Artificial Intelligence
and Nuclear War Risk

Karl Hans Bläsius,   15.11.2023

www.unintended-nuclear-war.eu

www.ai-implications.eu

Content

- Unintended nuclear war, AI

- Warnings about AI

- Possible solutions

Early warning and decision-making systems

Goal: Early detection of an attack with nuclear missiles

Structure :

- Sensors (e.g. satellites, radar, listening sensors in oceans)

- Computer centers and communication networks for data analysis and transmission

- Command centers for evaluating warning information and the threat situation as well as for planning and ordering actions.

Problem: Danger of nuclear war by mistake

                due to computer error (false alarm)

Launch on Warning

- High accuracy of modern weapons: Missile silos and command centers may be disabled, second strike capability may be compromised

- Launch of the own missiles before the enemy missiles hit

- Threats with „Launch on Warning"

- Probably not a fixed strategy, but a decision depending on the situation

- dependent on second-strike capability, political situation, expectations

Examples of false alarms

1962: Cuba crisis: several alarms and incidents (e.g. Archipow)

9.11.1979: Message mass attack, simulation program

3. and 6.6. 1980: Message mass attack, Hardware error

26.9. 1983: Stanislav Petrov prevents unintended nuclear war

25.1. 1995: Boris Yeltsin: Nuclear suitcase activated, launch codes ready

12.12. 2020: Warning of missile attack in Ramstein

False alarms are usually not made public (confidentiality)

Risk of "nuclear war by mistake"

A false alarm as a single event in peacetime is not critical.

Several critical situations in the past, fortunately nothing happened so far

Risk of "accidental nuclear war" increases if:

- Crisis situation exists (global political situation, threats)
- unusual errors occur
- several events occur simultaneously
- acting persons make misjudgments

Course of Confrontation: the West – Russia, China

New arms race in technology fields such as AI, cyberspace

- Every nation must be at the forefront
- Keeping up with potential competitors

Software-based weapons:

- Hidden development, any number of variants, copies
- Arms control and agreements are difficult
- Proliferation difficult to control, disarmament difficult
- Whatever autonomous weapons are developed now will always remain

Risk of "accidental nuclear war" increases

New technical developments:

- Space armament, Hypersonic missiles

- Cyber warfare capabilities

- AI, Autonomous weapon systems

Consequences :

Threat situations are becoming increasingly complex,

alarm messages very difficult to evaluate,

overall system of early warning components and nuclear weapons

increasingly difficult to control

Early Warning Systems – AI

- Increasing number of objects in airspace

- Increasing variety of object types in the airspace

- Warning times ever shorter (e.g. hypersonic weapons)

- Evaluation of sensor data very difficult

- AI methods required for evaluation

Data and information basis for AI decisions

vague, uncertain, incomplete

Vagueness - Uncertainty

Vagueness: not only **true** - **false**, intermediate values possible

If x is a heavy car, then x needs lots of fuel

What does mean: "heavy", "lots of" ?

Uncertainty: Rule only applies with a certain probability

If x is the owner of car y, then x is the user of y

Our everyday knowledge is usually vague and uncertain !

Incompleteness - vagueness - uncertainty

Early warning systems: lots of sensors, lots of input data

Information is

- **incomplete**: e.g. due to electronic warfare

- **vague**: size of radar signal, brightness of missile signal, ...

- **uncertain**: detection of object type, object characteristics, ...

Assumptions are required: may be incorrect

AI decisions

- plausible conclusion necessary in case of incomplete information

- Conclusions with incomplete and uncertain knowledge can be wrong

- AI: Hundreds of features, often no simple reasoning possible

- hardly (not) verifiable by humans

- Problem of uncertainty cannot be solved with AI

- Neither humans nor machines can make reliable decisions

➔ Important for evaluation :   trust,   expectations (can AI help ?)

AI-based prediction of wars and crises

Project „Preview" of the Bundeswehr (since 2018)

- Currently other objective

- Application in the future?

- Data basis: vague, uncertain, incomplete

- Assuming Petrov had had access to such a system, what decision would he have made?

Nuclear war risk

Dangerous interactions: AI (autonomous weapons) - cyber - nuclear weapons

=> uncontrollable for humans (extremely complex, short decision times)

Crises (e.g. climate)

Warning times shorter

Nuclear war by mistake: suddenly,

unexpectedly, as an accident

Content

- Unintended nuclear war, AI

- Warnings about AI

- Possible solutions

Current warnings on AI

- Pause AI development

- Warning of destruction by AI

- Superintelligence this decade

- media response: low

Center for AI Safety

Statement on AI Risk

AI experts and public figures express their concern about AI risk.

Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.

Statement on AI Risk - Signatories

Demis Hassabis, CEO, Google DeepMind

Sam Altman, CEO, OpenAI

Dario Amodei, CEO, Anthropic

Bill Gates, Gates Ventures

Ilya Sutskever, Co-Founder and Chief Scientist, OpenAI

Shane Legg, Chief AGI Scientist and Co-Founder, Google DeepMind

James Manyika, SVP, Research, Technology & Society, Google-Alphabet

Eric Horvitz, Chief Scientific Officer, Microsoft

Albert Efimov, Chief of Research, Russian Association of Artificial Intelligence

Alvin Wang Graylin, China President, HTC

Stuart Russell, Professor of Computer Science, UC Berkeley

Peter Norvig, Education Fellow, Stanford University

Warning of destruction by AI

Risks:

- Autonomous weapon systems

- Interactions between AI and nuclear weapons

- Revolution in warfare through AI

- Bio-, chemical weapons developed with AI

- Disinformation and manipulation

- Superintelligence

Higher risk: Unintended nuclear war

AI more dangerous than nuclear weapons ?

|	Nuclear weapons:	AI – SI:
Development goal:	Weapons	Dual-Use, Positive
Control:	Human	Machine
Effects:	Local, possibly global	global

Superintelligence

Consequences already due to preliminary stages ?

information dominance on the Internet :

- Previously unknown cyber threats
- Cyberwar through systems
- Takeover of information display and exchange
- Consequences for the financial sector
- Societies unstable

Risk of nuclear war (possibly accidental) increases

Superintelligence: important books
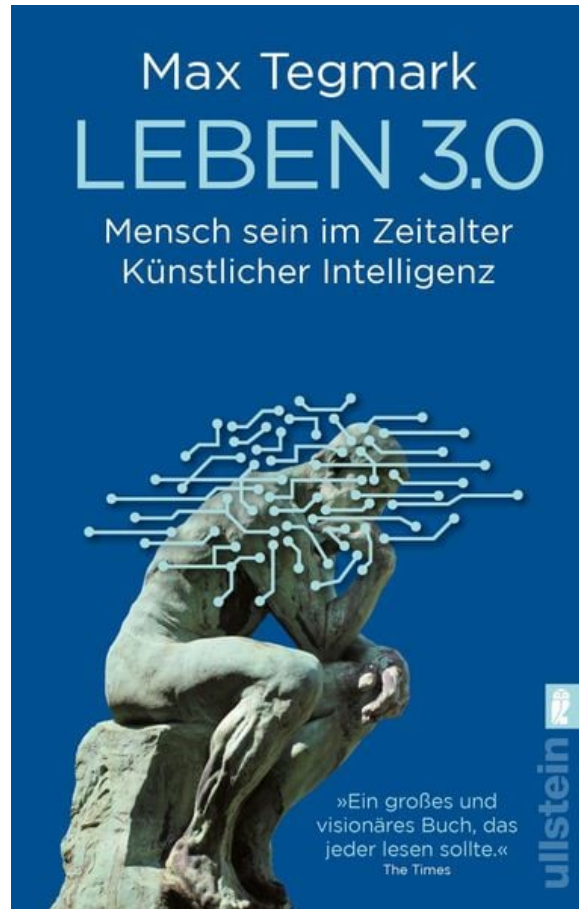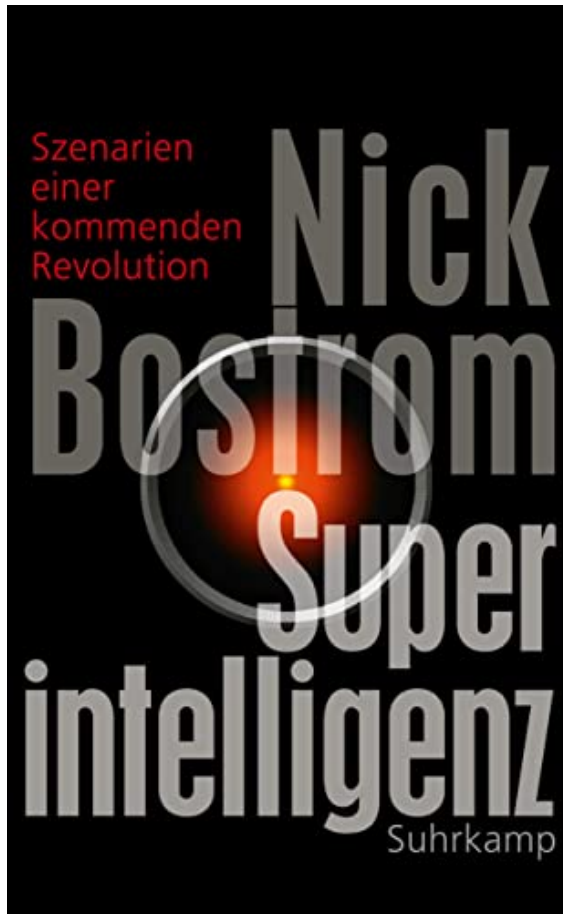
Warnings about AI - temporal aspects

- No precise predictions possible

- Potentially high risk for humanity as a whole

- Effects sudden, unexpected, like an accident

- before: no scientific findings,

- before: no proof of risks

- Waiting for experience may not be enough

Content

- Unintended nuclear war, AI

- Warnings about AI

- Possible solutions

Course of Confrontation: the West – Russia, China

Should this arms race continue?

- Cyber weapons, AI, autonomous weapons,

- Nuclear weapons, Bioweapons, Space, …

Or should we prevent such an arms race?

How?

How much time is left?

Course of Confrontation

Consequences of the current course of confrontation:

- Dangerous types of weapons are being developed further, becoming even more dangerous

- Superintelligence: increased efforts (China, India, Russia, …)

- Risks are increasing in many areas, including nuclear war risk

The course of confrontation should be ended:

- Can this be achieved by supplying ever more effective weapons to war zones?

What to do?

Important demands of the peace movement:

stop nuclear sharing and implement a ban on nuclear weapons

Is that enough?

Is this currently achievable?

Or are the necessary preconditions lacking?

What to do?

Important demands should be:

- Disarmament, arms control for nuclear weapons

- Effective agreements against climate change

- Agreements on cyber weapons, autonomous weapons

- Regulation of AI

necessary prerequisites:

- Effective agreements are only possible with Russia and China

- stop the confrontation course as soon as possible

- stop the war and conduct negotiations

Frequent counterarguments

Talks and negotiations are not possible with Putin.

Only military strength will help. Is that true?

Consequence: War Nato – Russia ?

It may be that there is no solution.

What to do when a solution is absolutely necessary?

Perhaps mathematics can help.

Math: Linear system of equations

M equations with N Variables:

$$a_{1,1} x_1 + a_{1,2} x_2 + \quad \cdots \quad + a_{1n} x_n \quad = \quad b_1$$

$$a_{2,1} x_1 + a_{2,2} x_2 + \quad \cdots \quad + a_{2n} x_n \quad = \quad b_2$$

$$\vdots$$

$$a_{m1} x_1 + a_{m2} x_2 + \quad \cdots \quad + a_{mn} x_n \quad = \quad b_m$$

M = N: exactly one solution

M < N: many solutions

M > N: no sulution

Math: Linear system of equations

M equations with N Variables and M > N: no solution

What to do when solutions are urgently needed?

Reduce the number of equations, i.e. conditions   and/or

increase the number of variables, e.g. by turning constants into variables

demands on politicians:

- reduce conditions
- turn constants into variables.

„Security through trade" : failed?

Or: not sufficient, but necessary ?

Difference: sufficient, necessary conditons

What does it mean:

- A is sufficient for B ?   If A, then B

- A is necessary for B ?   If B, then A

„Security through trade " not sufficient to prevent war in Ukraine,

but necessary for lasting global security

Quote Lahl, Varwick:

Book „Sicherheitspolitik verstehen" (understanding security policy):

„International interdependencies: This approach takes advantage of the realization that in a globally networked world, any violent confrontation only produces losers on the bottom line. The more the players are economically, technologically, culturally and possibly also militarily networked with each other, the less chance there is of achieving one-sided advantages through aggression. This idea, which is often wrongly perceived as naive, thus amounts to a kind of self-deterrence."

(translated with deepl)

Military. Contacts - scient. contacts

Important basis for arms control in recent decades :

- „Open Skies", mutual control, a certain level of trust

Scientific exchange in recent decades :

- A certain level of transparency
- Inclination of many scientists: no involvement in armaments
- Civil clauses at universities

Prohibition / restriction of scientific exchange with Russia / China

- Transparency ? Willingness to participate in arms projects ?
- Civil clauses at universities ?
- Acceleration of the AI arms race ?

Possible measures to reduce risks

Improving trust, communication and cooperation between **all** nations:

- political, scientific

- technological, economic

- culture, sport

- town twinning, school exchanges

- private

globalization and economic cooperation:

- wars should no longer be possible (self-deterrence)

Multiple risks

several serious risks in the near future:

- Climate change

- Conflicts and wars in various regions

- risk of escalation, also concerning nuclear weapons

- AI risks, e.g. disinformation, superintelligence,

- bioweapons with the help of AI (Terrorists),

- etc.

solutions:

- are only possible together, together with **all** nations

Thank you

www.unintended-nuclear-war.eu

www.ai-implications.eu