

Z-INSPECTION®

MAGNUS WESTERLUND

Z-INSPECTION® INITIATIVE

[HTTP://Z-INSPECTION.ORG](http://z-inspection.org)

ARCADA UNIVERSITY OF APPLIED SCIENCES, HELSINKI, FINLAND

KRISTIANIA UNIVERSITY COLLEGE, OSLO, NORWAY

Z-INSPECTION® IS A REGISTERED TRADEMARK.

THE CONTENT OF THIS WORK IS OPEN ACCESS DISTRIBUTED UNDER THE TERMS AND CONDITIONS OF THE CREATIVE COMMONS
(ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE
CC BY-NC-SA) LICENSE (**[HTTPS://CREATIVECOMMONS.ORG/LICENSES/BY-NC-SA/4.0/](https://creativecommons.org/licenses/by-nc-sa/4.0/)**)



PHOTO RVZ

Z-INSPECTION®: A PROCESS TO ASSESS TRUSTWORTHY AI

Based on a view of contemporary Western European democracy, and fundamental values of respect for others, expressed through support for fundamental human rights.

It's a participatory process that helps teams of skilled experts assess the ethical, technical, domain-specific, and legal implications of using AI Products/services within given *contexts*.

Z-inspection® is a registered trademark.
This work is distributed under the terms and conditions of the Creative Commons (Attribution-NonCommercial-ShareAlike CC BY-NC-SA) license.

WE USE THE EU FRAMEWORK FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE

THE EU HIGH-LEVEL EXPERT GROUP ON AI DEFINED ETHICS *GUIDELINES* FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE, WITH THREE **PILLARS**:

- **LAWFUL** - RESPECTING ALL APPLICABLE LAWS AND REGULATIONS
- **ETHICAL** - RESPECTING ETHICAL PRINCIPLES AND VALUES
- **ROBUST** - BOTH FROM A TECHNICAL PERSPECTIVE WHILE TAKING INTO ACCOUNT ITS SOCIAL ENVIRONMENT

FOUR **ETHICAL PRINCIPLES**, ROOTED IN FUNDAMENTAL RIGHTS

- RESPECT FOR HUMAN AUTONOMY
- PREVENTION OF HARM
- FAIRNESS
- EXPLICABILITY

THERE MAY BE **TENSIONS** BETWEEN THESE PRINCIPLES.

SOURCE: *ETHICS GUIDELINES FOR TRUSTWORTHY AI*. INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. EUROPEAN COMMISSION, 8 APRIL, 2019.

WE USE THE SEVEN REQUIREMENTS AND SUB-REQUIREMENTS FOR TRUSTWORTHY AI



PRIOR WORK FROM AI HLEG



THEY OFFER A **STATIC CHECKLIST** AND WEB TOOL (ALTAI) FOR SELF-ASSESSMENT, BUT **THEY DO NOT VALIDATE CLAIMS OR ACCOUNT FOR CHANGES IN AI OVER TIME.**

THE AI HLEG TRUSTWORTHY AI **GUIDELINES ARE NOT A LAW AND ARE NOT CONTEXTUALIZED BY THE DOMAIN THEY ARE INVOLVED IN.** THE MEANING OF SOME OF THE SEVEN REQUIREMENTS (E.G., FAIRNESS, WELLBEING, ETC.) IS NOT ANCHORED TO THE CONTEXT.

ASSESSMENT OF THE THREE PILLARS OF TRUSTWORTHY AI (TAI) – LAWFUL, ETHICAL, AND ROBUST – REQUIRES A MORE DYNAMIC APPROACH THAN A ONE-OFF CHECKLIST COMPLETED BY AN INDIVIDUAL. IT CALLS FOR AN ONGOING AND MULTI-FACETED EVALUATION PROCESS TO TRULY ENSURE THE TRUSTWORTHINESS OF AI IN HEALTHCARE.

Z-INSPECTION® PROCESS CAN BE APPLIED TO THE ENTIRE AI LIFE CYCLE

- DESIGN
- DEVELOPMENT
- DEPLOYMENT
- MONITORING
- CO-DESIGN ASSESSMENT
 - ASSESS THE SYSTEM IN COLLABORATION WITH DESIGN.
- POST HOC ASSESSMENT
 - ASSESS AI SYSTEMS ALREADY DESIGNED, IMPLEMENTED AND DEPLOYED.

EXAMPLES OF OUR RESEARCH WORK ON BEST PRACTICES

POST HOC ASSESSMENTS

- **ASSESSING TRUSTWORTHY AI. BEST PRACTICE:**
 - **AI FOR PREDICTING CARDIOVASCULAR RISKS** (*JAN. 2019-AUGUST 2020*)
- **ASSESSING TRUSTWORTHY AI. BEST PRACTICE:**
 - **MACHINE LEARNING AS A SUPPORTIVE TOOL TO RECOGNIZE CARDIAC ARREST IN EMERGENCY CALLS.** (*SEPTEMBER 2020-MARCH 2021*)
- **ASSESSING TRUSTWORTHY AI IN TIMES OF COVID-19.**
 - **DEEP LEARNING FOR PREDICTING A MULTI-REGIONAL SCORE CONVEYING THE DEGREE OF LUNG COMPROMISE IN COVID-19 PATIENTS.** (*APRIL- DEC. 2021*)

CO-DESIGN ASSESSMENTS

- **ASSESSING THE INCEPTION STAGE**
 - **CO-DESIGN OF A TRUSTWORTHY AI SYSTEM IN HEALTHCARE: DEEP LEARNING BASED SKIN LESION CLASSIFIER.** (*NOVEMBER 2020-MARCH 2021*)
- **ASSESSING RESEARCH PROJECT REQUIREMENTS**
 - **EU HORIZON VALIDATE PROJECT: HEALTHCARE** (2023 --)
- **FULL CO-DESIGN ASSESSMENT**
 - **EU HORIZON MANOLO PROJECT: THE CLOUD-EDGE CONTINUUM** (2024 --)

Z-INSPECTION® SET UP STAGE

MAGNUS WESTERLUND

Z-INSPECTION® INITIATIVE

[HTTP://Z-INSPECTION.ORG](http://z-inspection.org)

ARCADA UNIVERSITY OF APPLIED SCIENCES, HELSINKI, FINLAND

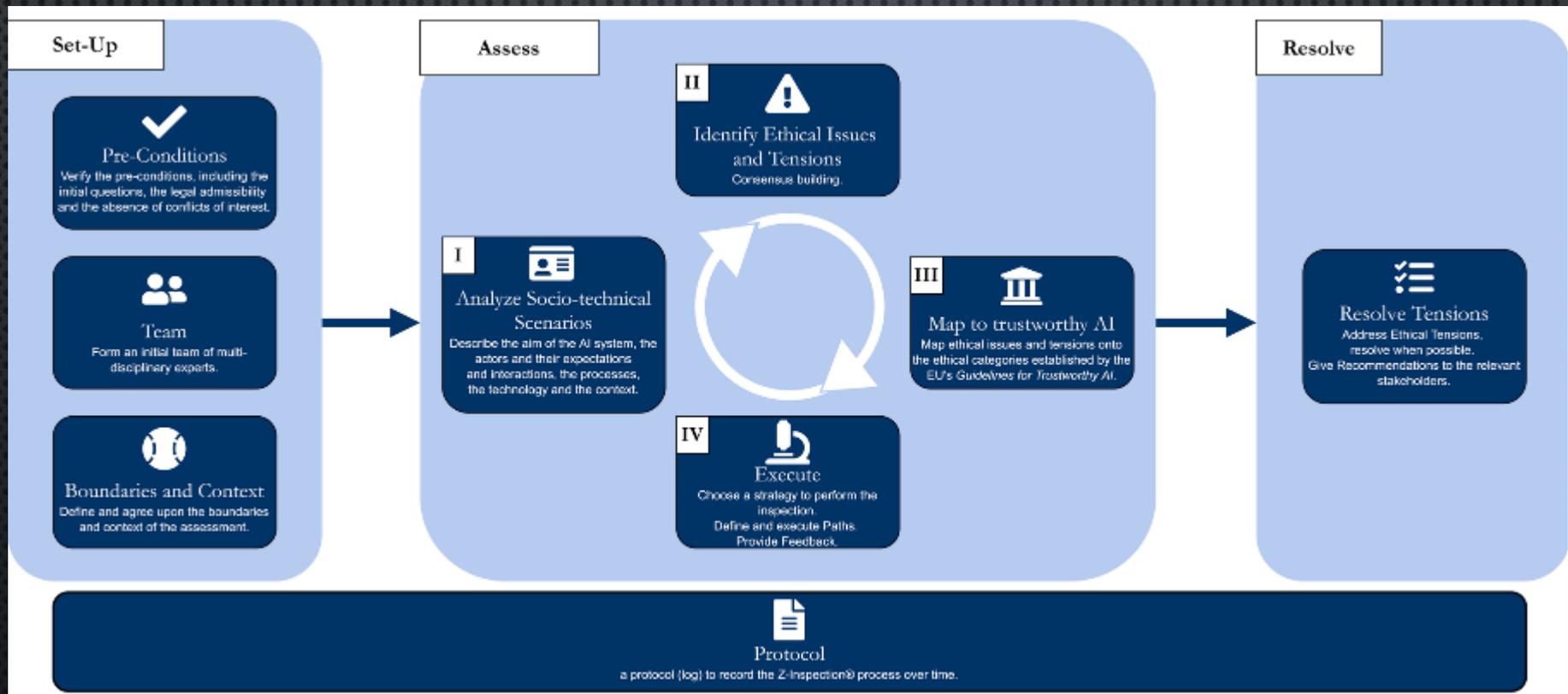
KRISTIANIA UNIVERSITY COLLEGE, OSLO, NORWAY

Z-INSPECTION® IS A REGISTERED TRADEMARK.

THE CONTENT OF THIS WORK IS OPEN ACCESS DISTRIBUTED UNDER THE TERMS AND CONDITIONS OF THE CREATIVE COMMONS
(ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE

CC BY-NC-SA) LICENSE ([HTTP://CREATIVECOMMONS.ORG/LICENSES/BY-NC-SA/4.0/](http://creativecommons.org/licenses/by-nc-sa/4.0/))

Z-INSPECTION® PROCESS: SET UP



1) PRE-CONDITIONS

Verify the pre-conditions, including the initial questions, the legal admissibility and the absence of conflict of interests.

Who requested the inspection?

Why carry out an inspection?

For whom is the inspection relevant?

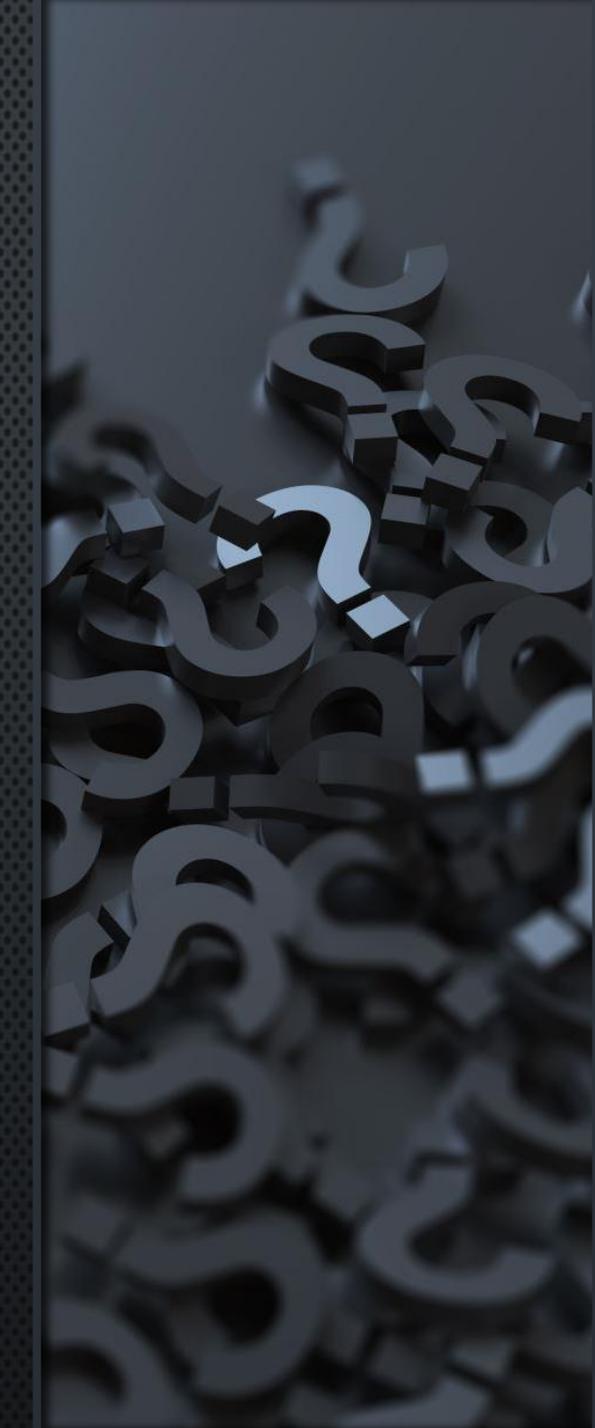
Is it recommended or required (mandatory inspection)?

What are the sufficient vs. necessary conditions that need to be analyzed?

How are the inspection results to be used?

Will the results be shared (public) or kept private?

Are there conflict of interests?



PRE-CONDITION ANALYSIS

DEFINE THE IMPLICATIONS IF CONDITIONS ARE NOT SATISFIED. FOR EXAMPLE:

- WHICH STAKEHOLDERS (IF ANY) HAVE BEEN LEFT OUT OF SCOPE? FOR WHAT REASON(S)?
- HOW WILL CONFLICTS OF INTEREST BE ADDRESSED BETWEEN PARTICIPANTS?
- WILL THE INSPECTION BE REVISITED AT A LATER DATE?
- WILL THE PARTICIPANTS CHANGE?

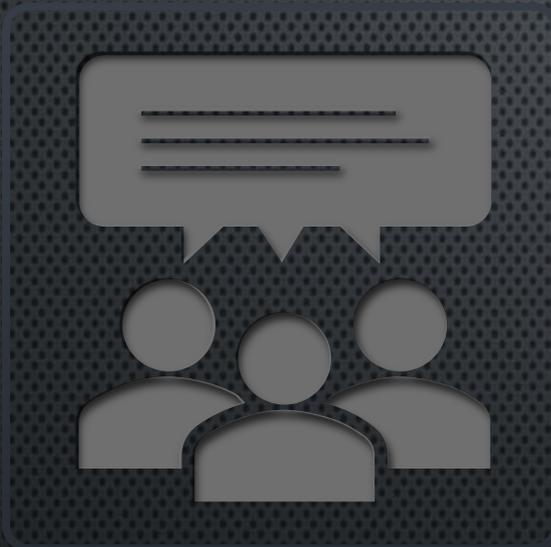
2) CREATION OF AN INTERDISCIPLINARY TEAM

- IN THE SET UP PHASE WE CREATE **AN INTERDISCIPLINARY** ASSESSMENT TEAM COMPOSED OF A DIVERSE RANGE OF EXPERTS.
- DEPENDING ON THE USE CASE (AND DOMAIN), THE TEAM MAY INCLUDE: PHILOSOPHERS, HEALTHCARE ETHICISTS, HEALTHCARE DOMAIN EXPERTS (SPECIALISTS, SUCH AS RADIOLOGISTS AND OTHER CLINICIANS, AND PUBLIC HEALTH RESEARCHERS), LEGAL RESEARCHERS, ETHICS ADVISORY, SOCIAL SCIENTISTS, AI ENGINEERS, AND PATIENT REPRESENTATIVES.
- IT IS IMPORTANT TO BRING TOGETHER A BROADER SET OF STAKEHOLDERS AT ALL STAGES OF THE AI LIFE CYCLE.
- THIS INTERDISCIPLINARITY IS ONE OF THE MOST IMPORTANT ASPECTS OF OUR APPROACH **TO ENSURE THAT A VARIETY OF VIEWPOINTS ARE EXPRESSED** WHEN ASSESSING THE TRUSTWORTHINESS OF AN AI SYSTEM.
- **THE CHOICE OF THE EXPERTS HAS AN ETHICAL IMPLICATION!**

SUGGESTION ON TEAM FORMATION

- TEAM MEMBERS SHOULD BE SELECTED BASED PRIMARILY ON SKILLS REQUIRED / EXPERTISE – AVAILABILITY AND INTEREST IN THE CASE
 - **LEAD:** COORDINATES THE PROCESS;
 - **RAPPORTEUR:** APPOINTED TO REPORT ON THE PROCEEDINGS OF ITS MEETINGS.
 - **ETHICIST(S)** : HELP THE OTHER EXPERTS;
 - **DOMAIN EXPERT(S):** BETTER MORE THEN ONE WITH DIFFERENT VIEWPOINTS;
 - **LEGAL EXPERT(S):** RELATED TO THE DOMAIN;
 - **TECHNICAL EXPERT(S):** MACHINE LEARNING, DEEP LEARNING;
 - (SOCIAL SCIENTISTS, POLICY MAKERS, COMMUNICATION, OTHERS)
 - **REPRESENTATIVE OF END USERS.**
- MOTIVATION IS ESSENTIAL BUT SHOULD NOT BE #1 CRITERIA FOR INVOLVEMENT.
- LATER ADDITIONS OF EXPERTS TO THE TEAM SHOULD BE LIMITED.

CHALLENGE



- THE MAIN CHALLENGE IS TO MAKE SURE THAT ALL EXPERTS HAVE A HOLISTIC VIEW OF THE PROCESS AND A GOOD UNDERSTANDING OF THE USE CASE.
- FOR THAT, ALL TEAM MEMBERS AND RELEVANT USE CASE STAKEHOLDERS NEED TO BE TRAINED OR TRAIN THEMSELVES ON THE EU REGULATION / Z-INSPECTION® PROCESS.

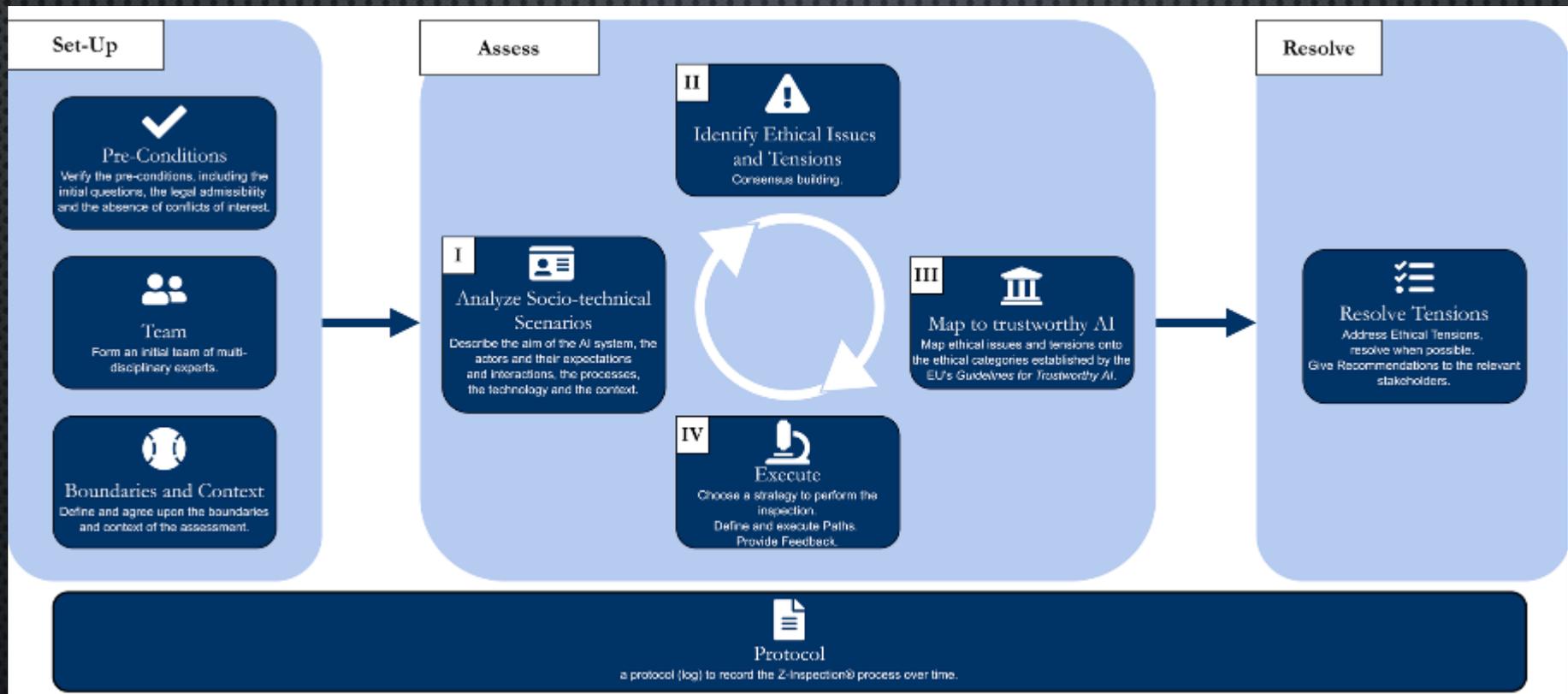
THE ROLE OF PHILOSOPHERS / ETHICISTS

- **APPLIED ETHICS**
- THEY SHOULD ACT AS “ADVISORS” TO REST OF THE TEAM, BE PART OF THE PROCESS TO IDENTIFY OF ETHICAL TENSIONS, BE PART OF THE MAPPING TO THE TRUSTWORTHY AI FRAMEWORK AND BE AVAILABLE FOR ETHICS RELATED QUESTIONS.
- IF THEY HAVE USE CASE SPECIFIC PRACTICAL EXPERTISE (E.G. HEALTH / MEDICAL ETHICS) THEY COULD LEAD THE PART OF THE PROCESS THAT IS TO IDENTIFY OF ETHICAL TENSIONS.

3) DEFINITION OF THE BOUNDARIES AND CONTEXT

- THE SET-UP PHASE ALSO INCLUDES **THE DEFINITION OF THE BOUNDARIES OF THE ASSESSMENT**, TAKING INTO ACCOUNT THAT WE DO NOT ASSESS THE AI SYSTEM IN ISOLATION BUT RATHER CONSIDER **THE SOCIAL-TECHNICAL INTERCONNECTION WITH THE ECOSYSTEM(S)** WHERE THE AI IS DEVELOPED AND/OR DEPLOYED.
- SOME OF THE MOST IMPORTANT ETHICAL AND POLITICAL CONSIDERATIONS OF AI DEVELOPMENT REST **ON THE DECISION TO INCLUDE OR EXCLUDE PARTS OF THE CONTEXT** IN WHICH THE SYSTEM WILL OPERATE.
- CONSIDER THE AIMED FOR **MATURITY** LEVEL. FOR EXAMPLE, THE TECHNOLOGY READINESS LEVEL (TRL) HAS GREAT IMPLICATIONS FOR WHAT SHOULD BE INCLUDED IN THE ASSESSMENT.
- **ADDITIONAL ORGANIZATIONAL SUPPORT PROCESSES** FOR MAINTAINING THE TRUSTWORTHINESS OF THE AI SOLUTION CAN ALSO BE INCLUDED.

Z-INSPECTION® PROCESS: ASSESS



Z-INSPECTION® SOCIO-TECHNICAL SCENARIOS

MAGNUS WESTERLUND

Z-INSPECTION® INITIATIVE

[HTTP://Z-INSPECTION.ORG](http://z-inspection.org)

ARCADA UNIVERSITY OF APPLIED SCIENCES, HELSINKI, FINLAND

KRISTIANIA UNIVERSITY COLLEGE, OSLO, NORWAY

Z-INSPECTION® IS A REGISTERED TRADEMARK.

THE CONTENT OF THIS WORK IS OPEN ACCESS DISTRIBUTED UNDER THE TERMS AND CONDITIONS OF THE CREATIVE COMMONS
(ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE

CC BY-NC-SA) LICENSE ([HTTP://CREATIVECOMMONS.ORG/LICENSES/BY-NC-SA/4.0/](http://creativecommons.org/licenses/by-nc-sa/4.0/))

WHY USE SOCIO-TECHNICAL SCENARIOS?



ONE SPECIFIC ASPECT OF THE METHODOLOGY INVOLVES USING WHAT WE CALL SOCIO-TECHNICAL SCENARIOS TO ANTICIPATE POSSIBLE USES AND PROBLEMS OF THE SYSTEM UNDER REVIEW.

SCENARIOS HELP TO BREAK DOWN THE TECHNICAL INPUT AND TEST THE SET BOUNDARIES

WHAT ARE SOCIO-TECHNICAL SCENARIOS?



- REAL-LIFE SCENARIOS BASED ON POTENTIAL EVERYDAY USE OF TECHNOLOGY
- IDENTIFYING POTENTIAL PROBLEMS THROUGH ANTICIPATION OF EXPERIENCES
- VIEWING SITUATIONS FROM MULTIPLE PERSPECTIVES FOR A WELL-ROUNDED APPRAISAL
- TEAM COLLABORATION WITH DIVERSE BACKGROUNDS IN TECHNOLOGY ASSESSMENTS
- FOCUSING ON SPECIFIC CONTEXT TO AVOID ABSTRACT OPPOSITION BETWEEN GENERAL PRINCIPLES

WHY ARE SOCIO-TECHNICAL SCENARIOS USEFUL?



SCENARIO BUILDING PROVED USEFUL IN ALLOWING LIVELY DISCUSSION OF SPECIFIC PROBLEMS RELATED TO THE USE CASE WHILE SIMULTANEOUSLY ENGAGING TENSIONS BETWEEN FUNDAMENTAL PRINCIPLES AND, ULTIMATELY, VARIOUS UNDERSTANDINGS OF OUR MORAL OBLIGATIONS (UNDERSTANDINGS THAT CAN BE LINKED TO THE MAIN ETHICAL THEORIES).

WHY ARE SOCIO-TECHNICAL SCENARIOS USEFUL?

- SCENARIOS CRAFTED CONSIDERING ETHICAL IMPLICATIONS OF SYSTEM SPECIFICATIONS AND ITS SOCIO-TECHNICAL CONTEXT
- INTENDED OPERATIONAL IMPACT ON VARIOUS STAKEHOLDER GROUPS TAKEN INTO ACCOUNT- DEPLOYMENT TERMS ASSESSED FOR REFLECTION AND ENACTMENT OF STAKEHOLDER CONSENSUS ON APPROPRIATE USE
- Z-INSPECTION® TEAMS APPLY JUDGMENT IN SCENARIO CREATION AND INTERNAL DISCUSSIONS DURING THE INSPECTION PROCESS
- INITIAL STEP RESULTS IN PINPOINTING KEY ISSUES FOR FURTHER ASSESSMENT

WE USE SOCIO-TECHNICAL SCENARIOS TO IDENTIFY *ISSUES*

BY COLLECTING RELEVANT RESOURCES, A TEAM OF INTERDISCIPLINARY EXPERTS CREATE SOCIO-TECHNICAL SCENARIOS AND ANALYZE THEM TO DESCRIBE:

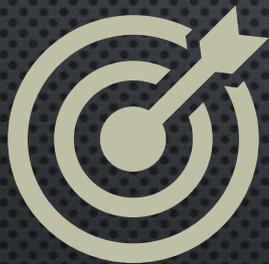
- 1. THE AIM OF THE AI SYSTEMS,**
- 2. THE ACTORS AND THEIR EXPECTATIONS AND INTERACTIONS,**
- 3. THE PROCESS WHERE THE AI SYSTEMS ARE USED,**
- 4. THE TECHNOLOGY,**
- 5. THE BROADER ECOSYSTEM.**

RESULTING IN A NUMBER OF *ISSUES* TO BE ASSESSED.

HOW TO START?

- INITIALLY, THE TEAM OF EXPERTS MEETS WITH THE STAKEHOLDERS OWNING THE USE CASE IN A NUMBER OF WORKSHOPS (VIA VIDEO CONFERENCE) TO DEFINE **SOCIO-TECHNICAL SCENARIOS OF THE USE OF THE AI SYSTEMS.**
- WE USE THE TERM **STAKEHOLDERS** TO **DENOTE** THE **ACTORS** WHO HAVE DIRECT **OWNERSHIP** OF THE DEVELOPMENT AND DEPLOYMENT OF THE AI SYSTEM.

1) AIM IN SOCIO-TECHNICAL SCENARIOS



What is the **Aim of the system**



Goal of the system,
context **WHY** it is used

2) ACTORS IN SOCIO-TECHNICAL SCENARIOS

ACTORS (PRIMARY, SECONDARY, TERTIARY)

- WHO IS USING THE SYSTEM,
- WHO IS INFLUENCED BY DECISIONS OF THE SYSTEM,
- WHO HAS INTEREST IN THE SYSTEM BEING DEPLOYED, ...

ACTORS' CONCERNS AND WORRIES

- WHAT PROBLEMS / CHALLENGES CAN THE ACTORS FORESEE?
- DO THEY HAVE CONCERNS REGARDING THE USE OF THE SYSTEM?

2) ACTORS IN SOCIO-TECHNICAL SCENARIOS

ACTORS' EXPECTATIONS AND MOTIVATION

- WHY WOULD THE DIFFERENT GROUPS OF ACTORS WANT THE SYSTEM?
- WHAT ARE THEIR EXPECTATIONS TOWARDS THE SYSTEM'S BEHAVIOR?
- WHAT BENEFITS ARE THEY EXPECTING FROM USING THE SYSTEM?
- ARE THERE ANY CONFLICTS?

3) PROCESS IN SOCIO-TECHNICAL SCENARIOS

THE PROCESS WHERE THE AI SYSTEM IS USED

- ADDITIONAL CONTEXT FOR THE SITUATIONS WHERE THE AI SYSTEM IS USED, HOW IT COULD BE USED IN THE FUTURE?

INTERACTION WITH THE AI SYSTEM

- WHAT IS THE INTENDED INTERACTION BETWEEN THE SYSTEM AND ITS USERS?
- WHY IS IT LIKE THIS?
- HOW ARE DECISIONS BEING MADE WHEN THE AI IS ACTIVE?

4) TECHNOLOGY IN SOCIO-TECHNICAL SCENARIOS



The relevant Technology employed for the AI System



Technical description of the AI system, so technically inclined people get an intuition of how it is working

5) ECOSYSTEM FOR THE SOCIO-TECHNICAL SCENARIOS

FIELD TESTS / CLINICAL STUDIES

- WAS THE SYSTEM'S PERFORMANCE VALIDATED IN FIELD STUDIES?
- WHAT WERE THE RESULTS OF THESE STUDIES?
- ARE THE RESULTS OPENLY AVAILABLE?

INTELLECTUAL PROPERTY

- WHAT IP REGULATIONS NEED TO BE CONSIDERED WHEN ASSESSING/DISSEMINATING THE SYSTEM?
- IS IT OPEN ACCESS?
- DOES IT CONTAIN CONFIDENTIAL INFORMATION THAT MUST NOT BE PUBLISHED?

LEGAL FRAMEWORK

- WHAT IS THE LEGAL FRAMEWORK FOR THE USE OF THE SYSTEM?
- WHAT SPECIAL REGULATIONS APPLY?

5) ECOSYSTEM FOR THE SOCIO-TECHNICAL SCENARIOS

ETHICS OVERSIGHT AND/OR APPROVAL

- HAS THE AI SYSTEM ALREADY UNDERGONE SOME KIND OF ETHICAL ASSESSMENT OR OTHER APPROVAL?
- IF NOT - WHY NOT? IF SO, WAS THIS INTERNAL/EXTERNAL, VOLUNTEER/REGULATED, WHAT WAS COVERED?
- DID THEY GET A WAIVER? WAS THERE A CLEARING, BUT IT WAS VERY LIGHT OR INTERNAL AND NOT CONSIDERED SUFFICIENT?