Impressions from the 63rd Pugwash conference in Hirishima

Tom Børsen

AALBORG UNIVERSITY

Dag 1: 80. årsdagen for bombningen af Hiroshima

- Filmvisning: Hiroshima: A Mother's Prayer fra 1990. Kan ses på Youtube.

- Hiroshima Peace Memorial Museum
  - Fotos af, artefakter fra, og kunst om bombningen af Hiroshima
- Vi husker – og skal huske – ofrene
- Message: Humanity and nuclear weapons cannot co-exist!
- Ingen byer, nationer, eller folk skal opleve det som Hiroshima oplevede
- Teknologisk fremskridt skal følges af en moralsk revolution!
- Total afskaffelse af WMD

President of Pugwash

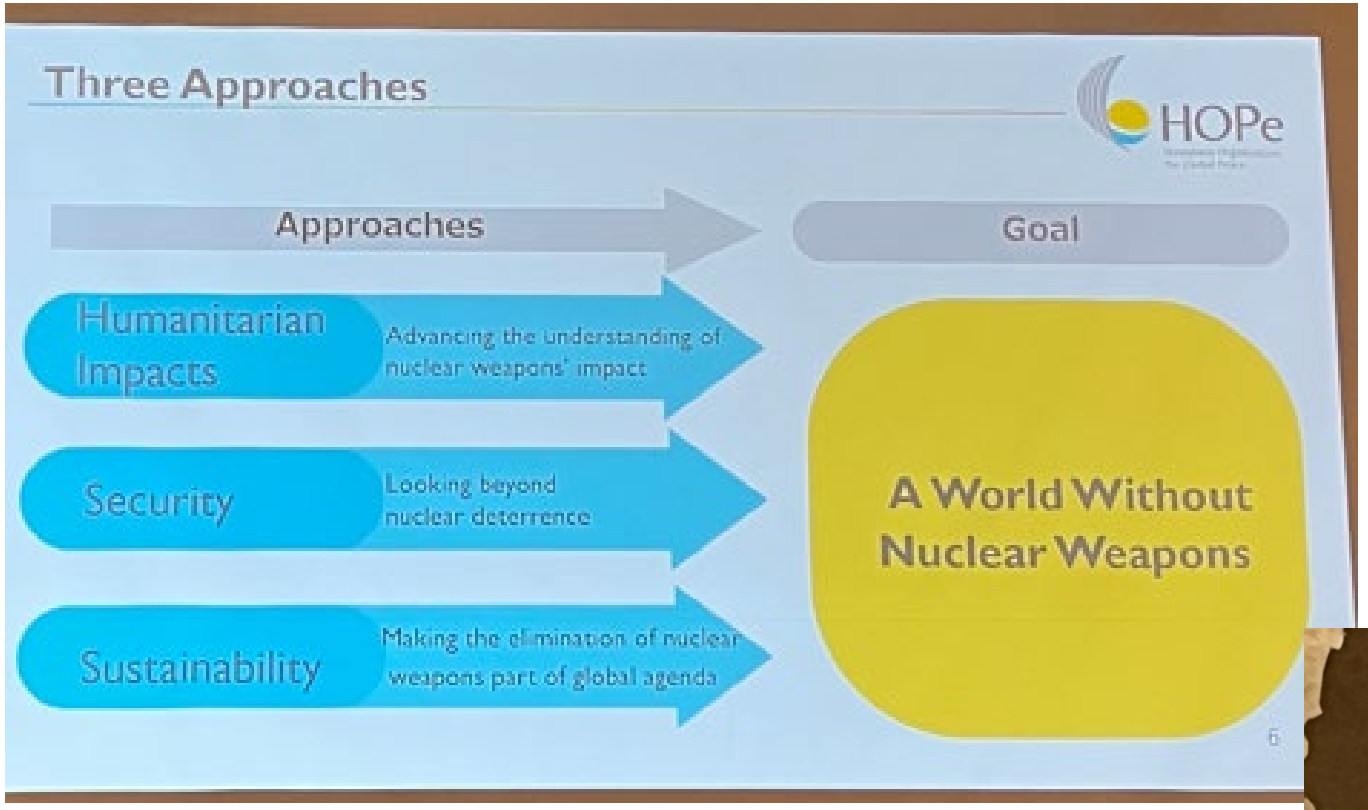
Museumsdirektør for Hiroshima memorial museum

Fra Hiroshima memorial museum

AALBORG UNIVERSITY

1. dag: Hiroshima og Pugwash

- Hiroshima: et symbol for fred. Morals chock -> Pugwash
- Horrors of nuclear must not be forgotten (can be described by science)
- Verdens ledere har ikke lært lektion fra Hiroshima
- Kommunen og amtet arbejder målrettet med at invitere ledere fra hele Verden til Hiroshima for lære af historien. Derfor er Museet blevet bygget.
- Understøtter også folkeoplysning: Hibakusha'er rejser rundt og fortæller om deres oplevelser.
- Vi mødte en 4. generation Hibakusha, som løfter arven.
- Pugwash repræsenterer "Science for Peace"
- Russell-Einstein manifested
- Science skal gavne menneskeheden. Skal bygge bro mellem forskellige interesser
- Værdier: Trust, understanding, collaboration, human dignity
- Remember your humanity and forget the rest
- At Pugwash we meet as humans face to face
- Promote dialogue. Bring together different viewpoints
- Turn responsibility into action ved at studere effekterne af kernevåben



To participate
in the Nobel Peace Prize ceremony
as younger generations

Hvordan afskaffer vi kernevåben?

- Tatsujiro Suzuki: Via Science-based policy-advice on nuclear weapons. Forskere har samtaler på tværs af politiske og nationale skæld. Viden er objektiv / fælles konstruktion og udgør et fælles grundlag.
- Der eroprettet en videnskabelig gruppe der skal rådgive om TPNW. FN har nedsat et panel om effekterne af atomvåbenkrig.
- Film om Sort Regn: Videnskabelig kontrovers om hvor den faldt. Kun anerkendte ofre kan få gratis lægehjælp og offentlig understøttelse.
- Ruth Mitchell: Kernevåben er blevet testet på oprindelige folks land. Fokus på ny teknologi – særligt AI og atomkrig. Vi skal arbejde imod det militær-industrielle kompleks.
- Tillman Ruff: Gennemgang af den internationale lovgivning om kernevåben.
- NATO-lande har anlæg der indgår i NATOs afskrækningspolitik.

Science without policy is science, but policy without science is gambling

— David Grey, International Institute for Applied Systems Analysis(IIASA)

Spirit of Pugwash

First Pugwash Conference held in Pugwash Village, Canada (1957)
https://pugwash.org/history/resources/

- **"Dialogue Across Divides"**
  - Scientists can speak over national borders
  - Exchange views and explore solutions based on science (and mutual interests)
- **"Social Responsibility of Scientists"**
  - Scientists are the ones who know potential impacts of science and thus should inform the policy makers and the public

Pugwash

Scientific Advice in Nuclear Disarmament

Scientific Advisory Group (SAG) of Treaty on the Prohibition of Nuclear Weapons (TPNW)

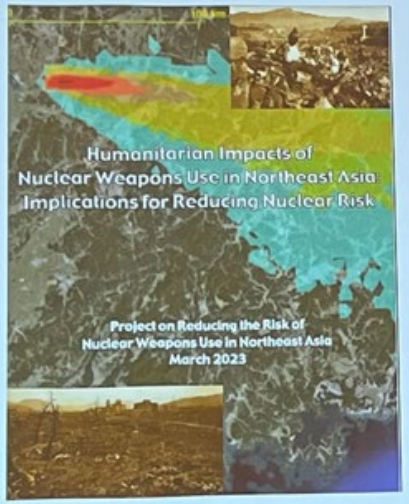
- Established in Feb. 2023 based on the decision at the First Meeting of States Parties
- 15 members (Chaired by Zia Mian and Patricia Lewis)
- Two reports ("Current status of nuclear weapons risks" and "humanitarian impacts of nuclear weapons") at the second MSP
- Tasked to make recommendations at the Review conference (involving other relevant stakeholders)

UN Independent Scientific Panel on Effects of Nuclear War

- The Panel was established pursuant to General Assembly resolution 79/238 in December 2024.
- SG Guterres appointed 21 members including Prof. Masao Tomonaga of Japan. He was appointed as co-deputy chair.
- Seven WGs will conduct research on *"examine the physical effects and social consequences of a nuclear war" including "climatic, environmental and radiological effects, and their impacts on public health, global socio-economic systems, agriculture and ecosystems"*

Humanitarian Impacts of Nuclear Weapons Use in Northeast Asia

by Research Center for Nuclear Weapons Abolition, Nagasaki University (RECNA), Nautilus Institute, APLN (2023)



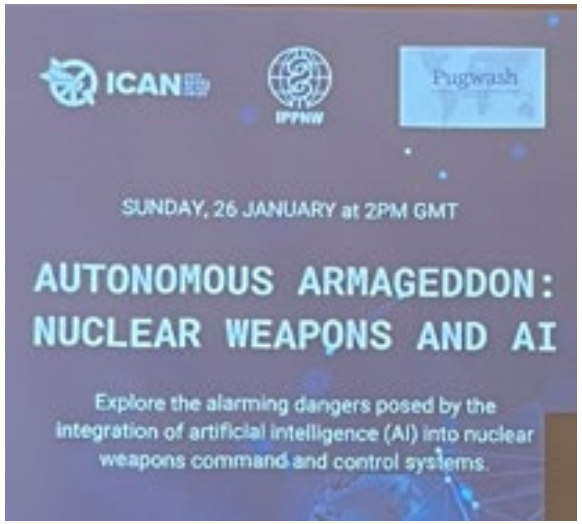
5 cases were chosen out of 30 cases (25 plus 5 new cases after the Ukraine War) for quantification of the impacts.

Quantify only physical impacts (short term and long term deaths, radioactive cloud and possible radiation dose) of nuclear weapons use.

Draw possible lessons from the humanitarian impacts of nuclear weapons use in the region for policy recommendations

https://www.recna.nagasaki-u.ac.jp/recna/bd/files/Year_2_NU-NEA_Report_E_2303

Pugwash



What is the novel risk?

- The risk is autonomy
- Little or no oversight by humans
- The accountability gap challenges international humanitarian law

Human oversight is of critical importance

- Russia has a new underwater vehicle, the Poseidon submarine: autonomous, nuclear powered, nuclear tipped
- China has displayed a similar looking underwater vehicle in Tiananmen Square

Complex systems are unpredictable

- AI is not adaptable in the ways that we are
- We cannot guarantee outcomes/outputs

AI increases the speed, scale and deadliness of war

- Gospel
- Lavender

- With nuclear weapons we have to exclude AI from the command and control chain
- Global Commission for the Responsible Use of AI in the Military
- Rome Call for AI Ethics

- BUT this is not enough – the pervasive nature of AI must increase our resolve to eliminate nuclear weapons. Because nobody should be in control of the worst weapons of mass destruction.

Open Letter in support of the Treaty on the Prohibition of Nuclear Weapons

"We must show courage and boldness - and join the treaty. As states parties, we could remain in alliances with nuclear-armed states, as nothing in the treaty itself nor in our respective defence pacts precludes that."

- 56 former heads of government, foreign and defence ministers of nuclear-allied nations, including:
  - 20 NATO members
  - Ban Ki-moon
  - 2 former NATO secretary-generals (Willy Claes, Javier Solana)
  - Hatoyama Yukio, Tanaka Makiko & Naoki

Joint Humanitarian Statement
NPT RevCon 22 – 147 yes

"It is in the interest of the very survival of humanity that nuclear weapons are never used again, under any circumstances."

UNGA 2024: 142 Yes incl Japan, India, Greece, Philippines, Iran, Belarus

NATO
27 Sep 2022

"any use of nuclear weapons is absolutely unacceptable"

Jens Stoltenberg, Secretary-General

International Court of Justice

Para 47: "The notions of "threat" and "use" of force under Article 2, paragraph 4, of the Charter stand together in the sense that if the use of force in a given case is illegal – for whatever reason – the threat to use such force will likewise be illegal."

INTERNATIONAL COURT OF JUSTICE

REPORTS OF JUDGMENTS, ADVISORY OPINIONS AND ORDERS

LEGALITY OF THE THREAT OR USE OF NUCLEAR WEAPONS

ADVISORY OPINION OF 8 JULY 1996

International Court of Justice

Re NPT Article 6 obligation:

264 THREAT OR USE OF NUCLEAR WEAPONS (ADVISORY OPINION)

The legal import of that obligation goes beyond that of a mere obligation of conduct; the obligation involved here is an obligation to achieve a precise result — nuclear disarmament in all its aspects — by adopting a particular course of conduct, namely, the pursuit of negotiations on the matter in good faith.

100. This twofold obligation to pursue and to conclude negotiations

No differentiation between states:

267 THREAT OR USE OF NUCLEAR WEAPONS (ADVISORY OPINION)

F. Unanimously,

There exists an obligation to pursue in good faith and bring to a conclusion negotiations leading to nuclear disarmament in all its aspects under strict and effective international control.

Indiscriminate inhumane weapons banned by international treaty

- [Expanding bullets 1899]
- Biological weapons 1972
- Chemical weapons 1993
- Landmines 1997
- Cluster munitions 2008

- The main basis for bans on all these has been humanitarian – unacceptable harm with any use
- Proven process: stigmatise - prohibit – eliminate
- No weapon has been controlled without being banned

Chancellor Scholz, President Xi

4 Nov 2022

- "President Xi and I agree: nuclear threats are irresponsible and incendiary," Scholz said after the meeting. "By using nuclear weapons, Russia would be crossing a line that the community of states has drawn together."

- Xi agreed that both leaders "jointly oppose the use or threat of use of nuclear

The obligation to disarm

Nuclear non-proliferation Treaty(NPT)

- 191 states parties
- entered into force 1970

Article VI

"Each of the Parties to the Treaty undertakes to pursue negotiations in good faith on effective measures relating to cessation of the nuclear arms race at an early date and to nuclear disarmament …. "

General Comment No. 36 (2018) on ICCPR Article 6: The right to life

"66. The threat or use of weapons of mass destruction, in particular **nuclear weapons**, which are indiscriminate in effect and are of a nature to cause destruction of human life on a catastrophic scale is **incompatible with respect for the right to life and may amount to a crime under international law**. …

They must also respect their international obligations to **pursue in good faith negotiations in order to achieve the aim of nuclear disarmament** under strict and effective international control and to **afford adequate reparation to victims** whose right to life has been or is being adversely affected by the testing or use of weapons of mass destruction, in accordance with principles of international responsibility.

67. Article 6 is included in the list of non-derogable rights …"

States without nuclear weapons

- Most act responsibly:
  - Foreswear NW, support disarmament:
    - join applicable treaties including NPT, CTBT, nuclear free zones (104)
    - Provide no justification for NW or threat and no hosting or assistance for their possible use
    - De-legitimise, stigmatise and ban nuclear weapons: join the TPNW (99)
- Some contribute to NW dangers:
  - with US
    - NATO (29), Australia, South Korea, Japan
  - with Russia
    - Armenia, Belarus
  - Kazakhstan and Marshall Islands - hosting test sites used as target for nuclear-capable ICBM launches
    - » Sary-Shagan missile range; Ronald Reagan range, Kwajalein Atoll

Nuclear "weasel" states

For undermining the NPT; obstructing efforts to ban nuclear weapons; general duplicity and insincere conduct; sundry offences against reason and logic.

Dag 2: The role of Scientists in Reducing Nuclear Threats

- Frank v. Hippel: Previous Successes. Will policy-makers listen? Depends on public opinion.
- Partial Nuclear Test Ban Treaty (1963), Involved Linus Pauling and Andrei Sakharov. Global Fallout Effects engaged the Public. An environmental rather than an arms-control treaty.
- Treaty Limiting Anti-ballistic Missiles (1972), Missile interceptors in Moscow and US cities. Had nuclear warheads. Surburbanities: Not in my Backyard! Congras began to listen. American scientists persuaded Russian scientists that ABM-systems were counter productive. Reagan revitalized it. Star Wars. Public was frightened. Chose nuclear disarmament.
- Comprehensive Test Ban Treaty. Gorbachev moratorium (1985). Onsite inspection in Kazakstan by US scientists. US voted to join Soviet moratorium. CTBT opened in 1996. Only DPRK has tested since then.

Russell-Einstein…Yukawa…Manifesto (1955)

"scientists should assemble in conference to appraise the perils that have arisen as a result of the development of weapons of mass destruction, and…discuss a resolution"

But will the politicians listen?

It depends:

If the public is engaged, then it is more likely.

At the moment, the public is concerned about other issues and the military-industrial complexes are hard at work.

That is one of our challenges.

My examples

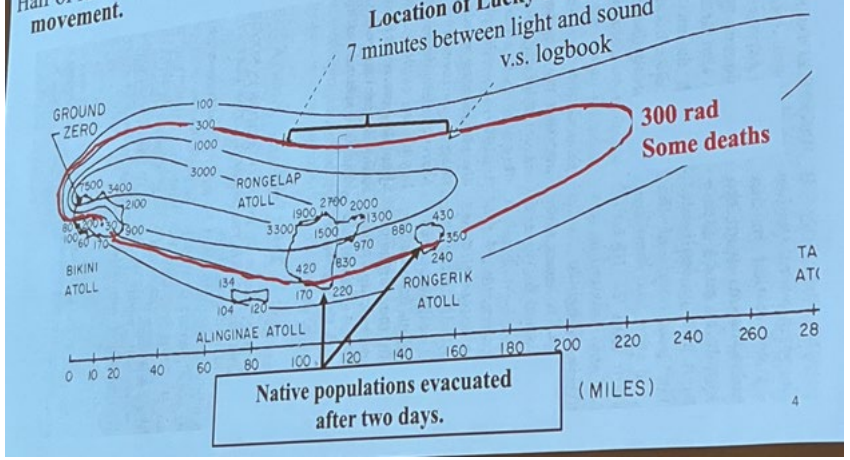
- Partial Test Ban Treaty of 1963
- The Anti-Ballistic Missile Treaty of 1972
- Comprehensive Test Ban Treaty of 1996
- INF Treaty of 1987, the START I Treaty of 1991…

Partial Nuclear Test Ban Treaty (1963)

1954: "Bravo," Bikini Atoll: **15 *million* tons TNT equivalent** (1000 Hiroshimas!),

Radioactive fallout on Japanese fishing boat, *Lucky Dragon 5*. One crew-member died.

Half of radioactivity went into global fallout. **Triggered a global anti-testing movement.**



The role of scientists

- Anton Khlopkov: Another example, Joint Comprehensive Plan of Action 2015 – JCPOA. Strengthen the global nuclear non-proliferation (Iran). Solutions proposed by a scientist and technical experts. E.g., use the facilities to produce stabile isotops. Will not be used for Uranium enrichment. Aftalen findes ikke mere.
- Open Science. Scientific advise should be flowing freely. More presidents of Scientific Bodies should be invited to Pugwash. Funding agencies, too. Same with individual scientists. Forskere skal skrive breve til politikerne.
- Minichiji Konuma: Var med til de Første Pugwash-konferencer. Var chorkeret over det der var sket I Hiroshima og Nagasaki. Eks. Doomsday Clock (Bulletin of the Atomic Scientists). The panel for study of the effects of nuclear weapons (chair – Ana Maria Cetto).
- How to mobilize young scientists to address the destructive side of science: Collaborate with Global Young Academy. Pugwash invite local young scientists to Pugwash meetings.
- How do scientists relate to Trump? Gaza. Academic boycott helped end apartheid. Sign calls. What about the assignations of scientists. Scientists must be protected. Se dokumentarfilmen The Lab (military research).

Origin of the solution: Zelenogosrk, Eastern Siberia

- Distance between Fordow and Zelenogorsk - ~5,500 kilometers
- Electrochemical Plant (JSC "PA "ECP"), Zelenogorsk (Krasnoyarsky Krai, Siberian Federal District); former name: Krasnoyarsk-45
- Industrial-scale gas centrifuge enrichment plant (since 1964)
- Stable isotope production facility (since October 1971)
- Production of 95 stable isotopes of 19 chemical elements (C, Si, S, Ar, Fe, Ni, Zn, Ge, Se, Kr, Mo, Cd, Te, Xe, Ir, W, Os, Pb)

7

Center for Energy and Security Studies

Multiple technically-intense issues to resolve as part of the JCPOA negotiations

- Uranium enrichment capacity (Natanz, Fordow, R&D)
- LEU stocks
- The Arak heavy water reactor
- Heavy water stocks

The Salehi-Moniz framework was established to address technical issues during the negotiations

1 Brief self-introduction

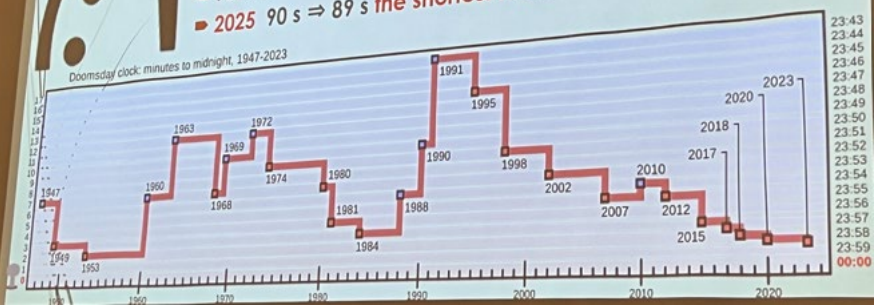
3

- 1931 born in Tokyo theoretical physicist
- 1957 August I followed Sin-itiro Tomonaga to meet Joseph Rotblat in Tokyo and heard directly his report on the 1st Pugwash Conference.
- a founding member of "Pugwash Japan" (then a graduate student).
- The first task: discussions on the agenda of the 2nd Conference. We sent "A statement of 21 Japanese physicists who support the Pugwash Conferences" to the 2nd Conference (Lac Beauport, Canada)
- The second task: Tomonaga organized weekly meetings to study the Minutes and Documents of the 2nd and 3rd Pugwash Conferences. (311 pages and 295 pages)
- I attended first the 14th Pugwash Conference (Venice, Italy, 1965).
- Pugwash Council member from 1992 to 2002

Bulletin of the Atomic Scientists Doomsday Clock

10

- 1947~ The start was 7 minutes to the midnight which corresponds to the nuclear apocalypse of the earth.
- The clock shifts forward for increasing nuclear danger, and backward for decreasing.
- Forward shifts 18 times and backward shifts 8 times.
- 2025 90 s ⇒ 89 s the shortest to the midnight.



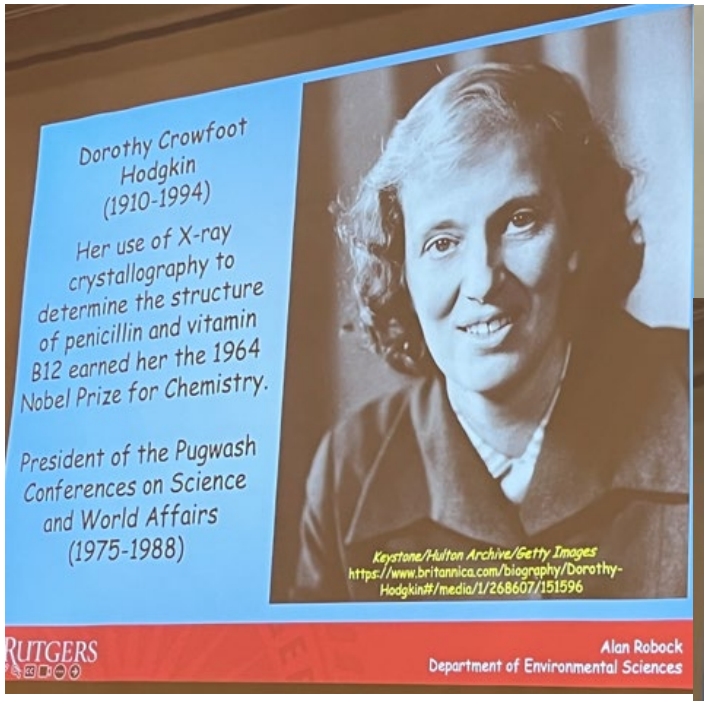
8 times of reducing nuclear threats

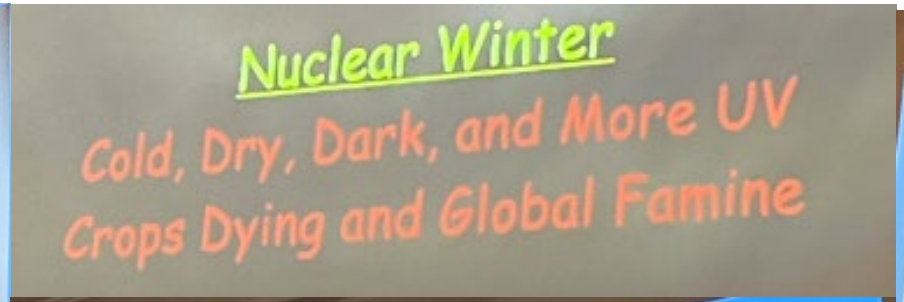
11

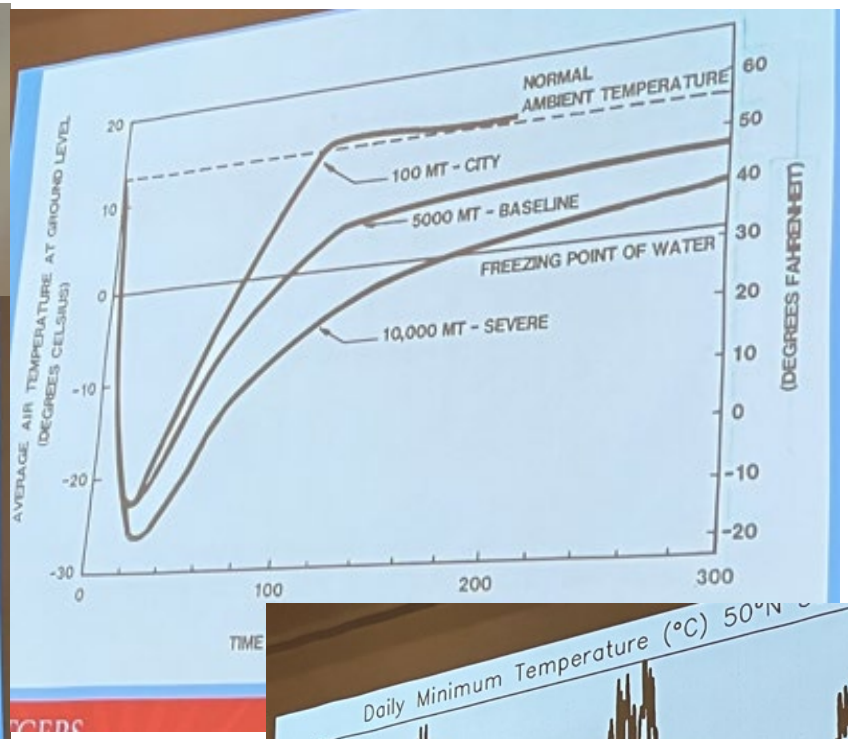
- 1960 International Geographic Year (IGY); Pugwash Conf.
- 1963 Partial Test Ban Treaty (PTBT); Cuban Missile Crisis
- 1969 Treaty on the Non-Proliferation of Nuclear Weapons (NPT) 1972 Strategic Arms Limitation Treaty (SALT); Anti-Ballistic Missile Treaty (ABM)
- 1988 Intermediate-Range Nuclear Forces Treaty (INF)
- 1990 Fall Berlin Wall; End of the Cold War
- 1991 Strategic of the Arms Reduction Treaty (SALT)
- 2010 UN Climate Chage Conference in Copenhagen
- 2018 Treaty on the Prohibition of Nuclear Weapons (TPNW); misuse of information technology; misuse of social media (2 ½ minutes ⇒ 2 minutes)

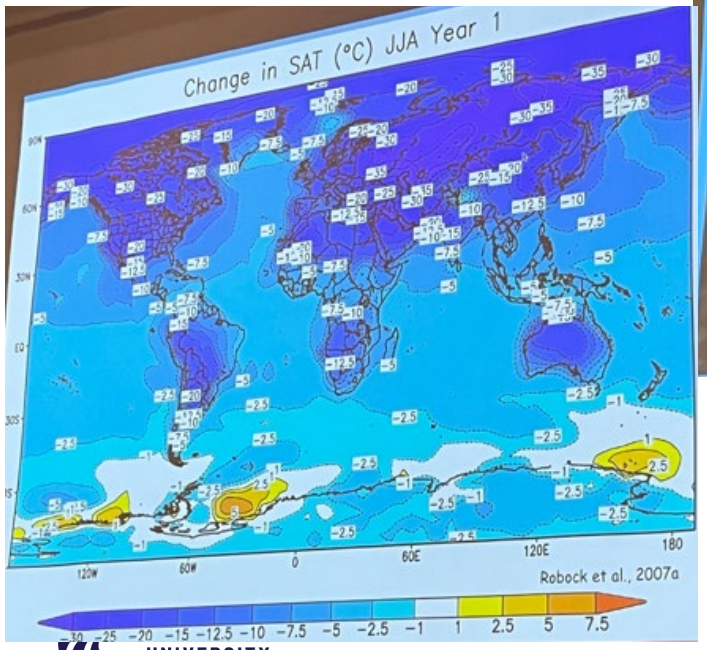
Dorothy Hodkins Lecture: Preventing Global Famine after Nuclear War

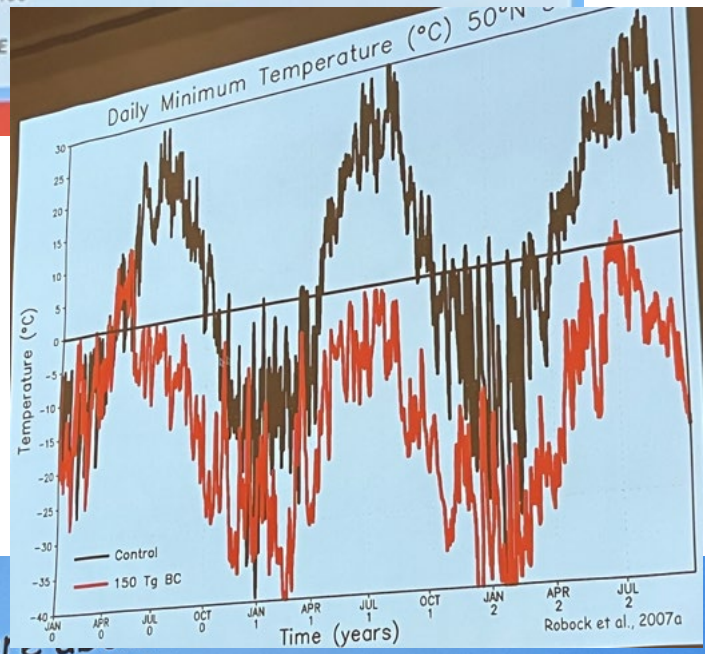
- Alan Robock: om Nuclear winter theory.
- Påvirkede også politik -> Ingen vindere af atomkrig -> alt liv forsvinder på grund af nuclear winter.
- Afskrækkelse er en myte. Kaster kun den ene side en atombombe, så dør de selv p.g.a. nuclear winter. Afskrækker ikke konventionel krig.
- Noget af soden fra brændene efter en atombombesprængning skylles ud som sort regn. Andet skydes op i stratosfæren og giver anledning til nuclear-vinter. Bliver i stratosfæren i flere år.
- Klimamodeller bruges til at beregne temperaturen på jorden efter atomkrig. -> Klimaforandringer.
- Jordskælv. Skovbrænde. WWII. Vulkaner. Fortæller hvor soden ryger hen.
- Klima ændres (mørke, mindre nedbør, kemikalier i atmosfæren...). Landbrugsproduktionen falder drastisk. Fødevareflygtninge.
- Pugwash vil gerne bringe flere forskere og evidensbaserede perspektiver ind i organisationen. Også samfundsøkonomi.
- Husk også at arbejde med hvordan radioaktiviteten bevæger sig hen.











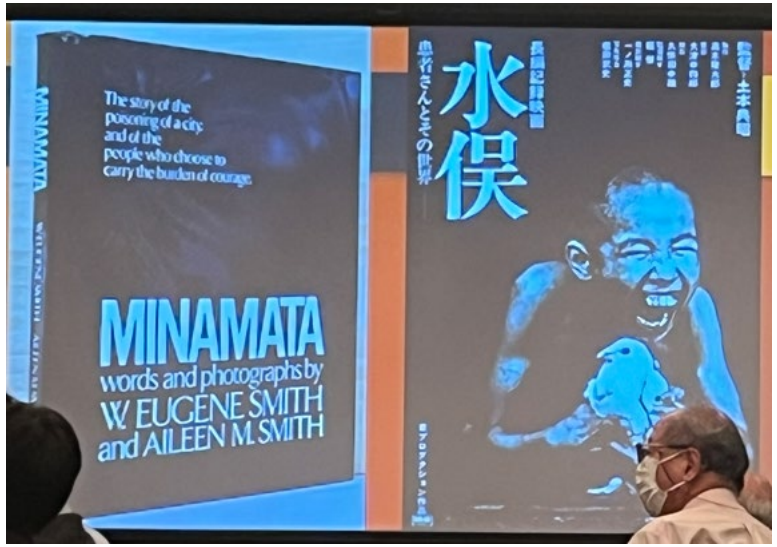
For more about

https://climate.envsci.rutgers.edu/nuclear/

UNIVERSITY

Briefing on Fukushima + suggestion for a treaty to protect from the air

- There is a controversy in Japan regarding increased thyroid cancer occurrence in Fukushima is caused by low radiation from the accident.
- How is the controversy portrayed in mainstream media?



The Future of the European Security Architecture

- Thomas Countryman (Arms Control Association):
  - Sikkerheden afhænger af om Rusland tager land fra Ukraine: Ja – den stærkest ret -baseret sikkerhedsarkitektur. NATO-tropper i grænsestater. Vanskeligt at lave aftaler om konventionelle våben = oprustning. Incitament til at anskaffe sig kernevåben. Selv Europæiske lande vil begynde at overveje at anskaffe sig kernevåben. Under fransk kernevåben paraply.
- Pavlo Klimkin (Ukraine):
  - Rusland ser sig ikke længere som del af Europa. Arkitekturen er ændret. I 2013 var det ingen planer om, Ukraine skulle være en del af NATO – kun aftale med EU. Krigen handler om identitet: Er Ukraine russisk eller ukrainsk. Det handler om territorium. Rusland siger de vil følge Folkeretten, men har en anden fortolkning af folkeretten end EU. Ukrainsk sikkerhed = Europæisk sikkerhed. Ny sikkerhedsarkitektur handler om afskrækning (flere våben) og resiliens (klare uforudsete kriser). Mere militær-teknologi.

- Sergey Oznobshchev (Russia):
  - Sikkerhedsarkitekturen skal bygges på folkeretten. Tager udgangspunkt i national sikkerhed. Man taler sammen og finder et kompromis. Det vil så ende i aftaler. Rusland er en del af Europa – historisk, geografisk. Det er NATO-udvidelser der har smadret den europæiske sikkerhedsarkitektur. Efter Warszawa-pagten Faldt var der ingen grund til at opretholde en NATO-blok. Stort skepsis i det russiske parlament vedr. NATO. Der er ingen Russisk – Europæisk dialog. Der er dialog mellem Rusland og USA. Den vil definere den Europæiske sikkerhedsarkitektur.
  - Kommentarer – lande selv bestemmer om de vil være medlemmer af NATO. Sikkerhedsarkitekturen vil ikke blive bestemt af Rusland og USA uden EU..
- Wolfgang Richter (Germany):
  - Kan vi revitalisere våben-kontrol i Europa? Selv efter krigen i Ukraine afsluttes vil Europa føle sig truet af Rusland. En kold fred. Mange handlinger vil blive tolket som en provokation -> kan føre til eskalation og konfrontation. Afskrækning skal kombineres med dialog. Europa vil fortsat være i tvivl om russiske intentioner -> fortsat oprustning. Territorial integritet skal respekteres i en fremtidig aftale. Måske kan vi lære af Cuba-krisen, hvor afskrækning blev kombineret med kombineret med dialog. Risiko reducerende tiltag skal også igangsættes: Militær transperens. Ansvarlighed i retorik. Europa skal begynde at tale med Rusland.
  - Kommenter – ingen fejl i traktaterne. Det er overholdelsen det kniber med.

President's address

- **Memory → Responsibility:** Hiroshima reminds us to act now.
- **Problem:** Arms control eroding; arsenals growing; flashpoints multiplying; dialogue fading.
- **Reality:** Deterrence views persist; NPT under strain; trust low.
- **Compounding risks:** Climate, pandemics, inequality → instability.
- **Pugwash Response:** Create safe dialogue, reduce risks, rebuild norms, train the next generation in ethics, link security to broader global challenges.
- **Duty:** Keep the vision—and the venue for progress—alive.

Secretary General report

CHARACTERISTICS OF PUGWASH ACTIVITIES:

- Open and informal participation
- Personal capacity
- Confidential dialogue
- Expertise and experience
- Diverse representation

THE PUGWASH CONFERENCES ON SCIENCE AND WORLD AFFAIRS

MISSION

To bring scientific insight and reason to bear on threats to human security arising from science and technology in general, and above all from the catastrophic threat posed to humanity by nuclear and other weapons of mass destruction.

Pugwash above all is an expression of the social and moral duty of scientists to promote the beneficial applications of their work.

DESIRABLE OUTCOMES:

- Peaceful Resolution and Disarmament
- Reduced Global Risks
- Strengthened Multilateralism
- Science with Ethics and Responsibility
- Empowered Future Generations

GOVERNANCE, FINANCING AND STRUCTURE:

- Bylaws (2025)
- Pugwash Foundation in Geneva
- Thematic organization, Working Groups:
  WG on Nuclear Weapons
  WG on BCW
  WG on Emerging Technologies
  WG on the Middle East
- National and regional groups
- ISYP

OFFICERS

President: Dr. Hussain Al-Shahristani (Iraq)
Secretary General: Prof. Karen Hallberg (Argentina)
Council Chair: Prof. Götz Neuneck (Germany)

Program Director:
Dr. Poul Erik Christiansen (UK)

MEMBERS OF THE PUGWASH COUNCIL

Amb. Taous Feroukhi (Algeria)
Prof. Karen Hallberg (Argentina)
Dr. Jean Pascal Zanders (Belgium)
Dr. Jennifer Simons (Canada)
Mr. Cesar Jaramillo (Canada)
Dr. Tariq Rauf (Canada/Pakistan)
Prof. Li Bin (China)
Ms. Ana Jerković (Croatia)
Amb. Khaled Shamaa (Egypt)
Dr. Katariina Simonen (Finland)
Prof. Venance Journé (France)
Prof. Annick Suzor-Weiner (France)
Prof. Götz Neuneck (Germany)
Dr. Happymon Jacob (India)
Prof. Amitabh Mattoo (India)
Amb. Reza Ziaran (Iran)
Dr. Ali Salehi (Iran)
Dr. Hussain Al-Shahristani (Iraq)
Prof. David Menashri (Israel)
Amb. Jeremy Issacharoff (Israel)
Prof. Francesco Calogero (Italy)
Prof. Paolo Cotta-Ramusino (Italy)
Prof. Francesco Lenci (Italy)
Prof. Tatsujiro Suzuki (Japan)
Prof. Tomohiro Inagaki (Japan)
Dr. Mark Suh (Korea/Germany)
Mr. Eldar Mamedov (Latvia)
Amb. Jan Hoekema (Netherlands)
Dr. Moeed Yusuf (Pakistan)
Ms. Taghreed El-Khodary (Palestine)
Amb. Sergey Batsanov (Russia)
Prof. Alexander Nikitin (Russia)
Prof. Joelien Pretorius (South Africa)
Prof. Thomas Jonter (Sweden)
Mr. Arnold Luethold (Switzerland)
Prof. Mustafa Kibaroğlu (Türkiye)
Dr. David Ellwood (UK)
Dr. Cliff Kupchan (USA)
Dr. Steven Miller (USA)
Amb. Robin Raphel (USA)

Secretary General's Activities 2025
January – May

- Concluded the Nuclear Risk Reduction Dialogue series in Istanbul
- Co-hosted the "Autonomous Armageddon" webinar on AI and nuclear risks with other Nobel Peace Prize laureates.
- Participated in major international disarmament meetings, including the TPNW Meeting of States Parties and the NPT PrepCom in New York.
- Delivered the Max-von-Laue Lecture in Bonn (DFG)
- Led a delegation to Tehran for regional dialogue.

June – October

- Conducted official visits to China and Hiroshima, promoting the upcoming Pugwash Conference.
- Co-organized the 70th Anniversary of the Russell–Einstein Manifesto (with BP)
- Co-organized the Nobel Assembly on the Prevention of Nuclear War in Chicago ——> Declaration
- Engaged in global peace and disarmament dialogues in Erice, Beijing, Nagasaki, Moscow, New York, and Dubai, emphasizing scientific responsibility, nuclear risk reduction, and multilateral cooperation

November

The 63rd Pugwash Conference
80 Years after the Atomic Bombing:
Time for Peace, Dialogue and Nuclear Disarmament
Hiroshima, Japan

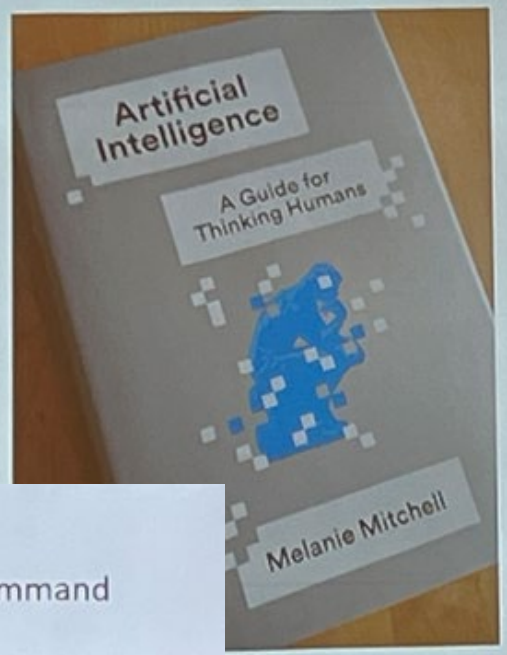
Disruptive technologies and nucelar weapons

- Herb Lin, Standford. AI og kernevåben. Introducerer AI: garage in garbage out. Reproducerer det der er lært. …
- Siger lidt om kernevåben systemer. … Kontrol af kernevåben: dag-til-dag. Episoder.
- Pointer: mennesker skal kontrollere kernevåben. Hvad betyder det?
- Man skal se AI i kernevåben i kontext. NC3 er ikke isoleret fra ROW.
- Internationale aftaler om AI I NC2 er ikke realistisk.
- Ricisi kan mitigeres (nedbringes - håndteres). Træk på kommercielle erfaringer. Strawman assessment.
- Hvad AI ikke kan.

Q&A to Herb

- Responsible AI in military application = people engaged in the topic in conversation. Ved ikke om hans principper følges i praksis (i USA).
- AI er ikke dårligt, men den skal kontrolles. Også når man burger ChatGPT.
- Skal alle lande engagere sig i Risico-mitigation? Hvert land kan gøre noget.
- Kan man ikke se ubåde? Fog of war. Det vil man nok kunne se dem bedre. Men ikke i morgen. Og tågen forsvinder aldrig helt.
- Global AI governance – initiative? Ja, de involverede parter skal tale sammen.
- Er house of dynamites – realistisk? Mennesket skal fortsat tage belsutningen.
- AI benyttes også i forskning og udvikle kernevåben. Bliver det studeret? Ja. Kan sende noget.

Background



Machine learning is the current paradigm for AI (1990s)

- Use of large amounts of data to enable computers to perform tasks without explicit instructions by generalizing (or "learning") from patterns in data with minimal human intervention.
- Multi-layer neural networks, deep learning, big data, compute-intensive
- Supervised, unsupervised, reinforcement learning
- Generative vs predictive
  - Generative: novel content
  - Predictive: forecast future outcomes
- Foundation models, large language models, chatbots
- non-uniform progress: LLM fast compared to speech recognition; computer vision; autonomous cars.

On operational risk

- Clearly benign: Let's use spell-check for emails at US Strategic Command
- Clearly bad: Let's give ChatGPT the launch codes
- In between these extremes (the clearly ridiculous to the clearly benign), there's a very large range, and to understand the proper relationship of AI to the nuclear enterprise, we need detailed definitions and scope for:
  - "AI"
  - "The nuclear enterprise"
- The intellectually thorny issues are concealed within these definitions.
- AI and the nuclear enterprise: MUCH broader than AI and launch decisions
- BLUF
  - Many applications of AI to the nuclear enterprise are either **benign or enhance** human control over nuclear weapons.
  - A narrow set of AI applications to the nuclear enterprise are **foolish, unwise, potentially catastrophic, and should be avoided at all costs.**

Key points to remember about machine learning

- Machine learning is still a way of programming computers
- "Garbage in, garbage out" still holds.
  - Bad training data yields bad outputs.
- ML doesn't enable you to violate the laws of physics.
  - AI can't create new facts where none existed before.
  - Insight regarding known facts isn't creating new facts
  - Can't address unknown unknowns problem any better than people can.
- ML internal operations are in general incomprehensible to humans.
  - Conventional programming involves the explicit coding of instructions
  - ML programming is not explicit; hence reasoning must be inferred
- Machine learning is at its root statistics, "correlation is not causation."
  - Statistical explanations can be useful, but
  - Sometimes mechanistic/causal explanations needed
  - Machine learning as statistical not smart.

Elements of the nuclear enterprise

- NW stockpile program
  - Design, production, and certification of modernized weapons, maintenance and assessment of all stockpile NWs
- Nuclear delivery systems and platforms
  - DVs: ballistic missiles, cruise missiles, gravity bombs
  - Platforms: silos, submarines, delivery aircraft
- The nuclear command and control system
  - NC2 is people, processes, policies for:
  - NC3 is technology infrastructure for NC2 (comms, computing)
  - Note: different functions within NC2 mean different ways of using AI within NC2.
    - Different kinds of AI for different applications. "AI" is not just one thing
- Computers are used everywhere; hence AI is usable everywhere.
- Just adding AI doesn't necessarily improve or harm the application.

Nuclear command and control decomposed (first level)

Day-to-day

- Force management: maintenance of continuous control over nuclear forces during peacetime and crises. Includes nuclear surety.
- Situation monitoring: real-time awareness of adversary actions, global events, and force readiness 24/7/365.
- Planning: development and update of operational plans for nuclear weapons deployment or use.

Episodic

- Decision-making: assessment, review, and consultation regarding use or movement of nuclear weapons.
- Force direction: implementation (preparation, dissemination, and authentication) of decisions regarding the execution, termination, destruction, and disablement of nuclear weapons.

1- Humans: the most essential element of nuclear command and control.

- "Human in the loop" - Many nations have made commitment in some form
- What is the definition of "in the loop"?
  - What if POTUS is being informed only by screens driven by GPT?
  - Intended scope of "in the loop"? - yes, ultimate decision of consequence (launch decision)
    - assessment of early warning information?
    - making recommendations for targeting or military operations using nuclear weapons?
- "meaningful human control" or "appropriate levels of human judgment" better?
- Political commitment to "maintain a human in the loop for all actions critical to informing and executing decisions by the President to initiate or terminate nuclear weapons employment" does not ensure "appropriate levels of human judgment" (cf., 3000.09) will be applied to such decisions.

The discussion of what "meaningful human control" means is critical.

2 - AI and nuclear nexus cannot be assessed in isolation

- AI/nuclear includes AI in
  - force structure (e.g., are forces survivable);
  - arrangements and infrastructure for NC2; (e.g., predelegation?)
  - doctrine (e.g., counterforce or minimum deterrence)
  - strategic priorities (e.g., preemption OK?)
- NC3 is not and has never been completely isolated from ROW
  - AI in conventional forces
    - Escalation pathways
    - Entanglement of C3 and NC3
  - AI in BMD/ASAT
  - AI in commercial assets (e.g., satellites) integrated into NC3
  - AI in critical infrastructure supporting nuclear forces
  - AI impact on nuclear decision makers including corrupted information feeds

∴ NC2 cannot be isolated from AI failures elsewhere

Possible example of concern of AI-driven escalation

- ASAT activities may be highly escalatory
- AI support to US ASAT activities
  - No political commitments to put "human in the loop" for ASAT
- SM-3 missile used from AEGIS destroyer in an ASAT mode Feb 2008.
  - Unmodified missile; changed software and added instrumentation
- AEGIS has highly automated capabilities
  - Currently integrates air and ballstic missile defense capabilities
- Lockheed/Martin currently integrating AI into AEGIS
  - https://www.lockheedmartin.com/en-us/news/features/2023/artificial-intelligence-and-aegis-the-future-is-here.html

Not hard to imagine that a Chinese early warning satellite interprets an AI-driven AEGIS SM-3 launch as being for ASAT purposes

3- International agreements on AI in NC2 unlikely

- Achieving verifiable agreements on AI in NC2 unlikely due to
  - Civilian-driven efforts
  - Unverifiability
  - Disagreements about what counts as AI
    - Russian view and Western view of AI very different
- May be possible to
  - Share concerns
  - Educate each other about AI vs AGI, for example
  - Obtain commitments in principle to put humans "in charge"
  - Increase timelines for decision making by collateral limits
    - Longer timelines may reduce pressures to use AI
    - Bans on ASAT, depressed trajectory SLBMs, space-to-surface weapons may increase timelines

4 - Nations can mitigate risks from their own AI use in nuclear contexts

- When the stakes involved are relatively low
  - Gain experience in low-stakes environments where mistakes don't matter very much.
  - Example: AI for predictive maintenance
- When the commercial world is doing it too
  - If commercial AI efforts are a close analog, military efforts can build on them.
  - Commercial has good metrics of success.
  - Civilian-only efforts are not well-matched to contribute to military-specific applications.
  - Example: AI for route optimization
- When a human has adequate time to review the output of the AI
  - Example: AI for pre-planned target selection vs AI for real-time target selection
- When outputs can be judged on ground truth in principle if not in practice
  - Assessments of intent vs assessments of fact
- When AI-enabled functionality can be separated from the remainder of the system
  - Requires ability to turn AI portions off
- When mechanisms are available to mitigate the worst consequences of failure
  - High consequence conditions must be recognizable but few in number
  - Need independent monitor to intervene

Strawman assessment for NC2

	Force mgmt (6)	Situation monitoring (~4)	Planning (5)	Decision-making (1)	Force direction (2)
When the stakes involved are relatively low	Yes	No	Yes	No	No
When the commercial world is doing it too	Yes	Yes	Yes	Yes, but	Yes
When a human has adequate time to review the output of the AI	Yes	Yes	Yes	No	No
When outputs can be judged on ground truth	Yes	Usually	Sometimes	No	Yes
When AI-enabled functionality can be separated from the remainder of the system	Yes	Depends	Sometimes	No	No
When mechanisms are available to mitigate the worst consequences of failure	Yes	Depends	Yes	No	No

Who should pay attention?

- Policy makers should pay attention to be informed about general perspectives on the topic of AI and nuclear weapons
- Program managers should pay attention to assess AI risk and benefit for programs they manage
- Policy makers and program managers have very different perspectives
  - Much of AI should be managed at the policy level but it is more likely to be introduced at the program level.

Knowing what AI can't do

- AI cannot fix the "test tape" problem.
  - Knowing that it was a test tape would require knowledge available only from outside the NC2 system.
- AI cannot fix the "fog of war" problem
  - Sensors provide more data
  - Data may be wrong/falsified
  - More data, even if good, entails more computational complexity
  - Solutions become intractable very fast with amount of data ingested

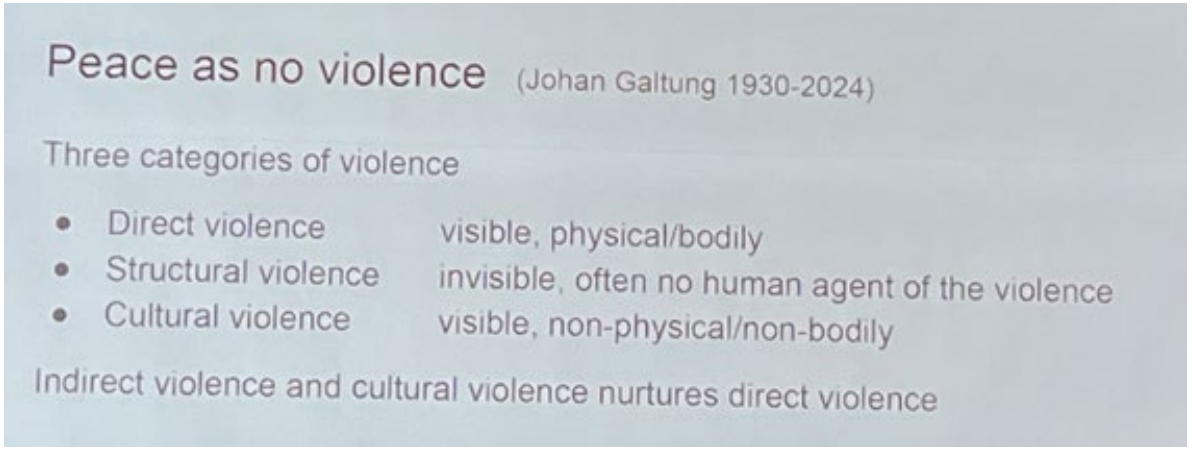
The cynic's view of AI

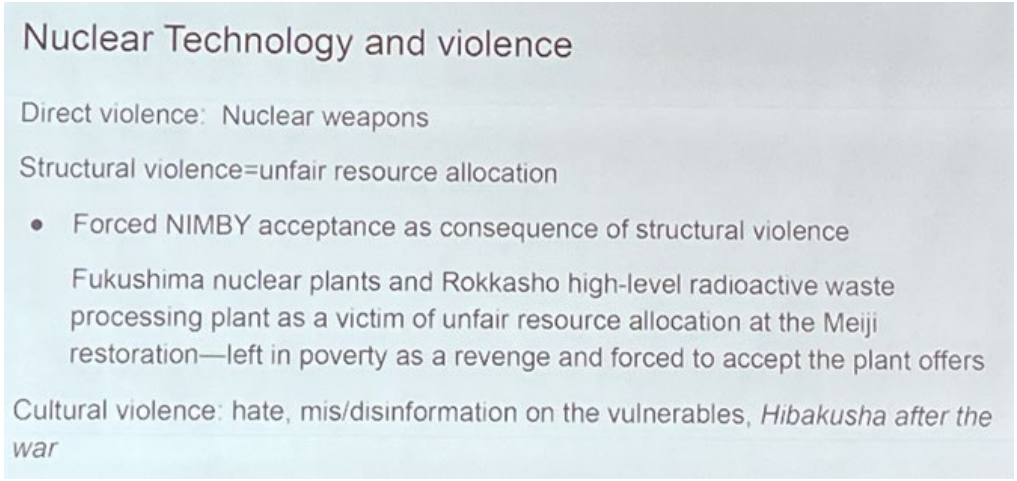
AI is what you call something that doesn't quite work.

- It works some of the time, maybe most of the time, but not all the time.
- You don't understand very well how it does work.
- You can't predict when it won't work
  - Corollary: you should not have high confidence that it does work.
- When it doesn't work, you often can't tell that it's not working.
- Even when you know it's not working,
  - you can't understand why it didn't work, and
  - you don't know how to fix it or make it work better.

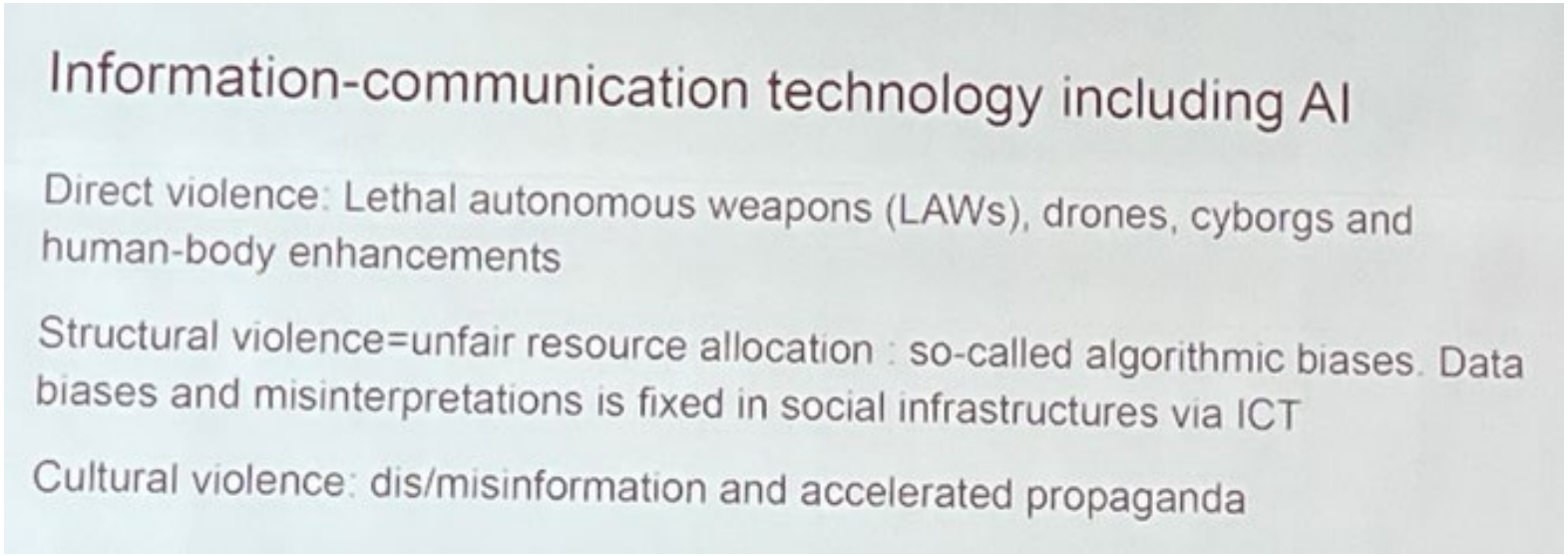
If these characteristics are OK, and they often are, then proceed.

If not, then more caution is advisable.





- Yoko Murakami: AI, nuclear tech and humanity risk. Peace as no violence. 3 slags vold. Erindring og reflexion vigtigt for at imødekomme vold. Er big science. Udøver AI og kernevåben: Ja alle tre typer.
- Der mangler etiske perspektiver i den professionelle træning af ICT exporter. Fremme Åben Source, fx Ubuntu.
- Pugwash kan hjælpe AI samfundet til at revidere deres forståelseshorizont til at reflektere og erindre.



Remembrance and reflections

We human beings learn from experiences, where experiences are from what was not expected nor planned.

Remembrance of the violent events and situations should not involve any exclusions or unfairness, as such injustices leads another structural violence.

Cycle of violence must be stopped.

- Remembrance of the history may control cultural violence.
- Reflections of the history may adjust structural violence.

Structural violence of "cognitive injustice" to oppress remembrance and reflections: deprivation of knowledge (interference of access to education)

ICT excluded from remembrance and responsibility

Information technology, esp. data science, has not been accused after holocaust, although it is the very technology behind the catastrophe.

No data scientists remember their responsibilities, because at that time computing power was poor. But now it equips big data, hardware power, and improved algorithms.

ICT-AI community itself is placed under structural violence. It is a background of cultural violence (fake news and propaganda) being rampant all over the information sphere now.

Easy shift from cultural violence to structural violence via data science

Automation may realize structural violence in the form of direct violence

Ubuntu

Computer tech community has recognized the American military-industrial complex issues since its early period.

1986- proposal of open source software distributions including Linux OS,

With the slogan "Ubuntu" after Zulu tradition: *we are for the others*

https://ubuntu.com/

Ubuntu against polarity of human and fragmentation

Fragmentation from national interests

- Needs of "us" often trespass their life and properties
- "They" excluded from us; their humanities are just ignored or not recognized

So there are competing cultures in tech: ubuntu and national interests,

BUT the dichotomy is being blurred: open software and cheap devices for drone attacks and home-made weapons to transform power structures among competing parties

Democracy of science gravitates violence

ICT-AI under cognitive injustice (possibly intentionally)

Self-regulation has not work— unlike Asiloma gene tech case and Pugwash on nuclear technologies

EU AI laws reflects the history of holocaust, but not other regions

Physical AI and generative AI will facilitate homemade AI with physical interference to the real world

- automated weapons soon as well as weapons in the space (or wherever hardly accessible for human)

Educational control does not fully work in ICT

- Professional training of ICT usually lacks ethics and/or compliance or even those without formal training can be ICT professionals

Proposal: AI supporting, and supported by, citizens— As well as Pugwash community

Research and development of AI and other big sciences are currently doomed to be sponsored by each nations and their military-industrial complexes. They have to maximize political and economical benefits of the nations, without consideration of the vulnerables, especially those who suffer from the consequences of those technologies

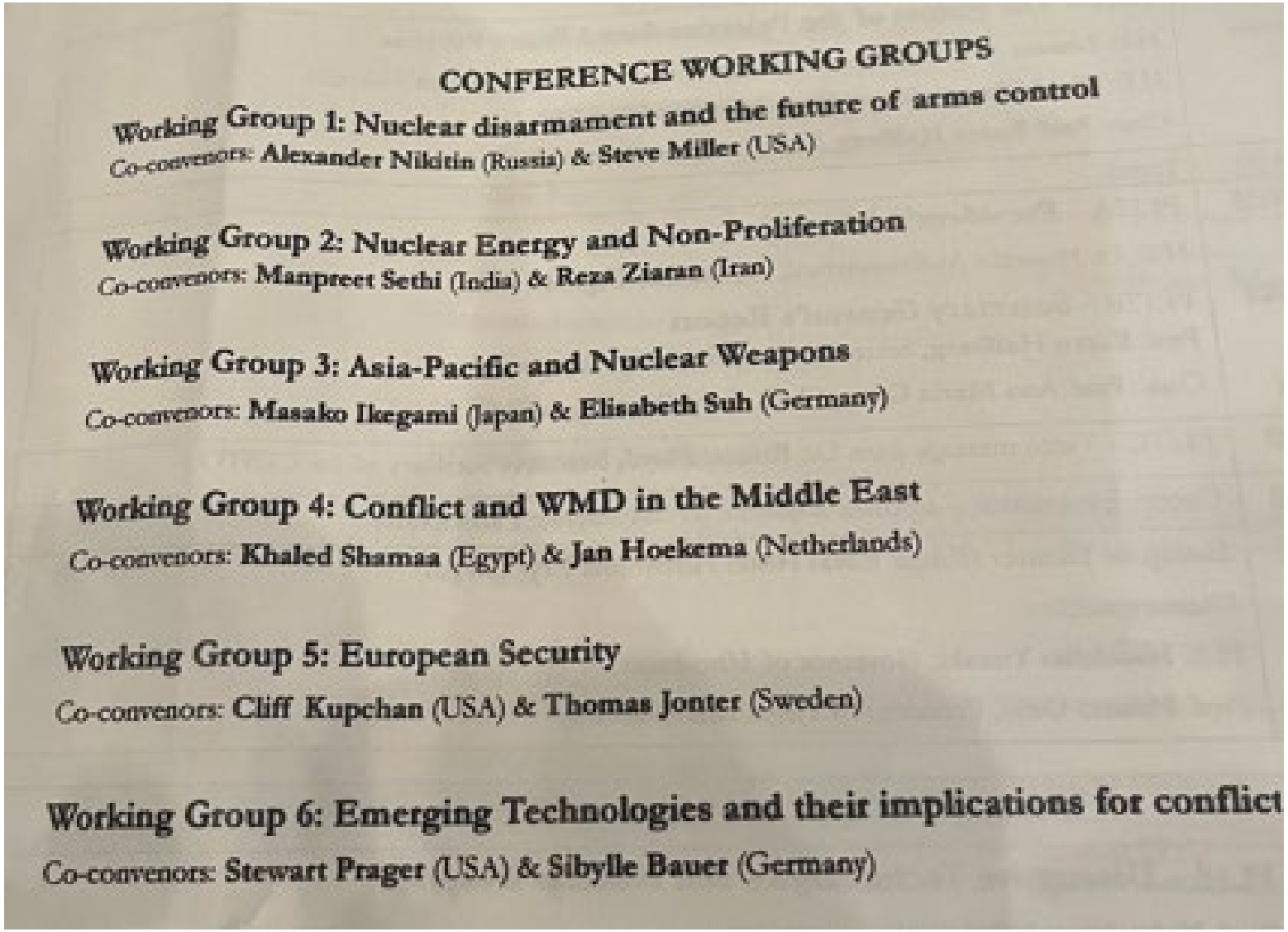
Supports for ICT community should involve reflection on the history— remembrance and memory are essential.

PUGWASH COMMUNITY CAN SUPPORT AI COMMUNITY TO CHANGE THEIR ATTITUDES, promoting dialogue among diverse cultures.

- Paul Meyer: Da kernevåben blev opfundet var de disruptive. Nu reguleres de af internationale aftaler. Men de er under pres. Andre teknologier påvirker også kernevåben og kan destabilisere (AI, cyber, rumtek). Mulighederne for diplomati. Det er der meget lidt focus på. Mange arbejder med at integrere disse tre typer teknologier inden for de tre områder:
- Rumtek. Sputnik. Prevention of weapons in outer space. I 50'erne. Offensive våben er ikke tilladt i rummet. Ja, her er der en traktat. Space security. Verifikations mekanismer.
- Cyber security: FN har overvejet cyberaftaler inden for international security. Hvad er ansvarlig state behavior inden for cyber security? FN Working group on cyber security. Langt fra en aftale om hvad der er uacceptabel adfærd.
- International AI regulering kan lære af ovennævnte.

- Modsætning mellem videnskab og krig (fra Russell-Enistein manifesto)
- Ny tænkning – I modsætning til “If you want peace, prepare for war”
- Der oprustes nu.
- Delocalisation of the killing zone.

Working groups



CONFERENCE WORKING GROUPS

Working Group 1: Nuclear disarmament and the future of arms control

Co-convenors: **Alexander Nikitin** (Russia) & **Steve Miller** (USA)

Working Group 2: Nuclear Energy and Non-Proliferation

Co-convenors: **Manpreet Sethi** (India) & **Reza Ziaran** (Iran)

Working Group 3: Asia-Pacific and Nuclear Weapons

Co-convenors: **Masako Ikegami** (Japan) & **Elisabeth Suh** (Germany)

Working Group 4: Conflict and WMD in the Middle East

Co-convenors: **Khaled Shamaa** (Egypt) & **Jan Hoekema** (Netherlands)

Working Group 5: European Security

Co-convenors: **Cliff Kupchan** (USA) & **Thomas Jonter** (Sweden)

Working Group 6: Emerging Technologies and their implications for conflict

Co-convenors: **Stewart Prager** (USA) & **Sibylle Bauer** (Germany)

Sessions jeg ikke tog noter til

- PL5, PL6: The urgency of nuclear reductions. Practical steps and Instruments for Nuclear disarmament.
- PL10, PL11, PL12: Mellemøsten.
- PL14: Averting Escalation and Working towards Stability in Northeast Asia.
- PL15, 16: Gaza.
- PL19: Renunciation of war.