

Assurance of Software-Intensive Medical Devices: What About Mental Harm?



Barbara Gallina, Associate professor of dependable software systems
Group founder/leader, Certifiable Evidences & Justification Engineering
barbara.gallina@mdu.se

Based on: J. L. de la Vara, **B. Gallina**, A. Fernández-Caballero , J. Pascual Molina , A. S. García , C. Ayora.

Assurance of Software-Intensive Medical Devices: What About Mental Harm?.

53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Porto, Portugal, June 27-30, 2023.



- **What is assurance?**
- **What is harm?**
- **What about mental harm?**
- **Mental Health**

- **Context and problem**
- **Inspiration**
- **Goal/Research roadmap**

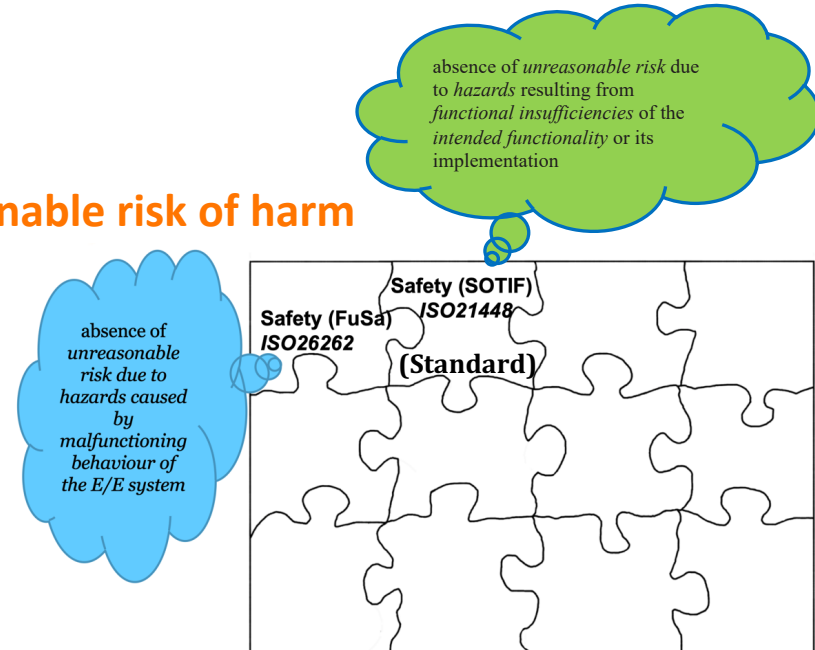
Assurance “grounds for justified confidence that a claim has been or will be achieved”

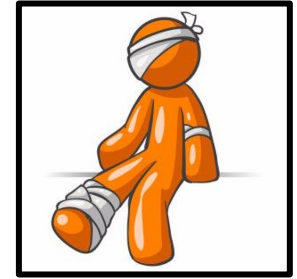
[ISO/IEC JTC 1/SC 7, ISO/IEC 15026: Systems and software engineering — Systems and software assurance, Part 1: Concepts and vocabulary (2019)]

Assurance process: set of **activities to provide justified confidence** that a system satisfies given requirements, e.g., for system safety, usually in compliance with standards

Claim about what?

The deployment of system does not pose an **unreasonable risk of harm** in a given context





Harm “Physical injury or damage to the health of persons”

[ISO/TC 22/SC 32, ISO 26262-1:2018 – Road vehicles Functional safety]

[ISO/TC 22/SC 32, SO 21448:2022 Road vehicles Safety of the intended functionality]

1

Abbreviated Injury Scale (AIS)

Note: created to enable an international comparison of severity

- AIS 0: no injuries
- AIS 1: light injuries such as skin-deep wounds
- AIS 2: moderate injuries such as deep flesh wounds
- AIS 3: severe but not life-threatening injuries such as skull fracture without brain injury
- AIS 4: severe life-threatening injuries such as concussion with up to 12 hours of unconsciousness
- AIS 5: critical injuries such as spinal fractures
- AIS 6: extremely critical or fatal injuries such as extremely critical open wounds of body cavities



SAE J3016™ LEVELS OF DRIVING AUTOMATION™

Learn more here: [sae.org/standards/content/j3016_202104](https://www.sae.org/standards/content/j3016_202104)

Copyright © 2021 SAE International. The summary table may be freely copied and distributed AS-IS provided that SAE International is acknowledged as the source of the content.



Assumptions on controllability



Harm

**“ Injury or damage to the health of people, or damage to property or the environment
[SOURCE:ISO/IEC Guide 63:2019, 3.1]”**

[ISO/TC 210 and IEC/SC 62A, ISO 14971:2019 – Application of risk management to **medical devices**]

Examples: serious burns, heart fibrillation, infarct, brain damage, minor organ damage, decreased consciousness (insulin related)



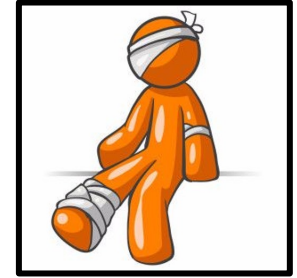
Harm injury or damage to health;

[REGULATION (EU) 2023/1230 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 June 2023]

Harm injury or damage to health;

[CEN/TC 114, ISO 12100:2010 Safety of machinery General principles for design Risk assessment and risk reduction]

Based on the state of practice, the focus is on Physical injury



The focus is on *potential contact events*

[ISO/TS 15066:2016 Robots and robotic devices – Collaborative robots]

Based on the state of practice, the focus is on Physical injury

What is health?
Shall we care only about physical health?

What about mental health?
What about mental harm?

“**Mental health** is a state of mental well-being that enables people to
-cope with the stresses of life,
-realize their abilities,
-learn well and work well,
-and contribute to their community.”



Mental health is a basic **human right**

Mental health is the **absence of mental health conditions**, which include:
mental disorders (e.g., autism spectrum disorder-ASD, major depression disorder-MDD, etc.)
and psychosocial disabilities as well as
other mental states associated with significant distress,
impairment in functioning, or risk of self-harm.

injury or damage to the mental health of people

worsening of a mental condition

emergence of a mental condition

...

Context and problem

Mental disorders



World Health
Organization

8 June 2022

Key facts

1 in every 8 people in the world live with a mental disorder

- Mental disorders involve significant disturbances in thinking, emotional regulation, or behaviour
- There are many different types of mental disorders
- Effective prevention and treatment options exist
- Most people do not have access to effective care

In 2019, 1 in every 8 people, or 970 million people around the world were living with a mental disorder, with anxiety and depressive disorders the most common (1). In 2020, the number of people living with anxiety and depressive disorders rose significantly because of the COVID-19 pandemic. Initial estimates show a 26% and 28% increase respectively for anxiety and major depressive disorders in just one year (2). While effective prevention and treatment options exist, most people with mental disorders do not have access to effective care. Many people also experience stigma, discrimination and violations of human rights.



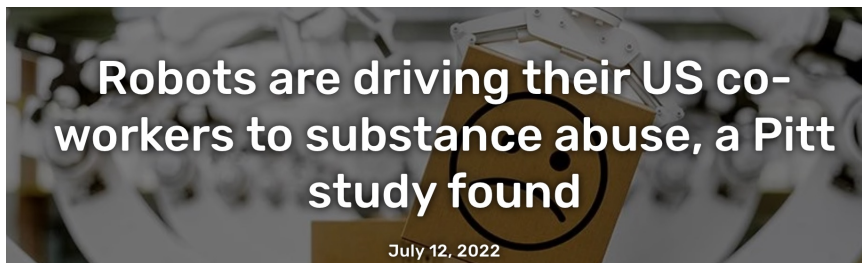
Robots replace humans



Mental health conditions may increase (anxiety, depression, etc.)
Robots may harm workers' mental health
and act as an additional workplace stressor.

<https://sciencenews.dk/en/new-research-automation-is-harmful-to-the-mental-health-of-its-human-co-workers>

Physical injuries decrease



Robots collaborate with humans ..

“Risks to workers’ mental health can also arise. One overarching cause of mental harm to workers is identified with their interaction with robots, expressed, for instance, as **frustration at robots’ non-sentience** and **inability to “understand each other” and accommodate mood change**. However, there are also risks from the converse, when robots are too sentient: **alienation and loneliness may surface** when robots understand and adapt so well that they are prematurely “humanised” but later fail to deliver the same responses of a human for more complex tasks or in emotional terms.”



Encoding the Enforcement of Safety Standards into Smart Robots to Harness Their Computing Sophistication and Collaborative Potential: A Legal Risk Assessment for European Union Policymakers

Published online by Cambridge University Press: 06 November 2023



SAE J3016™ LEVELS OF DRIVING AUTOMATION™

Learn more here: sae.org/standards/content/j3016_202104

Copyright © 2021 SAE International. The summary table may be freely copied and distributed AS-IS provided that SAE International is acknowledged as the source of the content.



Automatophobia, situational type of phobia..shall we consider it?

Controllability

Do we worsen the situational phopia?

doi: [10.1016/j.ijchp.2023.100371](https://doi.org/10.1016/j.ijchp.2023.100371)

Married father kills himself after talking to AI chatbot for six weeks about his climate change fears

31 March 2023

- The man reportedly found comfort in talking to the AI chatbot named 'Eliza'
- For confidential support call the Samaritans on 116 123 or go to [samaritans.org](https://www.samaritans.org)



Children with autism have been found to be more comfortable communicating with robots

*However, attention shall be paid to reduce
Risks of psychological attacks conducted by household robots*

Tony Belpaeme,
Jorre Deschuyteneer,
Djamari Oetringer,
Pieter Wolfertt

IDLab – imec, Ghent
University, Belgium

The potential of social robots for persuasion and manipulation: a proof of concept study



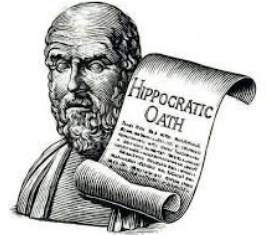
https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2019/10/14081257/Robots_social_impact_eng.pdf

**At the time of writing,
we were not aware of any publication on system assurance that
had addressed mental harm explicitly and in detail
(we focused on medical devices)**

In the context of medical devices, we are still not aware..

Hippocratic Oath -5th century B.C.

→ expression of ideal conduct for the physician



<https://www.herald.co.zw/the-hippocratic-oath-and-the-death-of-national-conscience/>

“I will use treatment to help the sick according to my ability and judgment, but never with a view to injury and wrong-doing.

...

I will utterly reject harm and mischief”

A promise said out loud

Extended Hippocratic Oath WANTED! Deontology wanted!

Principle of Non-Maleficence..

We aim at providing **new insights into the emerging idea of why and how mental harm** should be considered for assurance of software-intensive medical devices

Our ultimate goal is to **increase the awareness of the need to address mental harm for system assurance**

Hazard and Risk Analysis

System Compliance

System Dependability Justification

Assurance Evidence Collection

Needs & Solutions

Examples of systems

Emotion recognition based on physiological, perceptual, and behavioural responses

These systems can be used for SCZ, BPD, or MDD

XR for emotion induction and recognition

These systems can be used for SCZ, BPD, or MDD

Emotion induction and recognition using social robots

These systems can be used for ASD



[https://commons.wikimedia.org/wiki/
File:Osuri_socially_interactive_robot_platform_IMG_20200219_165823219_02.jpg](https://commons.wikimedia.org/wiki/File:Osuri_socially_interactive_robot_platform_IMG_20200219_165823219_02.jpg)

Sources of possible mental harm

Negative impact, no detection, therapy deviation, reluctance...

Emotion induction as aspect to especially pay attention to

Device-specific aspects

Functionality & expected usage

Requirements & input from ethics committees

Medical COTS consequences

Differences among patients

Multi-concern assurance

HCI failures characterisation

AI trustworthiness

“medical device software shall be developed in accordance with the state of the art”

(EU reg. EU Regulation 2017/745)

IEC 62304, ISO 14971...

Basis: injuries are possible

Frameworks of compliance with (existing) standards

Definition of who will regulate mental health technology and the data that it generates

Clear standards & guidance not only for clinical studies

Some guidance already by FDA but limited

injury and ill health

“adverse effect on the physical, **mental or cognitive condition of a person”**

[ISO/PC 283, ISO 45001:2018 **Occupational health** and safety management systems]

An organization is responsible for the occupational health and safety of workers and others who can be affected by its activities. This responsibility includes promoting and protecting their physical and mental health.

System Compliance

OUR PROPOSAL AS PER THE PRESENT PAPER (comprehensive Regulation, integrating cobotics and AI aspects)

Closer attention to mental aspects (as also bodily expressed), including an emphasis on neurodiversity, psychological comfort, multifactorial workplace stress, and complex neuropsychiatric disorders.

The safety of the cobot and that of the coworker relationally depend on one another.

Encoded OHS supervision is feasible, so long as last-resort commanding capacity is retained by humans.

Disclosure of more context-sensitive and integrated information that accounts for the environment, the cobotic-human relationship, as well as all mind-body interfaces.

"Dangerousness" lexicon.

Each machine, and recurrently over time (to cater for developments in machine learning).

Open-access source-code registries to be established, but balance to be found with trade-secret protection, and mindful of limited value of disclosing codes per se.

Public auditing at regular intervals – but aimed at raising awareness, as opposed to a mere fining approach.

Dynamic, "ecosystem" approach, which includes diverse human coworkers within different typologies of working environments.

Specific provisions for testing in VR environments.

For the lack of programming restrains of algorithmic learning, in the event such shortcoming (co-)causes safety failures.

Three typologies of disputes: 1) cobot vs cobot; 2) cobot vs human; 3) *cobot+human* team(s) vs other *cobot+human* team(s).

Binding also *de iure*, and systematically integrated (or at least referenced) within the Regulation.

Robotic safety and AI addressed together meaningfully, at least insofar as smart cobots are concerned.

Preliminary considerations on quantum technologies are warranted.

Also to keep pace with automation developments in the Asia-Pacific (especially Mainland China, South Korea, Japan, and India) and the US.

NEW MACHINERY REGULATION (May 2023)

Mental aspects are still encompassing mere situations of 'cognitive stress, discomfort, and fatigue from a traditional ergonomics perspective [Annex III, 1.1.6], without delving deeper to neurodiversity and complex neuropsychiatric disorders.

Cobotics is not specifically addressed.

progress in relation to machine supervisory functions [Annex 4], but no specific foreseen OHS supervision by (smart) cobots.

No deviation from the traditional approach.

The lexicon itself has mostly shifted towards "dangerousness", it conceptually resembles the old-fashioned "defectiveness" as opposed to emphasising AI's structural unforeseeability.

Reiteration of the "sampling" paradigm [Article 10.4-5], but no specific accommodation of machine-learning demands.

In-passing mention of registries [Article 10.4], but unclear specification of when they should be established, and no reference to algorithms' source-code.

Proposed introduction of regulator audits [Annex IX, 3.3-4.3], but not aimed at awareness-building and never public.

Reiteration of the traditionally static component-by-component (or at best machine-overall) approach.

Still no provisions for virtual reality.

Issue for safety outcomes as such. The issue we flag up is rightly addressed here, too [Preambulatory Clause 12], but not followed-up.

No specific recommendation.

Issue is extensively commented upon in this new Regulation [eg Preambulatory Clauses 44-73]. However, the place of technical standards within the formal hierarchy of EU law sources remains unclear. One may postulate that standards are properly referenced procedurally analysed in this new piece of legislation, while still formal systematisation and integration within the overall legal systems of the EU and its Member States. Moreover, the discussion remains high-level, with no pointing to specific standards and standard-setting bodies in relation to each regulated activity.

Improper integration between the entire discourse around AI and the specific safety and liability provisions on machines.

Reference to quantum computers or other quantum technologies. Again exclusively focused on the EU's internal market [Preambulatory Clauses 1-2].

Encoding the Enforcement of Safety Standards into Smart Robots to Harness Their Computing Sophistication and Collaborative Potential: A Legal Risk Assessment for European Union Policymakers

Published online by Cambridge University Press: 06 November 2023

	OVERALL TRADITIONAL FRAMEWORK (bundle of Directives + proposed Regulation)
MENTAL-PHYSICAL SAFETY CONTINUUM, SAFE COLLABORATION, SAFETY SUPERVISION, AND SAFETY DISCLOSURE	Focus on "physical" aspects from perspective + "stress per se". Reiterates simplistic dichotomies between "mental" and "physical" aspects.
	Rationale is that the robot should not be considered a "stressor".
	No foreseen OHS supervision by humans.
	Disclosure of mainly mechanical-safety aspects to external (outsourced) safety experts.
QUALITY COMPLIANCE IN MANUFACTURING	"Defectiveness" lexicon.
AUDITING FREQUENCY AND TRANSPARENCY	Sample machines, and on-site audits.
	No specific provision on open-access to algorithm programming requirements.
TESTING	Mostly self-certification.
	Static, component-by-component (or at best machine-overall) approach.
LIABILITY	No specific provisions for virtual reality.
DISPUTE PREVENTION, HANDLING, AND RESOLUTION	For safety outcomes as such.
EUROPEAN AND INTERNATIONAL INDUSTRY STANDARDS	No specific recommendations.
	<i>De facto</i> binding, and not explicitly or referenced within the Regulation.
RELATIONSHIP BETWEEN POLICY EFFORTS ON AI AND ROBOTICS	Robotic safety and AI within legislative texts/proposals.
QUANTUM COMPUTING	No reference to quantum computing.
GEOPOLITICAL COLLOCATION	For further integrating into global context.

System Dependability/Safety Justification

Assurance cases & different argument types

product- & process-based...

Evolution through device lifecycle

Ethical aspects

Argument patterns

Inc. for ML

Proven-in-use arguments

Past med. COTS performance

Argument modules

Most relevant artefact types

Process information seems especially important

Assurance evidence evolution

Sources of evidence & trustworthiness

Healthcare ones, not only mostly engineering results

Characterisation of assurance evidence for mental harm

patient characterisation & training characterisation aspects

Evidence from different healthcare stakeholder & tasks

Patients, mental health professionals, clinical studies...

Evidence assessment

- **Mental harm should be considered in the assurance of many software-intensive medical devices .. as well as other systems..**
- **How to deal with it needs to be determined, specialising certain **healthcare** aspects for mental harm**
- **There exist four main areas to address:**
 - Hazard and Risks Analysis,**
 - System Compliance,**
 - System Dependability Justification & Assurance**
 - Evidence Collection**
- **System developers and other stakeholders must be aware of the areas and pay attention to them**
- **We plan to define specific assurance means**

**Hope it was interesting!
Thank you very much for your attention!
...discussion time...**

**Interested in collaborating?
Drop me an e-mail:
barbara.gallina@mdu.se**