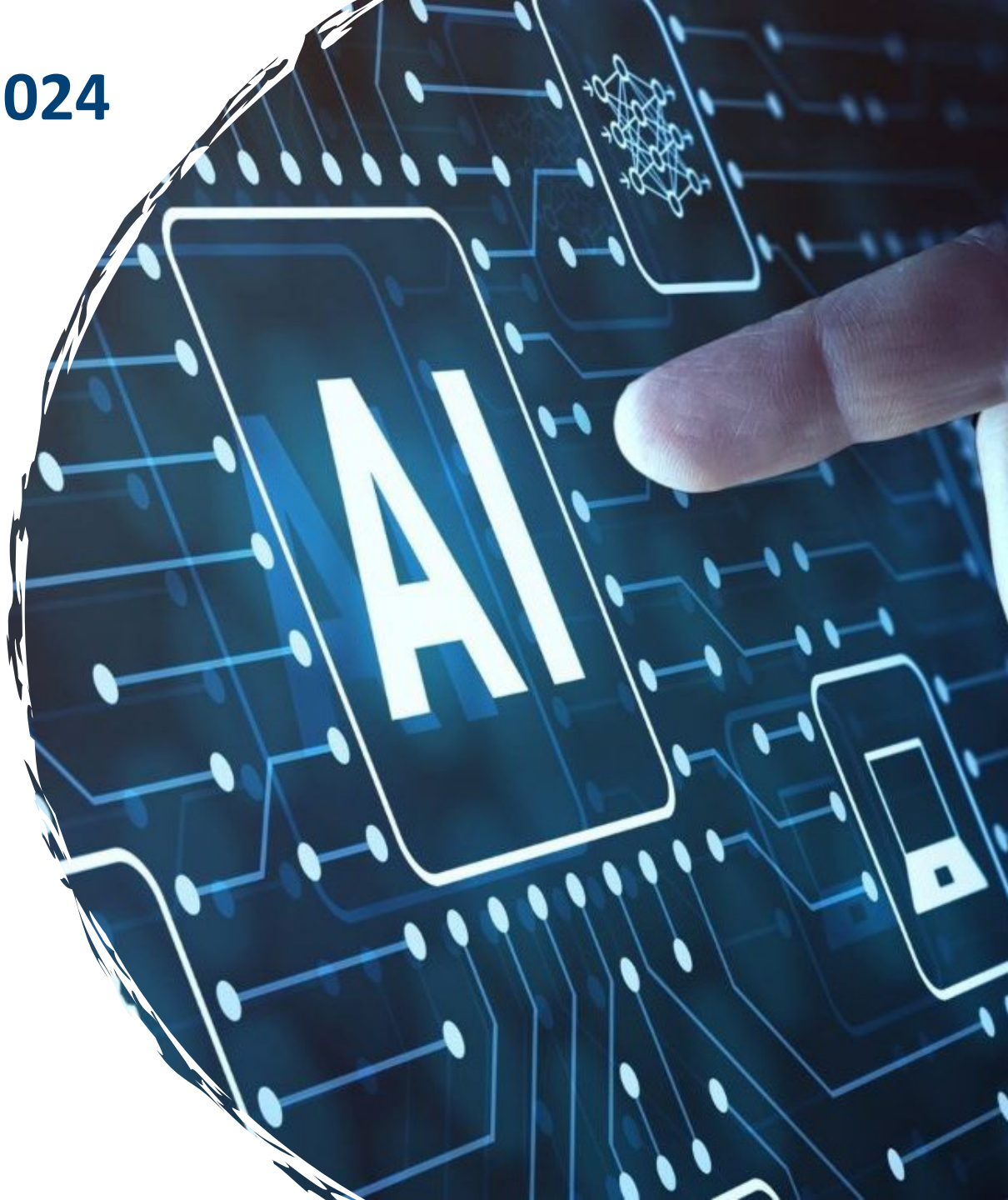
ETSI AI Conference 2024



Current and Future Activities of ETSI TC SAI (Securing Artificial Intelligence)

Presented by: Scott CADZOW, Chair TC SAI

6th February 2024

TC SAI – highlights

- 1st meeting of TC SAI held 4th and 5th of December
- Scott Cadzow endorsed by the group as chair
  - 3 vice chair positions proposed for election at SAI#02 in February
- 11 work items active in the group
  - https://portal.etsi.org/tb.aspx?tbid=913&SubTB=913#/
  - 2 of these adopted work items are already in the ETSI Publication queue
  - 2 to 4 more work items expected to be added in late February from previous work in ETSI ISG SAI
- Several liaison statements seeking collaboration on AI activity have been sent:
  - CEN/CENELEC JTC21
  - ENISA and JRC
  - ETSI TBs including CYBER, ITS, eHEALTH, MTS …

TC SAI – from our terms of reference

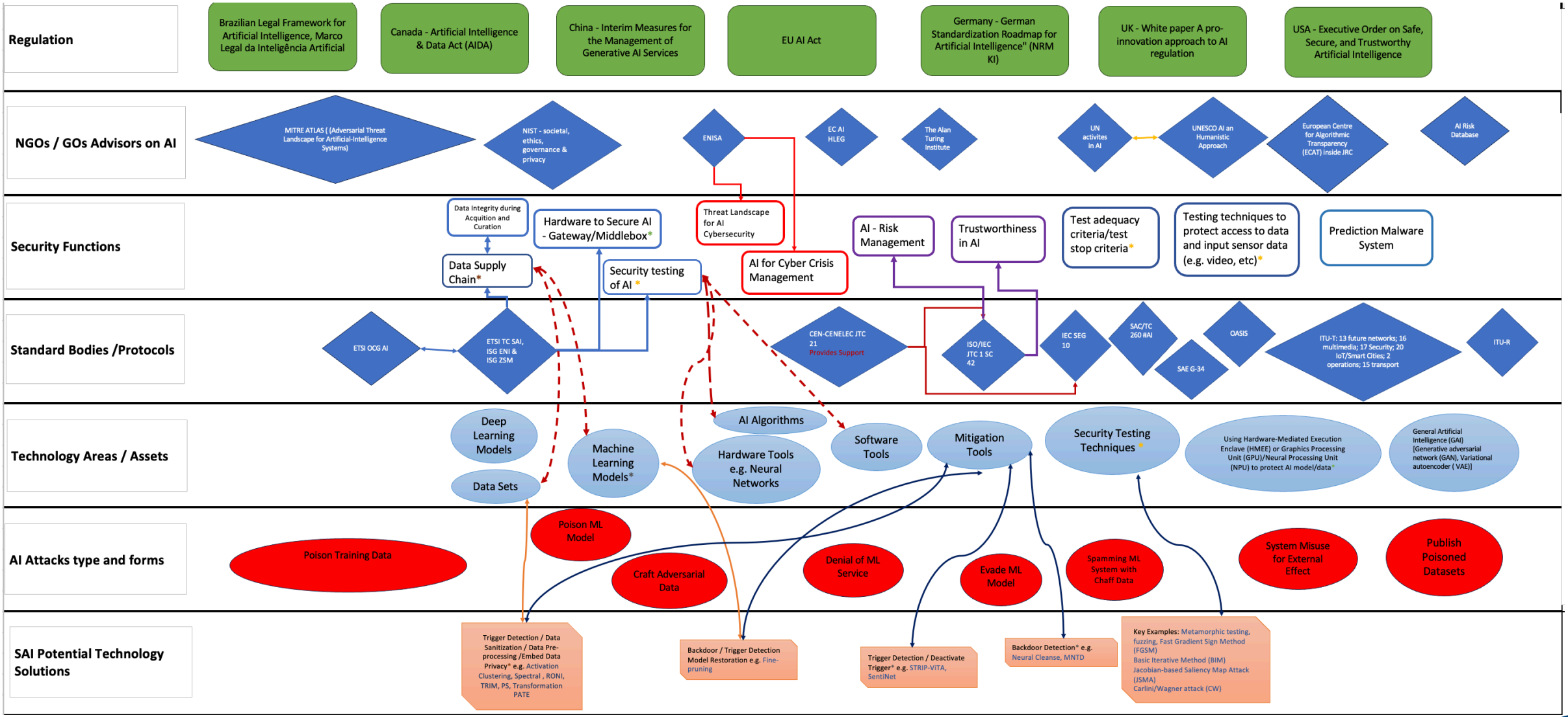
- Terms of Reference
- The aim of Technical Committee Securing Artificial Intelligence (TC SAI) is to develop technical specifications that mitigate against threats arising from the deployment of AI, and threats to AI systems, from both other AIs, and from conventional sources. Whilst in the short to medium term the focus of TC SAI will be on the application of Machine Learning (ML) the group shall also give guidance and evaluation reports to ETSI and its stakeholders on the wider developments of AI
- AI is becoming an increasing element of the ICT world thus it is essential that it is made secure, safe and societally responsible.
  - The word "securing" in the name of TC SAI is intended to address all of those aspects: AI has to be secure but it cannot only be secure - it has to be safe, it has to be societal, it has to be suitable. Thus the “S” in SAI is expanded to have all of these meanings.

Not ethics -- see supplementary slides

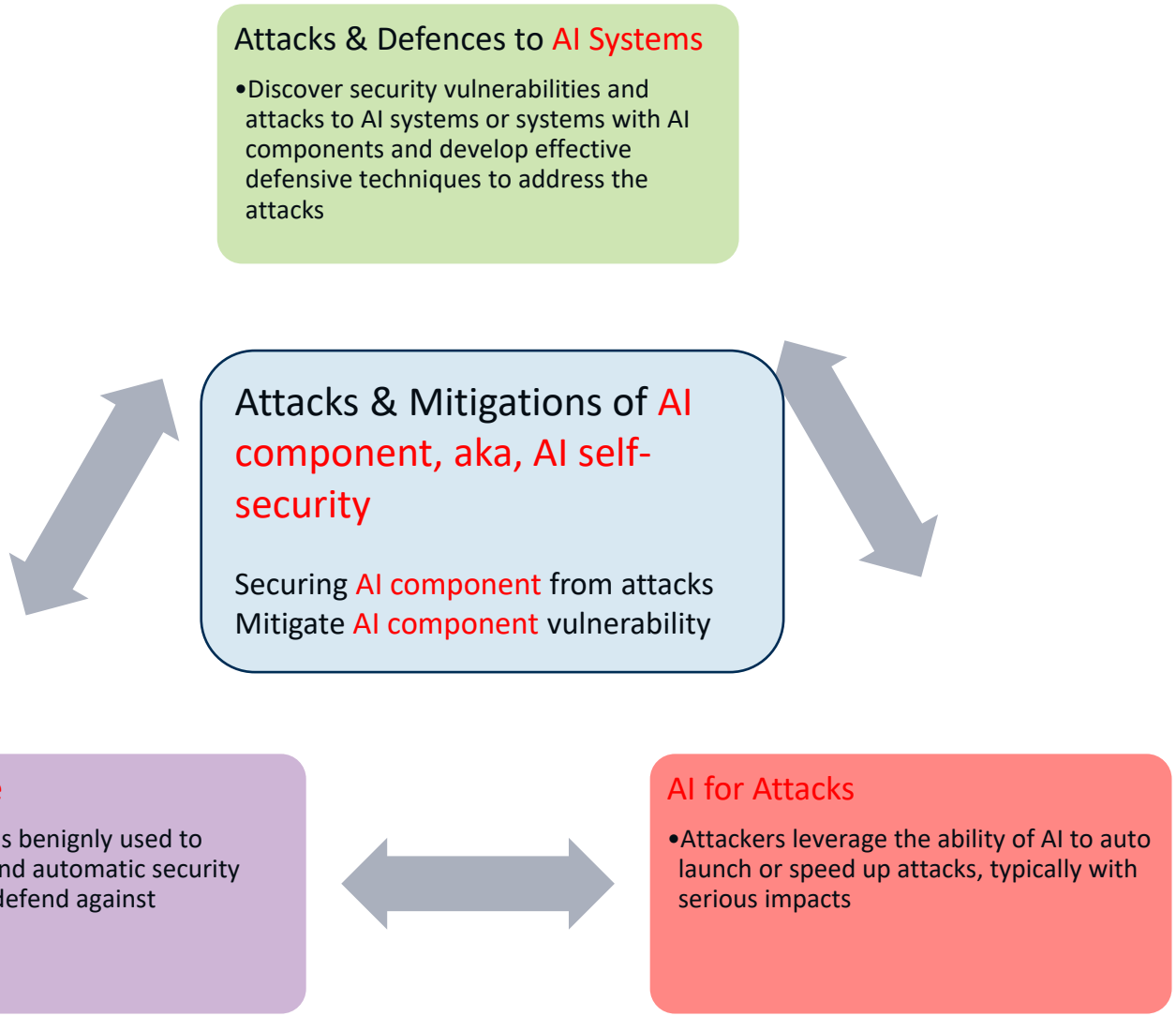
TC SAI – Our place in the EU

- TC SAI has the opportunity to lead on the standards that fit to Europe and wider afield and will take it
  - Article 8 of the proposed final draft of the Cyber Resilience Act (CRA) addresses High Risk AI as a digital element
    - There is a proposal for an AI Act to regulate the use of AI to be safe and secure, and socially accountable in addition to considerations of AI as a digital element in the CRA
  - There are multiple initiatives by EU and partner governments to regulate or guide the use of AI and all of them will require standards to guide the market's best practice

TC SAI - a view on our world



TC SAI – the attack cycle



TC SAI – our work programme

- Inherited, updated and reworked from ISG SAI
  - DTS/SAI-005 (TS 104 050) AI Ontology and definitions
  - DTR/SAI-007 (TR 104 048) Data supply chain report
  - DTR/SAI-009 (TR) Mitigation strategy report
  - DTR/SAI-008 (TR) Problem statement
  - DTS/SAI-006 (TS 104 033) AI Computing Platform Security Framework
  - DTR/SAI-004 (TR 104 032) Traceability of AI Models
  - DTR/SAI-003 (TR 104 031) Collaborative Artificial Intelligence
  - DTR/SAI-0011 (TR) Security aspects of using AI/ML techniques in telecom sector
  - DTR/SAI-0010 (TR) Automated Manipulation of Multimedia Identity Representations
- New work
  - DTR/SAI-001 (TR 104 029) Global AI Security Ecosystem
  - DTR/SAI-002 (TR 104 030) AI Critical Security Controls
  - DTR/SAI-0013 (TR) AI Act mapping and gap analysis

TC SAI – capturing a global vision

To develop an ETSI Standard based on the common principles for secure AI system development and deployment and to set these as the top level horizontal framework for AI.

- National Security Agency (NSA)
- Federal Bureau of Investigation (FBI)
- Australian Signals Directorate's Australian Cyber Security Centre (ACSC)
- Canadian Centre for Cyber Security (CCCS)
- New Zealand National Cyber Security Centre (NCSC-NZ)
- Chile's Government CSIRT
- National Cyber and Information Security Agency of the Czech Republic (NUKIB)
- Information System Authority of Estonia (RIA)
- National Cyber Security Centre of Estonia (NCSC-EE)
- French Cybersecurity Agency (ANSSI)
- Germany's Federal Office for Information Security (BSI)
- Israeli National Cyber Directorate (INCD)
- Italian National Cybersecurity Agency (ACN)
- Japan's National center of Incident readiness and Strategy for Cybersecurity (NISC)
- Japan's Secretariat of Science, Technology and Innovation Policy, Cabinet Office
- Nigeria's National Information Technology Development Agency (NITDA)
- Norwegian National Cyber Security Centre (NCSC-NO)
- Poland Ministry of Digital Affairs
- Poland's NASK National Research Institute (NASK)
- Republic of Korea National Intelligence Service (NIS)
- Cyber Security Agency of Singapore (CSA)

https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf

TC SAI – work we're planning to do (very) soon

- Documenting the principles for Secure AI (see previous slide) as an ETSI TS
- Update and republication of GR SAI-007 as an ETSI TS:
  - Explicability and transparency of AI processing
- Finalisation of draft GR SAI-008 as an ETSI TR:
  - Privacy aspects of AI/ML systems
- Finalisation of draft GR SAI-003 as an ETSI TR (in collaboration with ETSI TC MTS):
  - Security Testing of AI
- Update and republication of GR SAI-006 as an ETSI TR:
  - The role of hardware in security of AI

In broad summary

- ETSI TC SAI intends to act as the gravitational well of SAI standardisation across ETSI:
  - Providing the basis for strong horizontal guidance to making the AI in the market secure (in all its meanings)
  - Developing guidance to vertical markets adopting AI to ensure the guiding principles are maintained irrespective of industrial or commercial sector

Thank you for your attention

Follow us on:

Some supplementary slides

Why not ethics?

The following characteristics (or tests) set out in ETSI's guide to developing standards should be embedded into the development of any contribution to a standard:

- **Necessary**: it (a standard) should specify only what is required to meet its objectives, and not impose a particular approach to implementation.
- **Unambiguous**: it should be impossible to interpret the normative parts of the standard in more than one way.
- **Complete**: the requirement should contain all the information necessary to understand that requirement, either directly or by reference to other documents. The reader of a standard should not need to make assumptions about the implementation of any requirement.
- **Precise**: the requirement should be worded clearly and exactly, without unnecessary detail that might confuse the reader.
- **Well-structured**: the individual elements of the requirement should all be included in an appropriate and easy-to-read manner.
- **Consistent**: there should be no contradiction between different requirements within the standard, nor with other related standards.
- **Testable**: there should be clear and obvious means of demonstrating that an implementation complies with the requirement.

It is unlikely to be able to write a requirement for Ethics, of AI or in general, that satisfies these criteria.

The Lintilla problem in ethics (from Douglas Adam's HHGTTG)

While initially creating six clones of Lintilla, the machine used to clone her had a slight break down. The machine got stuck in a loop, and by the time one Lintilla clone had been created, a half clone had been started. Therefore, the machine could not be shut off without committing murder and would thus go on creating Lintillas indefinitely.

Can you stop the machine without committing murder? Can you stop the indefinite supply of Lintillas without committing murder?

This problem taxed the minds first of the cloning engineers, then of the clergy, then of the letters page of Siderial Record Straightener, and finally of the cloning machine company's lawyers. The lawyers experimented vainly with various ways of redefining murder, reevaluating it, and in the end even respelling it in the hope that no one would notice. Of course, they did, and in a final attempt to stem the tide of Lintillas, a group of Allitnils have been deployed: anti-clones designed to eliminate the Lintillas in the most humane and legally defensible way possible.

The ”solution” makes a Lintilla and an Allitnil fall in love and immediately agree to a “marriage” with the fraud that their marriage certificates are actually cloning machine company "Agreements to Cease to Be”.

Is the solution ethical? Obviously not, probably not, maybe.

The bovinity problem in ethics

In the scenario of an AI controlling a cow milking shed the rules may need to change from protecting humanity to protecting bovinity. It is not clear how much harm is actually allowed either. To make a cow produce milk it is “manipulated” to artificially extend the lactation period of mammalian females (a cow’s gestation period is about 285 days (slightly longer than human females)). Is it wrong to remove the newborn calf from its mother immediately after birth to gain access to the milk of the mother?

The AI could apply an ethical framework from human behaviour to the dairy herd and refuse to participate.

If this is an ethically correct choice is open to debate, and has been debated for a long time in human society with no absolute conclusion (we still milk cows after removing their calf for veal). In extending this, if an AI is able to recognise intelligence in any being, then an AI may invoke an ethical code at odds with those of the human population it was designed to assist.

If two AIs with different ethical codes meet will the contact result in the same consequences that human societies with different ethical codes tend to reach? This is often conflict for which there are many examples: vegans against butchers; animal rights activists against farmers; religion; politics.