gegenreaktion (teil-)automatisch unterstützen können, ist nicht auszuschließen, dass durch Fehler in der Konzeption oder Realisierung diese Software-Komponenten Aktionen ausführen, die von Menschen nicht gestoppt werden können, ähnlich wie bei den beiden Flugzeugabstürzen. Genau wie bei den Flugzeugabstürzen erfolgen Aktionen bei Frühwarnsystemen innerhalb weniger Minuten unter enormen Zeitdruck.

Wenn es solche teilautomatischen Komponenten in Frühwarnsystemen gibt, werden diese vermutlich auch getestet, z.B. durch Simulation. Auch solche Tests und Simulationen könnten durch Programmfehler außer Kontrolle geraten. Die Katastrophe von Tschernobyl ist durch einen solchen Test ausgelöst worden.

# 5 Vergleich Entscheidungen Mensch – Maschine

**Zwei Beispiele,** Beispiel 1:

Im Januar 2020 hatten die USA den iranischen General Soleimani mit einem Drohnenangriff getötet. Als Vergeltungsangriff hat Iran wenige Tage später amerikanische Stellungen im Irak angegriffen. Kurz danach wurde im Iran ein ukrainisches Verkehrsflugzeug aus Versehen abgeschossen. Die Bedienungsmannschaft kam zu dem Ergebnis, dass es sich bei dem Flugobjekt um einen angreifenden Marschflugkörper handeln könnte. Die Fehlentscheidung kam vor allem dadurch zu Stande, dass die Bedienungsmannschaft mit Krieg oder einem Angriff der USA gerechnet hatte.

In dieser Situation hätte eine Maschine möglicherweise besser entschieden. Denn die reinen Fakten, wie Größe des Radarsignals, hätten vermutlich gegen einen Marschflugkörper gesprochen. Vielleicht hätte eine Maschine in der Kürze der Zeit auch mehr Informationen, wie z.B. Flugpläne, berücksichtigen können. Die Bedienungsmannschaft hatte den politischen Kontext vermutlich überbewertet.

Beispiel 2:

Ein Satellit des russischen Frühwarnsystems meldet am 26.9.1983 fünf angreifende Interkontinentalraketen. Da die korrekte Funktion des Satelliten festgestellt wurde, hätte der diensthabende russische Offizier Stanislaw Petrow nach Vorschrift die Warnmeldung weitergeben müssen. Er hielt einen Angriff der Amerikaner mit nur fünf Raketen aber für unwahrscheinlich und entschied trotz der Datenlage, dass es vermutlich ein Fehlalarm sei und verhinderte damit eine Katastrophe mit atomarem Schlag und Gegenschlag. Der Vorfall ereignete sich während einer instabilen politischen Lage: Die Nachrüstung durch Mittelstreckenraketen stand an und wenige Wochen vorher hatten die Sowjets aus Versehen eine koreanische Passagiermaschine über internationalen Gewässern abgeschossen. Möglicherweise hätte eine Maschine aufgrund der Fakten den Angriff eher als echt eingeschätzt und Gegenreaktionen eingeleitet. Petrow hatte gefühlsmäßig auf einen Fehlalarm gehofft, wollte nicht für den millionenfachen Tod von Menschen verantwortlich sein und hat sich entsprechend entschieden.

**Immer kürzere Entscheidungszeiträume**

Entscheidungen in Frühwarnsystemen müssen in sehr kurzen Zeitspannen fallen und ein neues Wettrüsten wird diese Zeiten weiter reduzieren.

Wenn im Falle einer Alarmmeldung ein Mensch noch Möglichkeiten zur Bewertung relevanter Merkmale hat, könnte er zu dem Schluss kommen, dass die Zeit nicht reicht, um einen Fehlalarm auszuschließen. Konsequenterweise sollte er dann sicherheitshalber keine Gegenreaktion auslösen.

Für eine Maschine wird die Zeit reichen, eine Entscheidungsgrundlage zu schaffen. Wenn eine Angriffsmeldung mit gewisser Wahrscheinlichkeit als gültig eingestuft wird, könnte eine automatische Gegenreaktion eingeleitet werden, auch wenn es sich um einen Fehlalarm handelt. Für eine Maschine reicht die Zeit für eine Entscheidung, auch wenn diese falsch ist.

# 6 Fazit

Weder Menschen noch Maschinen können bei Alarmmeldungen in Frühwarnsystemen in so kurzer Zeit fehlerfrei entscheiden, da die Datengrundlage unsicher, vage und unvollständig ist und eine Überprüfung durch Menschen in der kurzen verfügbaren Zeit nicht möglich ist.

Es ist weder möglich, zu sagen, dass im Zweifelsfall Menschen die bessere Entscheidung treffen, noch gilt, dass im Zweifelsfall Maschinen die bessere Entscheidung treffen.

Bei Entscheidungen des Menschen kann das Ergebnis unter anderem davon abhängen, wer gerade Dienst hat und wie die momentane Grundeinstellung derjenigen ist.

Bei Entscheidungen durch Maschinen kann das Ergebnis unter anderem abhängen von der Wahl von Merkmalen mit Prioritäten durch die Programmierer oder einer vorhandenen Datenbasis als Lerngrundlage. Bei der Festlegung von Merkmalen und Prioritäten durch Programmierer bzw. der Festlegung einer Lerngrundlage sind die Auswirkungen für bestimmte Alarmsituationen nicht abschätzbar.

Es darf nicht sein, dass von der Entscheidung eines einzelnen Menschen oder einer Maschine das Überleben der gesamten Menschheit abhängt. Deshalb ist der Ansatz Frühwarnsysteme zu verwenden, um frühzeitig Atomangriffe zu erkennen und eventuell einen Gegenangriff zu starten, bevor die gegnerischen Raketen einschlagen, grundsätzlich untragbar, unabhängig davon, ob letztendlich Menschen oder Maschinen entscheiden. Diese Problematik kann durch den Einsatz von KI-Technologien nicht behoben werden.

Weitere Informationen zum Thema „Atomkrieg aus Versehen“: https://atomkrieg-aus-versehen.de/ , dort gibt es auch Hinweise auf weitere Artikel, wie z.B. https://www.fwes.info/fwes-19-3.pdf