

Inhalt

Vorwort	7
<i>Karl Hans Bläsius/Reiner Schwalb/Michael Staack</i>	
KI in Frühwarnsystemen für nukleare Bedrohungen	11
<i>Karl Hans Bläsius/Jörg Siekmann</i>	
Atomkrieg aus Versehen – Risiko und Folgen	23
<i>Karl Hans Bläsius</i>	
Keine Entwarnung	35
<i>Hans-Peter Bartels</i>	
Mehr Aufklärung und Transparenz	39
<i>Katja Keul</i>	
Künstliche Intelligenz – Gefahren und Chancen	43
<i>Alexander Müller</i>	
Atomkrieg aus Versehen? – Optionen der Prävention	49
<i>Wolfgang Richter</i>	
Atomkriegsrisiko und Russland-Ukraine-Krieg	65
<i>Michael Staack/Karl Hans Bläsius/Reiner Schwalb</i>	
Die Autorinnen und Autoren	71

KI in Frühwarnsystemen für nukleare Bedrohungen

Karl Hans Bläsius/Jörg Siekmann

Computergestützte Frühwarn- und Entscheidungssysteme sollen einen Angriff mit Atomwaffen rechtzeitig erkennen, damit bei entsprechender Situationsbewertung die eigenen atomaren Trägerraketen gegebenenfalls vor einem vernichtenden Einschlag aktiviert werden können. Die Zeit zur Bewertung einer potenziellen Angriffssituation ist gering und liegt im Bereich weniger Minuten. Neue Waffensysteme wie Hyperschallraketen werden diese Zeitspanne noch weiter verkürzen. Des Weiteren werden Anzahl und Vielfalt an Objekten im Luftraum weiter steigen, so dass die Analyse möglicher Angriffssituationen immer schwieriger wird. Für eine Analyse und Bewertung von Alarmmeldungen werden daher immer mehr automatische Systeme unter Verwendung von Methoden der Künstlichen Intelligenz (KI) erforderlich sein. Aber auch KI-Systeme können bei solchen Anwendungen keine sicheren Ergebnisse liefern, denn die zugrundeliegenden Daten sind begrenzt valide, vage und unvollständig. Automatische Erkennungsergebnisse gelten deshalb nur mit einer gewissen Wahrscheinlichkeit und können falsch sein. Die Komplexität möglicher Bedrohungslagen kann auch durch Cyberangriffe und autonome Waffensysteme noch weiter erhöht werden. Das Risiko eines Atomkriegs aus Versehen als Folge eines Fehlalarms und einer Fehleinschätzung der Gefährdungslage kann deshalb in den nächsten Jahren und Jahrzehnten erheblich steigen, insbesondere im Falle internationaler Krisen.

1. Abschreckungsstrategie

Die Sicherung der atomaren Zweitschlagfähigkeit ist die Grundlage der Abschreckungsstrategie, die bis heute jeden potenziellen Angreifer abgehalten hat, einen atomaren Angriff zu starten. Wer angegriffen wird, kann den Einschlag von Atomwaffen abwarten und hat danach immer noch genügend Zeit und Potenzial, einen vernichtenden Gegenschlag auszuführen. Im Schlagwort: „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.

Auch wenn die Abschreckungsstrategie weitere bewusste Atomwaffeneinsätze bisher verhindert hat, schützt sie nicht vor einem Atomkrieg aus Versehen, z.B. in Folge eines Fehlers in einem Frühwarnsystem. In der Vergangenheit gab es einige Situationen, in denen es nur durch großes Glück nicht zu einem Atomkrieg aus Versehen kam.

2. Frühwarnsysteme für nukleare Bedrohungen

Frühwarnsysteme für nukleare Bedrohungen dienen der Erkennung eines Angriffs durch Atomwaffen auf der Basis von Sensordaten. Sensordaten können Hinweise, aber keine sicheren Erkenntnisse über mögliche Bedrohungen liefern. Es kann also zu Fehlalarmen kommen, wobei ein Angriff mit Atomraketen angezeigt wird, obwohl keine Bedrohung vorliegt. Solche Fehlalarme können ganz unterschiedliche Ursachen haben. Neben unklaren oder falsch bewerteten Sensordaten haben z.B. auch Hardware-, Software- oder Bedienungsfehler zu falschen Alarmmeldungen geführt. Es ist grundsätzlich nicht möglich, hochkomplexe Softwaresysteme fehlerfrei zu entwickeln.

Die Datengrundlage für die Bewertung einer Alarmmeldung ist vage, unsicher und unvollständig. Zur Situationsbewertung sind deshalb neben den Sensordaten viele weitere Aspekte zu berücksichtigen, wie z.B. die weltpolitische Lage und die Einschätzung des „Gegners“. In Friedenszeiten und Phasen politischer Entspannung sind die Risiken gering, dass die Bewertung einer Alarmmeldung zu einem atomaren Angriff führt. Die Situation kann sich drastisch ändern, wenn politische Krisensituationen vorliegen, eventuell verbunden mit gegenseitigen Drohungen, oder wenn in zeitlichem Zusammenhang mit einem Fehlalarm weitere Ereignisse eintreten, die zur Alarmmeldung in Zusammenhang gesetzt werden könnten. In einer solchen Situation könnte eine Alarmmeldung als echt gedeutet werden und einen eigenen Angriff scheinbar rechtfertigen. So könnte es zu einem Atomkrieg aus Versehen kommen.

Eine steigende Anzahl verfügbarer Sensoren und Überwachungssysteme, auch im Weltraum, erhöht die zu verarbeitende Datenmenge, wobei gleichzeitig die Vorwarnzeiten weiter reduziert werden, z.B. durch neue Waffensysteme wie Hyperschallraketen. 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. Dabei ist es unerheblich, ob solche automatischen

Systeme als intelligent angesehen und dem Gebiet der Künstlichen Intelligenz zugeordnet werden oder nicht.

3. Automatische Entscheidungen

Entscheidungen in Frühwarnsystemen beruhen auf umfangreichen Daten, die von den Sensoren geliefert werden, sowie auf Kontext-Informationen, Gefährdungsanalysen oder Analysen der weltpolitischen Lage. Sowohl im Falle von menschlichen Entscheidern als auch bei automatischen Entscheidungen werden solche Daten und Informationen benötigt und müssen entscheidungsrelevant zusammengefasst werden. Diese Datengrundlage ist allerdings hoch fehleranfällig.

3.1 Unsicherheit, Vagheit, Unvollständigkeit

Zunächst ein Beispiel für sicheres Wissen:

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. Die Aspekte Unsicherheit, Vagheit, Unvollständigkeit treffen hier nicht zu.

Das nächste Beispiel zeigt eine unsichere Regel: 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. Diese Regel ist also unsicher und gilt nur mit einer gewissen Wahrscheinlichkeit.

Beispiel für einen vagen Zusammenhang: 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.

Bei vielen Anwendungen ist es nicht möglich, vollständige und sichere Informationen zu erlangen. Stattdessen 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.

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 mit gewisser Wahrscheinlichkeit.

3.2 Automatisches Schlussfolgern bei unsicherer Datengrundlage

In der Praxis gibt es viele Zusammenhänge, die unsicher sind, also nicht uneingeschränkt gelten, trotzdem sind auch in solchen Situationen Schlussfolgerungen möglich und eventuell notwendig.

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 und in diesen Fällen können Korrekturmaßnahmen vorgenommen werden. Solange keine Inkonsistenzen auftreten, kann automatisch nicht festgestellt werden, dass ein Schluss falsch ist.

3.3 Unsicherheit in Frühwarnsystemen

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 Raketenysteme gelten, die einer Erfassung ausweichen könnten. Des Weiteren sind für die elektronische Kampfführung Systeme entwickelt worden, die es ermöglichen sollen, eine Erkennung durch die gegnerische Flugabwehr abzuwehren. Im Falle einer Angriffsmeldung kann also nicht sichergestellt werden, dass die Daten auf Basis mehrerer unabhängiger Signalquellen überprüft werden.

3.4 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, die in Beziehung gesetzt werden können. 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 Friedenszei-

ten. Wenn Komponenten 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 u.a. eingesetzt werden, um solche Informationen zu verarbeiten und zu bewerten. Auch in anderen Staaten (z.B. Schweden, USA) gibt es solche KI-basierten Systeme zur Vorhersage von Krisen und Kriegen.

Auch wenn solche Vorhaben wie das Projekt *Preview* sinnvoll zur frühzeitigen Erkennung von potenziellen Krisen sein können und es derzeit keine Hinweise auf einen Zusammenhang mit Frühwarn- und Entscheidungssystemen für die Erkennung eines nuklearen Angriffs 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.

4. Entscheidungsvorschläge

In der Öffentlichkeit und insbesondere auch von KI-Forscher*innen wird gefordert, dass Entscheidungen zur Tötung von Menschen nicht automatisch erfolgen, sondern dass solche Entscheidungen nur unter echter menschlicher Kontrolle zulässig sein dürfen. Führende KI-Forscher*innen haben dazu einen offenen Brief verfasst, der u.a. hier veröffentlicht ist: www.autonomewaffen.org. Das Ziel, letale Autonome Waffensysteme zu verbieten, ist auch im neuen Koalitionsvertrag von SPD, Grünen und der FDP enthalten: „Letale Autonome Waffensysteme, die vollständig der Verfügung des Menschen entzogen sind, lehnen wir ab. Deren internationale Ächtung treiben wir aktiv voran“ (Koalitionsvertrag 2021: 115).