Z-inspection®
Claim, Arguments and Evidence

Roberto V. Zicari
Z-Inspection® Initiative
http://z-inspection.org

Graduate School of Data Science, SNU, Seoul

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/)

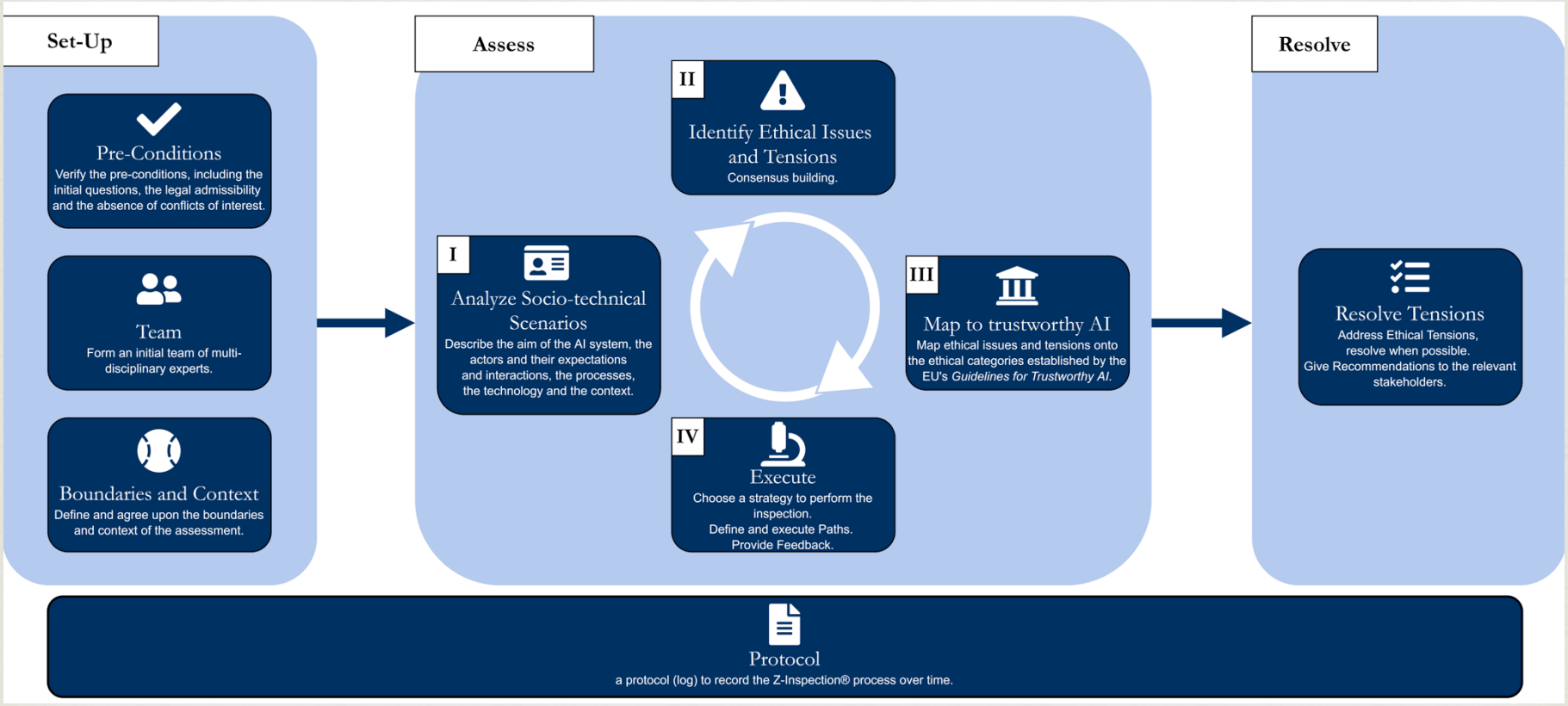
Structure of the Lesson

1. *Claim, Arguments and Evidence*

2. *Use Case:* Assessing Trustworthy AI Best Practice: Machine Learning as a Supportive Tool to Recognize Cardiac Arrest

Self Assessment together with the Emergency Medical Dispatch Center (EMS) of the City of Copenhagen.

Z-inspection® Process in a Nutshell



Infrastructure

- We have been using Zoom for meetings and Google docs for shared content, Google groups for communication to the entire team and e-mails.
- Good experience in using Zoom, with recorded most of the meeting.
- Mix experience with Google docs. When a document becomes too big with many comments, it is not the optimal tool for co-working
- We did not find a good solution for creating a joint "library" with all relevant articles, working documents etc.
- Mix experience with Google groups and e-mails.

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:

the aim of the AI systems,

the actors and their expectations and interactions,

the process where the AI systems are used,

the technology and the context (*ecosystem*).

Resulting in a number **of *issues* to be assessed**.

Concept building

„An important obstacle to progress on the ethical and societal issues raised by AI-based **systems is the ambiguity of many central concepts currently used to identify salient issues**.„

- Terminological overlaps
- Differences between disciplines
- Differences across cultures and publics
- Conceptual complexity

- Source: *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research.* Whittlestone, J. Nyrup, R. Alexandrova, A. Dihal, K. Cave, S. (2019), *London*. Nuffield Foundation.

Concept building

1. Mapping and clarifying ambiguities
2. Bridging disciplines, sectors, publics and cultures
3. Building consensus and managing disagreements

Source: *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research.* Whittlestone, J. Nyrup, R. Alexandrova, A. Dihal, K. Cave, S (2019), *London*. Nuffield Foundation.

We develop an evidence base

This is an iterative process among experts with different skills and background with goal to:

- Understand technological capabilities and limitations
- Build a stronger evidence base to support claims and identify tensions (*domain specific*)
- Understand the perspective of different members of society

Source: Whittlestone, J et al (2019)

Claims, Arguments and Evidence (CAE)

Claims – "assertions put forward for general acceptance. They are typically statements about a property of the system or some subsystem.

Claims that are asserted as true without justification become **assumptions** and claims supporting an argument are called sub claims. "

- Source: – Brundage et al. (2020) – Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims.

Claims, Arguments and Evidence (CAE)

Evidence "that is used as the basis of the justification of the claim.

Sources of evidence may include the design, the development process, prior field experience, testing, source code analysis or formal analysis", peer-reviewed journals articles, peer-reviewed clinical trials, etc.

Source: – Brundage et al. (2020) – Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims.

Claims, Arguments and Evidence (CAE)

Arguments link the evidence to the claim.

They are defined as Toulmin's warrants and are the "statements indicating the general ways of arguing being applied in a particular case and implicitly relied on and whose trustworthiness is well established", together with the validation for the scientific and engineering laws used.

Source: – Brundage et al. (2020) – Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims.

Toulmin Argument

„Stephen Toulmin, a modern rhetorician, developed a model for analyzing the kind of argument you read and hear every day, in newspapers and on television, at work, in classrooms, and in conversation. Toulmin's model focuses on identifying the basic parts of an argument."

"Toulmin identifies the three essential parts of any argument as the **claim**; the **data** (also called **grounds** or **evidence**), which support the claim; and the **warrant**. "

Source: https://www.blinn.edu/writing-centers/pdfs/Toulmin-Argument.pdf

Develop an evidence base

- Technology is generally designed for a highly specific purpose, however, it is not always clear what the technologies unintended harm might be.

- Therefore, an important part of our assessment process is **to build an evidence** base through the **socio-technical scenarios to identify tensions as potential ethical issues.**

Claims

- "AI developers regularly make **claims regarding the properties of AI systems they develop as well as their associated societal consequences. "**

- Source: Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims https://arxiv.org/pdf/2004.07213.pdf

Identify Claims

- **Claims** for technological capability (for example aim, performance, architecture, or functionality, etc.) serve as an important input in developing the **evidence base.**
- This is an iterative process among experts of the assessment team with different skills and backgrounds with a goal to understand technological capabilities and limitations

Verifiable Claims

- „**Verifiable claims** are statements for which **evidence** and **arguments** can be brought to bear on the likelihood of those claims being true.

- The degree of attainable certainty in such claims will vary across contexts. „

- Source: Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims https://arxiv.org/pdf/2004.07213.pdf

Examples of Claims

- We will *adhere* to the data usage protocols we have specified;
- The cloud services on which our AI systems run are *secure*;
- We will *evaluate* risks and benefits of publishing AI systems in partnership with appropriately qualified third parties;

Examples of Claims

- The AI system *is very accurate…*
- The AI system is *more accurate then….*
- The AI system is *98% accurate...*
- The AI *predicts with high quality ….*
- *Using the AI system results in saving XXX do*llars…

Examples of Claims

- We *will not create or sell AI systems that are intended to cause harm;*
- We *will assess and report any harmful societal impacts of AI systems that we build*; and
- Broadly, *we will act in a way that aligns with society's interests.*

“Keep your AI claims in check”
USA Federal Trade Commission Division

Are you exaggerating what your AI product can do?

Or even claiming it can do something beyond the current capability of any AI or automated technology?

For example, we’re not yet living in the realm of science fiction, where computers can generally make trustworthy predictions of human behavior.

Your performance claims would be deceptive if they lack scientific support or if they apply only to certain types of users or under certain conditions.

Source: Keep your AI claims in check, By Michael Atleson, Attorney, US Federal Trade Commision Division of Advertising Practices February 27, 2023

“Keep your AI claims in check” USA Federal Trade Commission Division

Are you promising that your AI product does something better than a non-AI product?

It’s not uncommon for advertisers to say that some new-fangled technology makes their product better – perhaps to justify a higher price or influence labor decisions. You need adequate proof for that kind of comparative claim, too, and if such proof is impossible to get, then don’t make the claim.

Source: Keep your AI claims in check, By Michael Atleson, Attorney, US Federal Trade Commision Division of Advertising Practices February 27, 2023

“Keep your AI claims in check” USA Federal Trade Commission Division

Are you aware of the risks?

You need to know about the reasonably foreseeable risks and impact of your AI product before putting it on the market. If something goes wrong – maybe it fails or yields biased results – you can’t just blame a third-party developer of the technology. And you can’t say you’re not responsible because that technology is a “black box” you can’t understand or didn’t know how to test.

- Source: Keep your AI claims in check, By Michael Atleson, Attorney, US Federal Trade Commision Division of Advertising Practices February 27, 2023

“Keep your AI claims in check” USA Federal Trade Commission Division

Does the product actually use AI at all?

If you think you can get away with baseless claims that your product is AI-enabled, think again. In an investigation, FTC technologists and others can look under the hood and analyze other materials to see if what’s inside matches up with your claims. Before labeling your product as AI-powered, note also that merely using an AI tool in the development process is not the same as a product having AI in it.

- Source: Keep your AI claims in check, By Michael Atleson, Attorney, US Federal Trade Commision Division of Advertising Practices February 27, 2023

It is highly desirable for claims made about AI development to be verifiable.

"First, those potentially affected by AI development–as well as those seeking to represent those parties' interests via government or civil society–deserve to be able to scrutinize the claims made by AI developers in order **to reduce risk of harm** or foregone benefit. "

- Source: Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims https://arxiv.org/pdf/2004.07213.pdf

It is highly desirable for claims made about AI development to be verifiable

- „Second, to the extent that claims become verifiable, various actors such as civil society, policymakers, and users can raise their standards for what constitutes responsible AI development.

- This, in turn, can **improve societal outcomes** associated with the field as a whole. „

- Source: Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims https://arxiv.org/pdf/2004.07213.pdf

It is highly desirable for claims made about AI development to be verifiable

- "Third, a lack of verifiable claims in AI development could foster or worsen a "race to the bottom" in AI development, whereby developers seek to gain a competitive edge even when this trades off against important societal values such as safety, security, privacy, or fairness "

- Source: Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims https://arxiv.org/pdf/2004.07213.pdf

Building a solid knowledge / evidence base

- We suggest **building a solid knowledge / evidence base** among all team members of the use case before the inspection starts and also a solid Q&A log during the inspection process.
- Experts may approach the use case quite differently:
- Interpretations of and expectations for the AI tool being inspected may differ
- Focus of interest may be very different

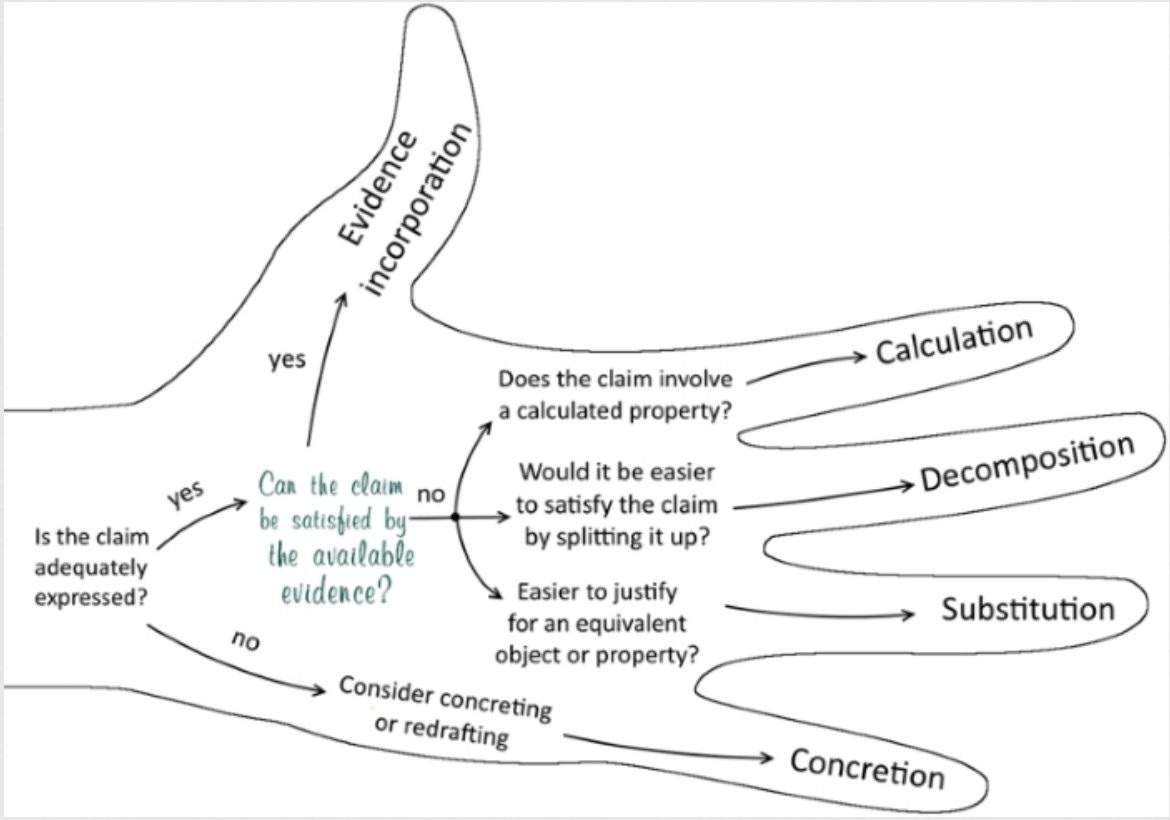
Claims, Arguments and Evidence

- **The claims, arguments and evidence (CAE) framework (*)** can help with the structuring of the use case in a clear and precise form that is supported by evidence.
- For example, each of the claims should be about only one specific property of the system and at the same time, it should be phrased in a way that is clearly verifiable or falsifiable. The CAE framework also provides guidance on how to disseminate complex claims into easier ones .

(*) https://claimsargumentsevidence.org

Claims, Arguments and Evidence

Source: https://claimsargumentsevidence.org



What is Evidence?

- Who is “qualified” to give strong evidence? We could introduce different levels of what constitutes “evidence”.
- Strong evidence is when testing is possible. However, testing is not always possible. We look at peer-reviewed journal articles supporting a claim. This is also evidence.
- When domain experts have different viewpoints, then we list such different viewpoints and related supporting evidence as tensions.

Different View Points

- Experts in different fields will see the AI system quite differently. What may be considered a lack of knowledge can just be a different lens. It's crucial the team understands that there will be very different perspectives based on the specific role or subdomain different experts represent.

Managing Different View Points

- Managing different viewpoints between experts composing the assessment team is an essential part of the process.

- One of the key lessons learned is that there may be tensions when considering what the relevant existing evidence to support a claim is.

Identifying “issues”

- When a Claim has no evidence it becomes an assumptions, and this could be a potential risk.

  We call them “issues”.

- How to describe “issues”?

- Use free text and an open vocabulary

Tensions in Evidence Base

- For example in the case of a skin cancer detection AI tool (*), there were *tensions* between the various arguments linking evidence to support the choice of a design decision derived from the different viewpoints expressed by domain experts.

- (*) – **Co-Design of a Trustworthy AI System in Healthcare: Deep Learning Based Skin Lesion Classifier.** *Front. Hum. Dyn. |Human and Artificial Collaboration for Medical Best Practices, July 13, 2021*

Tensions in Evidence Base

Claim:

This AI System helps dermatologists to early detection of malignant melanoma.

Argument **:**

Malignant melanoma is a very heterogeneous tumor with a clinical course that is very difficult to predict. To date, there are no reliable biomarkers that predict prognosis with certainty. Therefore, there exist subgroups of melanoma patients with different risks for metastasization, some might never metastasize and diagnosing them would be overdiagnosed.

Tensions in Evidence Base

- *View Point*: **Dermatologist**.
  Early detection of malignant melanoma is critical, as the risk of metastasis with worse prognosis increases the longer melanoma remains untreated.

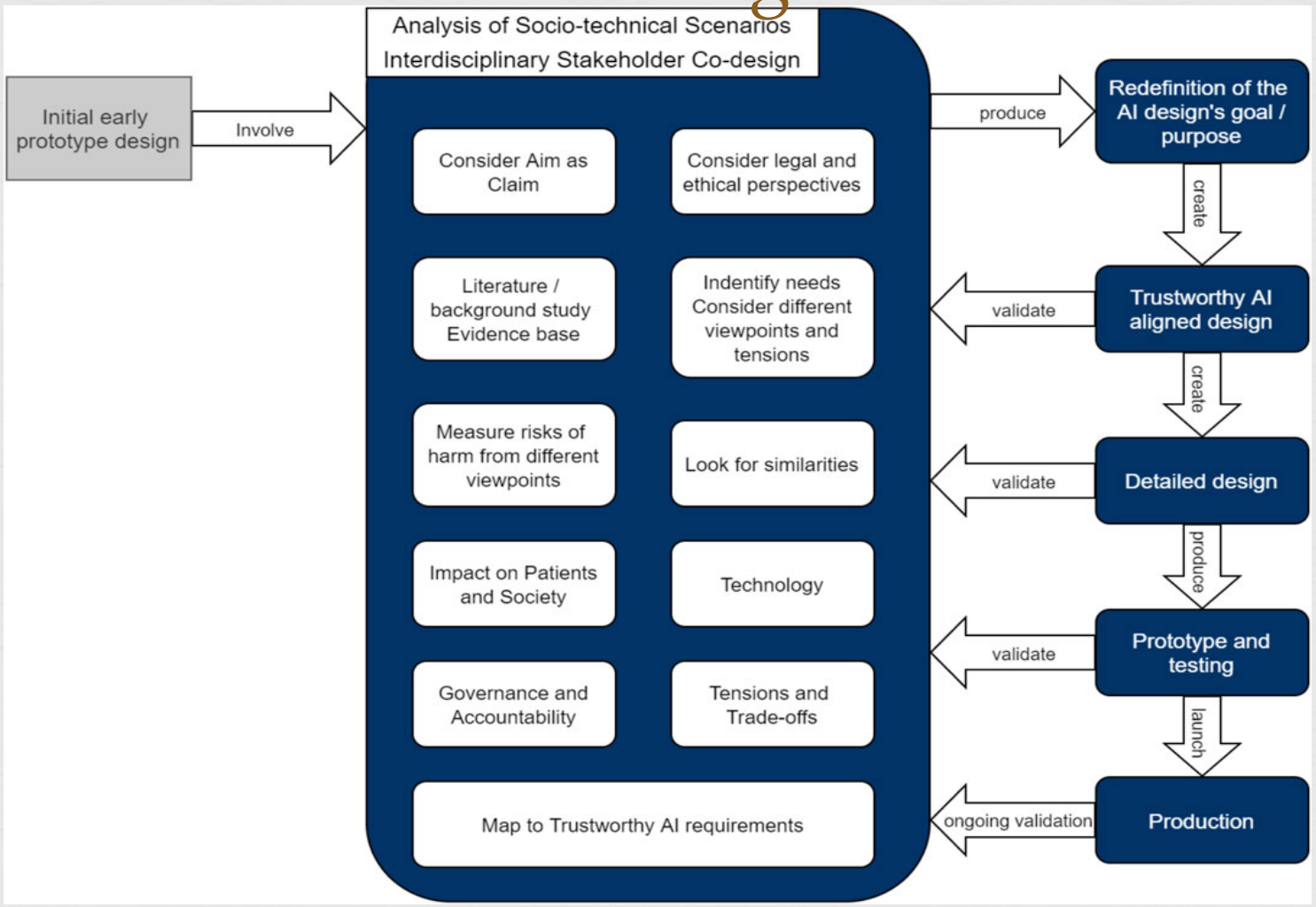
- *View Point:* **Evidence Based Medicine Professio**nal.
  There are no reliable biomarkers that can predict the prognosis of melanoma before excision. There are patients who survive their localized melanoma without therapy. **Therefore, the early diagnosis does not necessarily mean a better prognosis; on the contrary, there is a risk of poor patient care due to overdiagnosis.**

The Z-Inspection® process: Co-Design

In **design and development phases:**

Z- Inspection® can be used as a **co-creation process** to help AI engineers, domain experts to ensure that the design of their AI system meets the trustworthy AI criteria.

Co-Design


Initial early prototype design
Involve
Analysis of Socio-technical Scenarios
Interdisciplinary Stakeholder Co-design
Consider Aim as Claim
Consider legal and ethical perspectives
Literature / background study Evidence base
Indentify needs Consider different viewpoints and tensions
Measure risks of harm from different viewpoints
Look for similarities
Impact on Patients and Society
Technology
Governance and Accountability
Tensions and Trade-offs
Map to Trustworthy AI requirements
produce
Redefinition of the AI design's goal / purpose
create
Trustworthy AI aligned design
validate
create
Detailed design
validate
produce
Prototype and testing
validate
launch
Production
ongoing validation

When in Co-design.

- **Consider the AI *initial design as a Claim that needs to be verified with evidence.***

- Example: When designing, training and testing an AI-system (e.g. Machine-Learning algorithm) we do "embed" into the system notions such as "**good**", "**bad**", "**healthy**", "**disease**", etc. mostly not in an explicit/transparent way.

The Z-Inspection® process

In deployment and after deployment:

Z-Inspection® can be used as a validation process to assess the trustworthiness of the AI system being developed.

Additionally, it can form part of an AI certification, audit or **monitoring process**.

The latter can be considered a part of "*ethical maintenance*" for trustworthy AI.

When the AI is Deployed

- ***Verify Claims of the producer of the AI with Evidence***

- *Example:* "Embedded" Ethics in AI for healthcare: *Medical Diagnosis*

"Embedded" Ethics in AI for healthcare: *Medical Diagnosis*

"In case medical diagnosis or treatment recommendations are being deferred to machine learning algorithms, **it is the algorithm who sets the bar about how a disease is being defined."**

"The deployment of machine learning in medicine might resurge the debate between *naturalists* and *normativists*. "

-- Thomas Grote , Philipp Berens

Source: Grote T, Berens P.
J Med Ethics Epub ahead of print: [please include Day Month Year]. doi:10.1136/ medethics-2019-105586

Lessons Learned

There may be tensions in building a stronger evidence base on the current uses and impacts (*domain specific*)

- **Different View Points among Domain Experts**
- **Who is "qualified" to give a strong evidence?**

How to handle IP

- **Clarify *what is* and *how to handle* the *IP* of the AI and of the part of the entity/company to be examined.**

- Identify possible restrictions to the Inspection process, in this case assess the consequences (if any)

- Define if and when *Code Reviews* are needed/possible. For example, check the following preconditions (*):
  - There are no risks to the security of the system
  - Privacy of underlying data is ensured
  - No undermining of intellectual property

  Define the implications if any of the above conditions are not satisfied.

(*) Source: *"Engaging Policy Shareholders on issue in AI governance" (Google)*

Implication of IP on Trustworthy AI

- There is an inevitable trade off to be made between disclosing all activities of a trustworthy AI assessment vs. delaying them to a later stage.

Benjamin Haibe-Kains, et al. The importance of transparency and reproducibility in artificial intelligence research. (Submitted on 28 Feb 2020 (v1), last revised 7 Mar 2020 (this version, v2))
https://arxiv.org/pdf/2003.00898.pdf

Use Case

Assessing Trustworthy AI Best Practice: Machine Learning as a Supportive Tool to Recognize Cardiac Arrest

together with the Emergency Medical Dispatch Center (EMS) of the City of Copenhagen.

Health-related emergency calls (112)

- **Health-related emergency calls** (**112**) are part of the Emergency Medical Dispatch Center (EMS) of the **City of Copenhagen**, triaged by medical dispatchers (i.e., medically trained dispatchers who answer the call, e.g., nurses and paramedics) and medical control by a physician on-site (EMS).

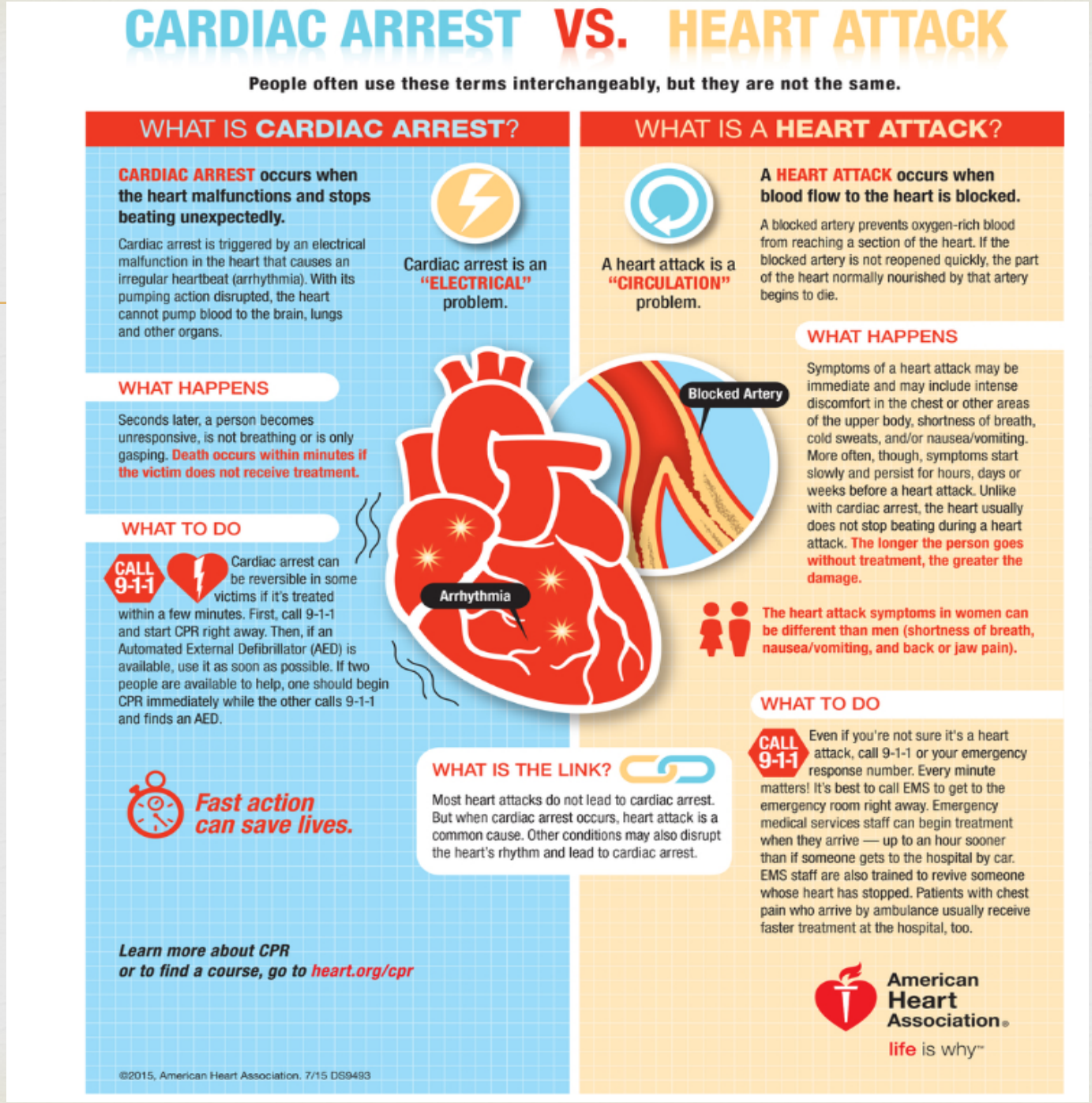
Health-related emergency calls (112)

Image https://www.expatica.com/de/healthcare/healthcare-basics/emergency-numbers-in-germany-761525/

The problem

- In the last years, the Emergency Medical Dispatch Center of the City of Copenhagen **has failed to identify approximately 25% of cases of out-of-hospital cardiac arrest (OHCA),** the last quarter has only been recognized once the paramedics/ ambulance arrives at the scene .

Image: CPR


CARDIAC ARREST VS. HEART ATTACK
People often use these terms interchangeably, but they are not the same.
WHAT IS CARDIAC ARREST?
CARDIAC ARREST occurs when the heart malfunctions and stops beating unexpectedly.
Cardiac arrest is triggered by an electrical malfunction in the heart that causes an irregular heartbeat (arrhythmia). With its pumping action disrupted, the heart cannot pump blood to the brain, lungs and other organs.
Cardiac arrest is an "ELECTRICAL" problem.
WHAT HAPPENS
Seconds later, a person becomes unresponsive, is not breathing or is only gasping. Death occurs within minutes if the victim does not receive treatment.
WHAT TO DO
CALL 9-1-1
Cardiac arrest can be reversible in some victims if it's treated within a few minutes. First, call 9-1-1 and start CPR right away. Then, if an Automated External Defibrillator (AED) is available, use it as soon as possible. If two people are available to help, one should begin CPR immediately while the other calls 9-1-1 and finds an AED.
Fast action can save lives.
Learn more about CPR or to find a course, go to heart.org/cpr
©2015, American Heart Association. 7/15 DS9493
WHAT IS A HEART ATTACK?
A HEART ATTACK occurs when blood flow to the heart is blocked.
A blocked artery prevents oxygen-rich blood from reaching a section of the heart. If the blocked artery is not reopened quickly, the part of the heart normally nourished by that artery begins to die.
A heart attack is a "CIRCULATION" problem.
Blocked Artery
Arrhythmia
WHAT HAPPENS
Symptoms of a heart attack may be immediate and may include intense discomfort in the chest or other areas of the upper body, shortness of breath, cold sweats, and/or nausea/vomiting. More often, though, symptoms start slowly and persist for hours, days or weeks before a heart attack. Unlike with cardiac arrest, the heart usually does not stop beating during a heart attack. The longer the person goes without treatment, the greater the damage.
The heart attack symptoms in women can be different than men (shortness of breath, nausea/vomiting, and back or jaw pain).
WHAT TO DO
CALL 9-1-1
Even if you're not sure it's a heart attack, call 9-1-1 or your emergency response number. Every minute matters! It's best to call EMS to get to the emergency room right away. Emergency medical services staff can begin treatment when they arrive — up to an hour sooner than if someone gets to the hospital by car. EMS staff are also trained to revive someone whose heart has stopped. Patients with chest pain who arrive by ambulance usually receive faster treatment at the hospital, too.
WHAT IS THE LINK?
Most heart attacks do not lead to cardiac arrest. But when cardiac arrest occurs, heart attack is a common cause. Other conditions may also disrupt the heart's rhythm and lead to cardiac arrest.
American Heart Association.
life is why™

The Problem (cont.)

- Therefore, the Emergency Medical Dispatch Center of **the City of Copenhagen loses the opportunity to provide the caller instructions for cardiopulmonary resuscitation (CPR), and hence, impair survival rates.**
- OHCA is a life-threatening condition that needs to be recognized rapidly by dispatchers, and recognition of OHCA by either a bystander or a dispatcher in the emergency medical dispatch center is a prerequisite for initiation of cardiopulmonary resuscitation (CPR).

Cardiopulmonary resuscitation (CPR)

Image :http://developafrika.org/compress-airways-breath-a-guide-to-performing-cardiopulmonary-resuscitation-cpr/?utm_source=ReviveOldPost&utm_medium=social&utm_campaign=ReviveOldPost

Liability

- Who is responsible is something goes wrong?
- Medical Dispatchers are liable.

The AI "solution"

- A team of medical doctors of the Emergency Medical Services Copenhagen, and the Department of Clinical Medicine, University of Copenhagen, Denmark worked together with a start-up and examined whether a **machine learning (ML) framework could be used to recognize out-of-hospital cardiac arrest (OHCA) by listening to the calls** made to the Emergency Medical Dispatch Center of the City of Copenhagen.

Context and processes, where the AI system is used

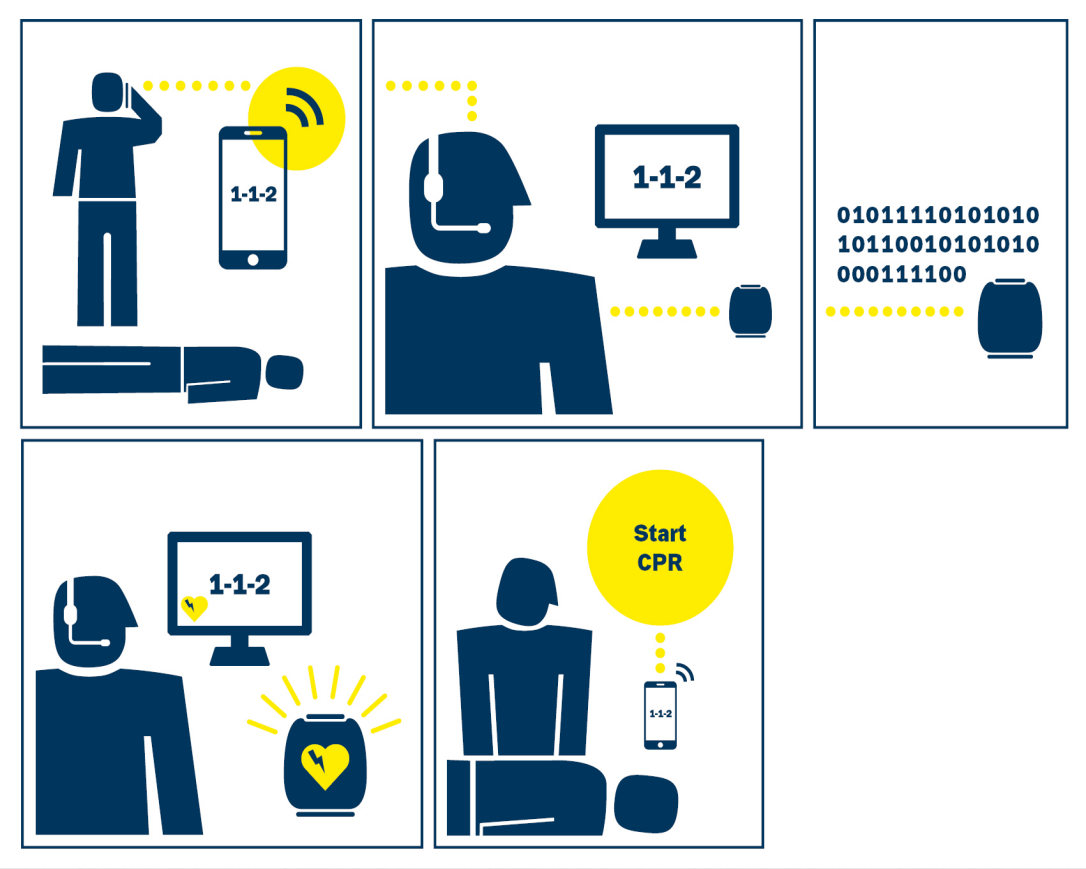


Figure . Ideal Case of Interaction between Bystander, Dispatcher, and the ML System. (with permission from Blomberg, S. N 2019b)

Retrospective study

- The AI system **performed well in a retrospective study** (AI analyzed 108,607 emergency calls audio files in 2014)

Retrospective study

- The machine learning framework had a significantly **higher sensitivity** (72.5% vs. 84.1%, $p < 0.001$) with **lower specificity** (98.8% vs. 97.3%, $p < 0.001$).
- The machine learning framework had a **lower positive predictive** value than dispatchers (20.9% vs. 33.0%, $p < 0.001$).
- **Time-to- recognition** was significantly shorter for the machine learning framework compared to the dispatchers (median 44 seconds vs. 54 s, $p < 0.001$).

- Source RESUSCITATION 138(2019)322–329 Published 2019

Randomized clinical trial

- In 2020 it was conducted a **randomized clinical** trial of 5242 emergency calls, a machine learning model listening to calls could alert the medical dispatchers in cases of suspected cardiac arrest.

Published January 2021, *JAMA Netw Open*. 2021;4(1):e2032320. doi:10.1001/jamanetworkopen.2020.32320

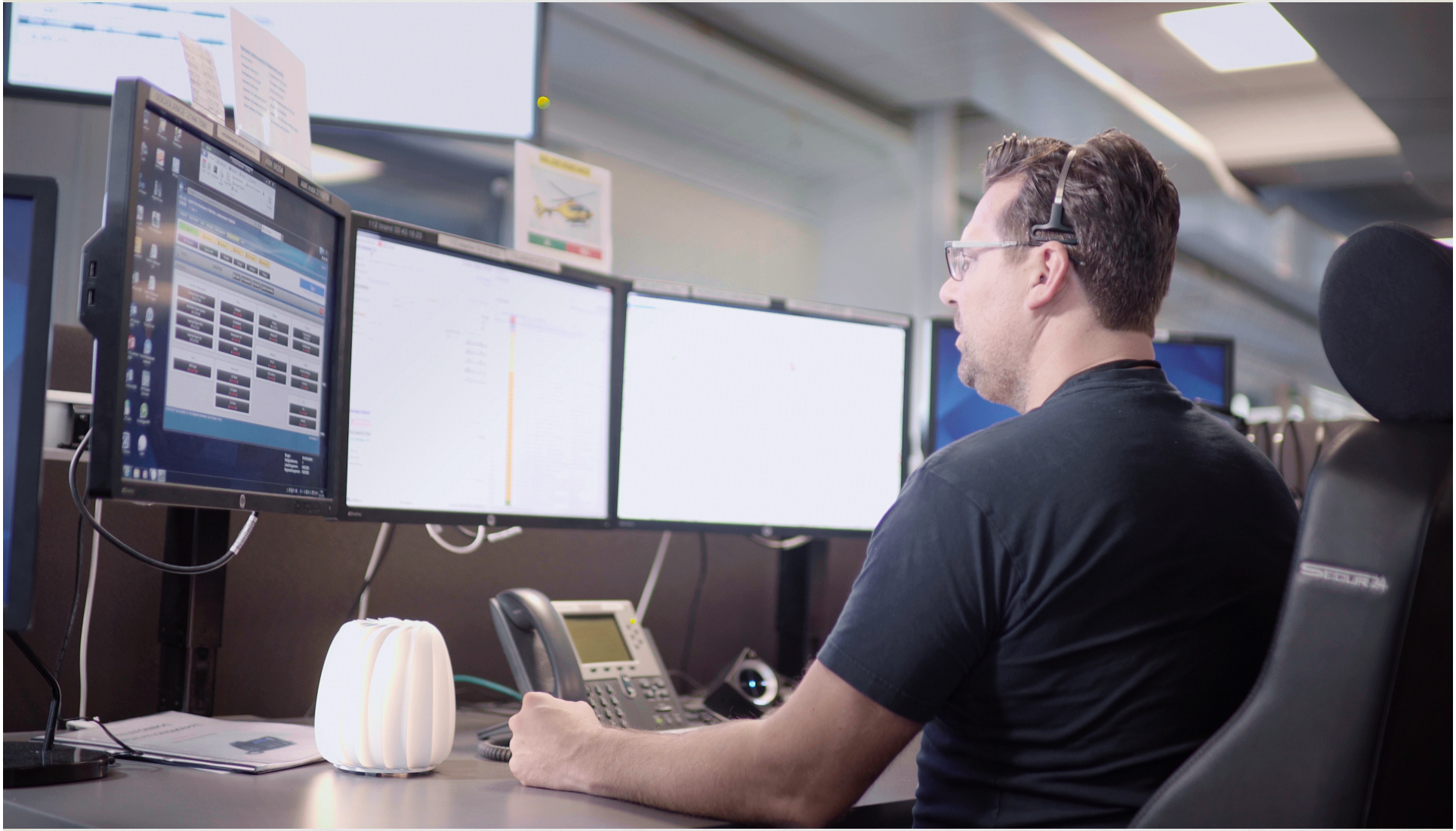
Randomized clinical trial (Cont.)

- **There was no significant improvement in recognition of out-of-hospital cardiac arrest during calls on which the model alerted dispatchers vs those on which it did not;** however, the machine learning model had higher sensitivity that dispatchers alone.

The AI system was put in production

- **The AI system was put into production during Fall 2020.**

- Note: A responsible person at the Emergency Medical Dispatch Center **authorized the use** of the AI system.



- https://cordis.europa.eu/project/id/823383/reporting

Motivation

- We agreed to conduct a *self-assessment* jointly by our team of independent experts together with the prime stakeholder of this use case.
- The main motivation of this work is to study if the rate of lives saved could be increased by using AI, and at the same time to identify ***how trustworthy is the use of the AI system*** assessed here, and to provide recommendations to key stakeholders.

Tensions in the evidence base

- There is **a tension** between:

- The conclusions from the **retrospective study** (Blomberg et al., 2019), indicating that **the ML framework performed better than emergency medical dispatchers** for identifying OHCA in emergency phone calls - and therefore with the expectation that the ML could play an important role as a decision support tool for emergency medical dispatchers- ,

Tensions in the evidence base

- and the results of **a randomized control trial** performed later (September 2018 – January 2020) (Blomberg et al., 2021), **which did not show any benefits in using the AI system in practice.**

Possible lack of trust

- For our assessment, it was important to find out **whether and how the ML system influences the interaction between the human actors**,
- i.e., how it influences the conversation between the caller/bystander and the dispatcher, the duration of the call, and the outcome, and why during the clinical trial the use of the AI system did not translate into improved cardiac arrest recognition by dispatchers (Blomberg et al. 2021).

Describing Issues

- **Use free text and an open vocabulary to describe the possible risks and issues found when analyzing the AI system.**

- The report may list the identified **ethical, technical, domain-specific (i.e. medical) and legal issues described using an open vocabulary**.

Example Ethical Issue

ID Ethical Issue: E4, Fairness in the Training Data

Description

The training data is likely not sufficient to account for relevant differences in languages, accents, and voice patterns, potentially generating unfair outcomes.

Narrative Response (*Open Vocabulary*)

- There is likely empirical bias since the tool was developed in a predominantly white Danish patient group. **It is unclear how the tool would perform in patients with accents, different ages, sex, and other specific subgroups.**

Narrative Response (*Open Vocabulary*)

- **There is also a concern that this tool is not evaluated for fairness with respect to outcomes in a variety of populations.** Given the reliance on transcripts, non-native speakers of Danish may not have the same outcome. It was reported that Swedish and English speakers were well represented but would need to ensure a broad training setmay not have a diverse enough representation.

Narrative Response (*Open Vocabulary*)

- It would also be important to see if analyses show any **bias** in results regarding age, gender, race, nationality, and other sub-groups. The concern is that the training data

Example

Possible Risks and Harm: False Positives and False Negatives

We could not find a justification for choosing a certain balance between sensitivity and specificity.

- If *specificity* is too low, CPR is started on people who do not need it and administered CPR over a longer period of time can lead to rib cage fractures, for example. However, it is unlikely that CPR would be performed on a conscious patient for a longer time, as the patient probably would fight back against it.
- If *sensitivity* is too low, cardiac arrests may not be detected. This results in no CPR being administered and the patient remains dead. In this context "too low" is when the AI system performs poorer than the dispatchers, hence will not be of any help. The AI system is evaluated against human performance, as this system is only useful if it can assist humans; otherwise, it is just a distraction.

Example

Lack of Explainability

- Our team of experts did not sign a Non Disclosure Agreement (NDA) with the vendor company, and that means that the AI system is considered a “black box,” with no details of the implementation of the AI algorithms and the AI model. To avoid possible conflict of interests, no direct communication between our team of experts and the vendor company was (and is) taking place.
- The prime stakeholder cooperates with the vendor company, and they have declared no conflict of interest with them.
- The main issue here is that it is not apparent to the dispatchers how the AI system comes to its conclusions. It is not transparent to the dispatcher whether it is advisable to follow the system or not. Moreover, it is not transparent to the caller that an AI system is used in the process.