An ethical framework for human-robot collaboration for the future people-centric manufacturing

Callari, Tiziana C.; Vecellio Segate, Riccardo; Hubbard, Ella Mae; Daly, Angela; Lohse, Niels

Published in:
Technology in Society

DOI:
10.1016/j.techsoc.2024.102680

Publication date:
2024

Technology in Society 78 (2024) 102680

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

An ethical framework for human-robot collaboration for the future people-centric manufacturing: A collaborative endeavour with European subject-matter experts in ethics

Tiziana C. Callari [a,b,*], Riccardo Vecellio Segate [c,d], Ella-Mae Hubbard [a], Angela Daly [c], Niels Lohse [a,b]

[a] Intelligent Automation Centre, Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, UK
[b] College of Engineering and Physical Sciences, School of Engineering, Department of Mechanical Engineering, University of Birmingham, Birmingham, UK
[c] Leverhulme Research Centre for Forensic Science, School of Science and Engineering & School of Law, University of Dundee, Dundee, UK
[d] Faculty of Economics and Business, University of Groningen, Groningen, the Netherlands

ARTICLE INFO

Keywords:
Qualitative Delphi
Collaborative robots
Ethics
Human-robot
Resilient operator
Industry 5.0
Roboethics

ABSTRACT

Envisioning humans and (smart) robots collaboratively working on the manufacturing shop floor, sharing spaces, tasks and objectives, reflects the ambitious goal that the ideal factory of the future aspires to attain. However, ensuring the effective implementation of this novel form of labour organisation remains an ongoing area of research. Key aspects such as the future role of workers, potential psychological risks, and the overall ethical considerations of human-robot (H-R) collaboration warrant further investigation until the underpinning safety challenges have been addressed. This study presents a novel ethical framework for H-R collaboration in manufacturing, which involved 30 subject-matter experts in ethics within the European context in a collaborative design process conducted through a year-long three-round data collection qualitative Delphi study. The ethical framework adopts a human-centric approach, recognising the influences that expand beyond the specific context of H-R dynamics on the shop floor, towards organisational and societal governance for a more responsible integration of (smart) robotics into the professional settings. Ethics, in this regard, aims to foster ethical awareness and accountability in the processes and practices of design and innovation, involving all stakeholders who play a role in shaping the future of Industry 5.0.

1. Introduction

Within the manufacturing sector, the pursuit for higher productivity, enhanced quality, and increased flexibility has long been a driving force behind technological innovation and digital transformation. This drive has culminated in what is now known as the Industrial Internet of Things (IIoT) or Industry 4.0 [1,2]. Within this transformative era, the integration of humans and robots collaborating as a mixed team on the shop floor holds the potential to unlock many opportunities and strategies for addressing unforeseen challenges that would otherwise remain inaccessible when each is operating individually [3]. For instance, humans are much better at responding to changes, using their wide perception to decode the environment in which they work, reasoning about uncertainty, and of course harnessing their creativity. On the other hand, robots excel in performing repetitive tasks consistently, lifting large loads, executing vast computations rapidly, and operating within hazardous environments.

Currently, at the forefront of research, the concept of human-robot (H-R) collaboration is being debated across disciplines, underscoring the interdisciplinary nature of ongoing efforts to make this novel collaboration a reality. In contrast to the traditional manufacturing factories, where automation has been confined within physical barriers and separated from humans primarily for safety measures, current research is exploring avenues to steadily integrate smaller industrial robots into collaborative manufacturing environments, to interact and work alongside human operators [4]. These collaborative robots currently have limitations in their load-bearing (payload), torque, and speed with only limited, if any, degrees of autonomy (Machine Learning

* Corresponding author. Intelligent Automation Centre, Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, UK.
E-mail address: T.Callari@lboro.ac.uk (T.C. Callari).

https://doi.org/10.1016/j.techsoc.2024.102680
Received 29 March 2024; Received in revised form 14 June 2024; Accepted 7 August 2024
Available online 14 August 2024

(ML)/Artificial Intelligence (AI)-enabled robots show some levels of adaptivity) to perform different tasks throughout the collaborative manufacturing process [4,5]. This is a reflection of their embedded safety systems [6–8] that enable them to stop or reduce their speed to prevent unintended harm to human operators within a designated collaborative H-R workspace [9]. The use of larger robots and higher degrees of autonomy is currently being explored and developed in engineering and computer science labs to expand the range of possible collaborations. However, this also presents new challenges for the safety and wellbeing of human operators, necessitating enhanced safety protocols on the production floor, perception sensors and cameras to monitor H-R interaction and trigger safety stops in risky situations [10–12].

Yet, while I.4.0 has been primarily concerned with the digital transformation occurring in manufacturing environments, the impact of these novel industrial changes on workers has been largely overlooked [13]. There is indeed a pressing need for a new paradigm in manufacturing embedding more human-centric, sustainable, and resilient considerations [14,15]. This new revolution is termed Industry 5.0 (I.5.0) [16]. In I.5.0, adopting a human-centric approach means that technologies will be designed to leverage the skills and knowledge (both tacit and explicit) that humans bring, rather than striving for complete automation solutions that would replace them [17]. This signifies not just a technological revolution but a societal evolution, to the extent that Society 5.0 has already been coined [18]. Society 5.0 is a vision; a vision of a 'super-smart society' that, guided by scientific and technological innovation, is able to address and solve societal challenges, inequalities, and usher in an era of unprecedented efficiency, sustainability, and wellbeing for all its members.

1.1. Motivation

As we stand on the precipice of this change, ethics play a crucial role in ensuring that the choices and decisions that are made in shaping Society 5.0 prioritise the needs of humans, communities, and society at large, safeguarding against unintended—and unethical—risks and hazards that could compromise our personal and professional lives. At this point, we note that we use 'ethics' not in the narrower sense of ethics as a non-binding alternative to binding laws (e.g. Ref. [19], but rather in the broad and general sense of ethics as relating to morality and virtue [20], and encompassing binding laws and industrial standards in principle – although these are not always 'ethical' in practice [21].

I.4.0 and I.5.0 technologies will alter the very fabric of work—how work is performed, managed, and regulated. Although these technologies will potentially support more flexible and worker-friendly environments, there is concern that they could lead to rigid and exploitative working conditions. The critical issue is the concern that *technological determinism* might shape social structures and cultural values, influencing the future of society and industry [22]. Instead, society/industry can and should set the terms for how these technologies are used, ensuring they contribute positively to workers' rights, equity, and overall wellbeing. Ethical frameworks serve this purpose: to influence future regulations, drive design and innovation, and offer principles that guide both individual and organisational behaviour and governance towards outcomes that align explicitly with ethical and social values, as well as responsibility for sustainable and environmental impacts while embracing the I.5.0 paradigm. Ethics, in this regard, should not be reduced to a mere checklist for industries to follow. Instead, it should foster awareness and accountability in the application and practices of H-R collaboration, such as the duty of designing meaningful work for workers, or the respect of human dignity and autonomy in H-R collaboration tasks.

How humans and robots will interact with each other has emerged as a core concern in the 'Robot Ethics' scientific community [23–29]. Research concerns have revolved around how to integrate ethical systems into robots (such as, built-in ethics in self-driving cars) [30], what ethical principles should guide designers when building robots (for example, those used in military applications), or in the way we should treat robots [31]. Further debate in the literature concerns whether robots themselves should have rights (e.g. Ref. [32]). However, as this is unlikely to occur in the near to medium term, we do not address this issue in more detail in our work, which is focussed on that timeframe. In these examples, ethical principles can encompass pre-existing legal frameworks and industrial standards, and can extend beyond them, including to recommending new interventions from legislators and/or technical communities. Further consideration of the relationship between ethics, law, and standards in a broad sense is beyond the scope of this paper; however, in our findings, we highlight where certain ethical principles or interventions overlap with laws and standards.

One of the first publications to provide guidance to designers, manufacturers, and users of robot systems for a sustainable development and employment of robotics was the EURON Roboethics Roadmap in 2006 [25,33]. Roboethics principles and issues were considered to shed light on the potential ethical challenges that may arise in the relationship between humans and technology. Based on the agentic nature exhibited by robots, four types of Roboethics have been proposed [26, 34]. Following this, in 2011 the EPSRC Principles of Robotics were introduced, encompassing a collection of rules and overarching messages, serving as a foundation for promoting responsible robotics for designers, builders and users of robots [35–37]. These principles have inspired specialised ethical frameworks to address the unique challenges and concerns that may arise in each domain of application (such as in social and healthcare settings, or for military purposes [38].

However, there has been a lack of substantial discussion within the context of manufacturing. As researchers, we wondered why this was the case, questioning if the future I.5.0 would require a heightened focus on responsible ethical design and innovation throughout the manufacturing processes, practices, and artefacts. Notably, in the evolving definition of the human workforce's role in I.5.0, ethical considerations take on a distinctive character, such as the nature of H-R agency on the shop floor [39], or the psychological impact that the novel H-R work design and expectation may have on workers (S. R. [40]).

1.2. Scope and objective

In order to address these issues, we therefore asked:

- ***How can ethical considerations be effectively integrated into the implementation of a human-centric design approach for addressing the specific challenges of H-R collaboration in the manufacturing industry?***

Aligned with the argument proposed by van Wynsberghe and Li [41], ethical considerations and implications extend beyond the specific H-R dynamics on the shop floor; indeed, they encompass the inter-dependent roles of organisations and society to support and facilitate this paradigmatic transition.

- ***What are the implications for responsibly shaping the future of organisational and societal governance?***

This article offers an original contribution by advancing a novel framework to guide the ethical design implementation of H-R collaboration for future people-centric manufacturing. To accomplish this, we engaged in a collaborative dialogue with subject-matter experts in ethics of ML/AI, technology, and robotics, within the European context. Given that this marks the first ethical framework addressing the unique challenges of I.5.0 using field-research data, at this stage we chose to involve participants primarily from an academic background to guide us in framing and outlining an ethical framework for H-R collaboration in the manufacturing sector. Through a year-long three-round data collection qualitative Delphi approach, our experts suggested points, provided feedback on their peers' inputs, refined ethical themes and emerging candidate principles, towards a consolidated ethical framework that we present in the results below. Although the ethical framework investigated ethical considerations at all three levels–micro (i.e. H-R interaction and collaboration on the shop floor), meso (i.e. at organisational level) and macro (i.e. at societal level)–this paper will only present the results of ethical principles at the micro and meso-levels, as the ethical considerations within the macro-level will require further refinement beyond the scope of this Delphi study. Furthermore, while we acknowledge that in the future collaborative robots will embed higher degrees of 'intelligence', this framework focuses on the H-R physical interaction and collaboration specifically.

1.3. Theoretical underpinnings

Technological advancements provide tools and opportunities but do not inherently dictate how work should be organised. The implementation of smart technologies (such as AI or collaborative robots) has the unprecedented capacity to reshape our future society (e.g. Ref. [42,43]). Such implementation must also consider organisational structures and labour rights regulations which will depend on underlying ethical business decisions. In this section, our objective is to outline the theoretical foundation of our research.

Roboethics and machine ethics provide a foundation for outlining the ethical requirements for the future implementation of industrial collaborative robots in H-R collaboration. These frameworks also reference the United Nations' Universal Declaration of Human Rights (UNDHR) [44] and the Charter of Fundamental Rights of the European Union [45] as a shared basis for developing the ethical requirements of human-centric manufacturing in the context of H-R collaboration. Roboethics involves applied forms of ethics [46,47]. Applied ethics is a branch of ethics that focuses on the practical application of moral principles to specific issues and contexts. It combines consequential approaches (which evaluates the morality of actions based on their outcomes), and nonconsequential approaches (which includes deontological ethics and virtue ethics), and considers actions that are motivated by factors, such as duties, rights, and moral principles) [48–50]. The EURON Roboethics Roadmap has incorporated principles from UNDHR and consequential approaches from applied ethics to guide the development and implementation of ethical standards for robotics (e.g. reviewing concepts such as the humanisation of the human-machine relationship, with cognitive and affective bonds towards machine; the environmental impact of technology; design for all; etc.) [34].

Within this school of thought, and in line with work by Ashok and colleagues [51], Roboethics lies at the intersection of techno-ethics and ML/AI ethics. Techno-ethics, alongside European and international guidelines on H-M (human-machine) and H-R interaction, examines the ethics in the development of new technologies, how these technologies behave, and how they alter the power dynamics din interactions with humans. In the context of H-R collaboration, this assumes a distinctive significance in concerns pertaining to human autonomy, agency, accountability, and power in the specific context of H-R collaboration on the shop floor. Indeed, these themes raise pivotal questions about the extent of control and decision-making authority retained by human workers in tasks performed in collaboration with robots, as well as humans' capacity to exercise choice and influence within the collaborative process. ML/AI ethics concerns include responsibility, transparency, auditability, incorruptibility, predictability, and the morality of machines [52]. In the European Union (EU), there are legislative proposals for further regulation of AI, which will impact upon H-R relations when the robots embed AI capabilities. Critically, the EU AI Act regulates AI applications based on the level of risk they pose but does not include liability rules [53], which is addressed in a separate proposal for an AI Liability Directive [54]. In the future, with AI-embedded robots gaining greater capabilities, the distribution of responsibility within complex socio-technical systems will become a key focus for the law, driven by the imperative of preventing harm caused by robots and ensuring accountability for human actions [55,56].

2. Methodology

2.1. Rationale of the qualitative Delphi approach

The Delphi method in its qualitative variation [57] was employed to gather expert insights and judgments regarding emerging relatively underexplored ethical concerns in H-R collaboration in manufacturing. The Delphi method is characterised by several key aspects: (1) it collects diverse perspectives and feedback from domain experts, who are invited to respond to questions related to the study's problem; (2) the data collection typically consists of multiple rounds, usually ranging between three to five sequential questionnaires. The design of the first questionnaire is often informed by a literature review or *ad-hoc* workshops defining the study scope; (3) the information gathered undergoes iterative assessment and review by the domain expert; and (4) the results tend to offer expert-derived solutions to the proposed problem [58–61]. Additionally, the Delphi method involves independent individuals, such as researchers, who are tasked to synthesising, summarising and sharing each iteration's interim results back to the participants until a consensus is achieved [62].

In the qualitative modifications, the Delphi method's key features such as experts' anonymity, iterative questionnaires, and researcher-controlled feedback are retained. However, instead of structured questions, open-ended and qualitative prompts are preferred, to allow participants to freely comment on emerging themes and collective perspectives [57]. This qualitative process concludes when the domain experts' responses exhibit a convergent trend that addresses the study's objective(s), or when sufficient information has been exchanged to attain information 'richness' (i.e. the "*information power*") [63,64] on newly developed themes [65].

The decision to employ the qualitative Delphi approach was based on several factors. First, considering the ethical challenges posed by the novel work configuration of H-R collaboration, we required a methodology able to support the collection of diverse expert opinions and insights, and support iterative revisions to align participant responses with our research objective. Central to this was the identification of subject-matter experts who could contribute fresh knowledge and ethical perspectives to both the overarching and context-specific ethical considerations. Emphasis was placed on recruiting experts actively engaged in research and discourse within the field, while preserving respondent anonymity to ensure that our selected participants felt unrestrained in expressing their views without the influence of known colleagues—indeed, anonymity is a key-feature of the Delphi methodology.

2.2. Study design

A review of the literature surrounding the investigated ethical concerns in H-R systems was initially performed to inform the design of our study. This study involved a three-round data collection process using the *onlinesurvey.com* platform. This took place between February and October 2023 (Table 1). The recruitment process involved subject-matter experts in ethics with a specialisation in technology, AI, and robotics, identified based on their reputation in the field.

Table 1
Data collection design.

Questionnaire #	Expert #	Data collection period	Data analysis period
1st round	34	February–March 2023	April–May 2023
2nd round	32	June–July 2023	July–August 2023
3rd round	30	September–November 2023	October–December 2023

Table 2
List of participating experts (in alphabetical order) contributing to all three rounds of research.

#	Name	Job Title	Field of research/application
1	Dr Mona Ashok	Associate Professor of Digital Transformation, Henley Business School, UK	Digital Ethics (ethical considerations in the use of AI and Digital Technology)
2	Dr Dorothea Baur	Consultant and Lecturer, Baur Consulting AG, Switzerland	Business Ethics, Ethics of AI, Environmental Ethics
3	Dr Nicole Duller	Senior Scientist, University of Klagenfurt, Austria	Ethics of Robotics, Space (Robot) Ethics, Mediatised Intimacies and Ethics
4	Prof Michael Fisher	Professor of Computer Science, University of Manchester, UK	Ethics of Robotics, Ethics of AI, Autonomous Systems
5	Dr Sarah Fletcher	Principal Research Fellow in Industrial Robotics Psychology, Cranfield University, UK	Ethics of Robotics, Research Ethics
6	Dr Bianca Himenez	Chief Data Officer, PhD in Ethical Machine Learning, PUR, France	Ethics in ML
7	Dr Willam Knight	Research Ethics and Integrity Officer, De Montfort University, Leicester, UK	Research Ethics, Ethics Compliance, Data Protection, Data Governance, Neuroethics
8	Dr Tobias Kopp	Postdoctoral researcher, Karlsruhe University of Applied Sciences, & Business consultant, esentri AG, Ettlingen, Germany	Ethics of Technology, Ethics of AI, Collaborative Robots, Technology Assessment
9	Dr Karen Lancaster	Academic researcher, University of Nottingham, UK	Ethics of Robotics, Human-Robot Interaction
10	Dr Natalie Leesakul	Assistant Professor in Law and Autonomous Systems, School of Law, University of Nottingham, UK	Ethics in Autonomous Systems
11	Dr Alex Leveringhaus	Lecturer in Political Theory, University of Surrey, UK	Ethics of Automation, especially in relation to weapons development and armed conflict
12	Dr Alireza Mortezapour	Postdoctoral researcher, University of Salerno, Italy	Ethics in Human-Robot Interaction with HF concerns
13	Prof Sven Nyholm	Professor of the Ethics of Artificial Intelligence, LMU Munich, Germany	Ethics of Robotics, Ethics of AI, General Ethical Theory, History of Ethics
14	Dr Fiorella Operto	Robotics Consultant, Scuola di Robotica, Genoa, Italy	Ethics of Robotics
15	Dr Pericle Salvini	Senior Researcher, University of Oxford, UK	Ethics of Robotics
16	Dr Filippo Santoni de Sio	Associate Professor Ethics and Philosophy of Technology, Delft University of Technology, The Netherlands	Ethics of AI, Ethics of Robotics
17	Dr Murray Sinclair	Visiting Scholar, Loughborough University, UK	Ethics in Engineering
18	Prof Neil Sinclair	Professor of Philosophy, University of Nottingham, UK	Ethical Theory, Meta-ethics, Ethics of AI
19	Prof Matthew Studley	Professor of Ethics and Technology, Bristol Robotics Lab, UWE Bristol, UK	Ethics of Technology, with a focus on Robotics
20	Dr Gianmarco Veruggio	Robotics Scientist, Scuola di Robotica, Genoa, Italy	Ethics of Robotics
21	Dr Steven Umbrello	Managing Director, Institute for Ethics and Emerging Technologies, University of Turin, Italy	Ethics of Technology, Ethics of AI, Quantum
22	Dr Anouk Van Maris	Senior Project lead, Zuyderland, The Netherlands	Ethics of AI, Ethics in Social Robotics
23	Dr Sara Wilford	Associate Professor, De Montfort University, Centre for Computing and Social Responsibility, Leicester, UK	Computer Ethics, Privacy, Data Protection
24	Dr Max L. Wilson	Associate Professor, University of Nottingham, UK	Neuroethics
25	Dr Blay Whitby	Technology Ethicist, Imperial College, London, UK	Ethics of Robotics, Ethics of AI, Social Robots, AI in care settings
26	Participant I.D.#01	[Industry]	Ethics of Robotics, Responsible Robotics, Machine Ethics
27	Participant I.D.#02	[Academia]	AI and Robotics in Manufacturing
28	Participant I.D.#03	[Academia/Industry]	Ethics of Technology Regulation
29	Participant I.D.#04	[Academia/Industry]	Ethics of AI
30	Participant I.D.#05	[Academia]	Ethics of Robotics

2.3. Domain experts

We engaged domain experts primarily from academic backgrounds, with a focus on the European region. This choice was intentional, as we sought to investigate specific characteristics of this region, encompassing aspects such as law, policies, economy, manufacturing activity, society, and culture. Critically, we wanted to understand, and collaboratively design, a framework that integrates their insights into the potential challenges associated with the introduction of smart technologies in both life and work contexts, areas in which they are actively engaged in studying, debating, and conceptualising.

Although initially 34 domain experts agreed to participate in our study, over the multiple stages of data collection we experienced a minor attrition rate resulting in the loss of 4 participants. The list of participants who have concluded the three rounds of data collection is provided in Table 2 with an overview of their expertise.[1]

[1] Here, only the names of the experts who concluded all three rounds of the data collection are provided. Notably, at the conclusion of the third-round data collection phase, we requested permission from each participant to disclose their names to signify their contribution to the co-creation process of this research. For those who preferred to remain anonymous, we just refer to the body of expertise they brought in the research.

2.4. Reflexivity statement

In this study, we adopted a multi- and interdisciplinary approach to incorporate and interpret the insights from the participating experts throughout the study's three rounds of data collection and analysis. From one side, the diverse academic and professional background experience of this paper's authors, (from fields such as engineering robotics, systems ergonomics, work and organisational psychology, and law) lay the foundations for shaping the ethical framework in its consolidated outline (as presented in section 4). For instance, the ecosystemic-layer structure—micro, meso, (and macro, although not fully developed here due to reasons of scope)—was suggested by the author with a background in organisational study, to ensure that ethical considerations be woven into the fabric of the organisation at every level. On the other hand, ethical principles requiring specific legal explorations (e.g. 'liability, responsibility and accountability'), were reviewed and integrated by the authors with a background in law.

Additionally, we acknowledge that our ethical framework's results are predominantly influenced by Eurocentric views, as the authors, and the participating experts, are based in European institutions. While this may be seen as a limitation for the ethical framework's replicability outside of the European space, we believe it does not significantly detract from the broader applicability of our results. Critically, the question here encompasses more the experts' epistemic foundation of credentials and socialised specialism, rather than their geographical representativeness. The experts' knowledge is shaped by doctrines and ideologies rooted in Western techno-social structures, particularly those of the US. Although automation-intensive jurisdictions like China are challenging these epistemic sources, they have yet to develop full-fledged alternatives. Therefore, experts in ethics and law from other world regions might not necessarily result in substantially different perspectives at this time.

Finally, methodologically, quality criteria included rigour and transparency in the analytic strategies adopted throughout the Delphi iterative rounds, time for the assessment and review of the collected inputs, and an in-depth scrutiny of the insights from our participating experts. A computer-based qualitative data analysis software (NVivo 12 Pro for Windows, ©Lumivero) was used to support empirical data storage and analysis. Additionally, it served as a social platform for this paper's co-authors to review, add comments, and make annotations during the analytic processes. Critically, in each subsequent questionnaire presenting the analytical outcome of the participants' responses gathered in the previous one (questionnaire 2 and questionnaire 3), we consistently asked our experts whether the suggested results and identified ethical issues adequately represented the potential structure of the ethical framework for H-R collaboration in manufacturing. Their responses helped us be reflexive on our understanding of their perspective, ensuring that our data interpretation accurately captured their inputs.

3. The process of building the ethical framework

3.1. Mapping the problem area

The problem area investigated with our participants in the three-round Delphi study was identified through a literature review on ethics in H-R systems. Literature was searched in the main scientific databases (such as: Ebsco, Science Direct, Scopus, Scholar) using the following key-terms: ethic* AND human-robot AND interaction OR collaboration; ethic* AND human-robot AND manufacturing; ethic* AND industry 5.0 OR industry 4.0. In this review, four overarching ethical concerns pertaining to H-R collaboration in manufacturing—workforce skilling, safety-related concerns, psychological impact on workers, and accountability—are presented. Each theme prompted a series of questions, which served as the foundational pillars to guide the definition of the scenarios presented in section 3.2.1.

In the I.5.0 paradigm, the integration of collaborative robots alongside human workers will form novel mixed teams on the shop floor, altering the organisation of labour and the very fabric of work. As industry and academia are pushing towards the highest form of *collaboration* in H-R interaction, which occurs when the human worker and the robot perform a task together within the same workspace [66], other forms of interaction can happen on the shop floor. These could be: *coexistence*, when the human worker and collaborative robot are in the same environment but generally do not interact with each other; *synchronised*, when the human worker and collaborative robot work in the same workspace, but at different times; and *cooperation*, when the human worker and collaborative robot work in the same workspace at the same time, though each focuses on separate tasks [67,68]. Within the different configurations, it is expected that tasks will be allocated in such a way that robots take on physically demanding tasks, while human workers are assigned higher-value upstream or downstream manual tasks [4,5].

The design of future task allocation in I.5.0 work settings raises significant concerns regarding unemployment and the evolving role of the future industrial workforce. With collaborative robots being allocated tasks that can be injury-prone—thus mitigating physical ergonomic risks that often contribute to work-related musculoskeletal disorder (WMSD) injuries [69]—the human workforce is expected to cover more sophisticated cognitive and supervisory roles [3]. While not all tasks can be automated, there is a growing consensus that some will continue to rely on human involvement [70]. The existing literature is pointing to creative and problem-solving abilities [71,72], although the essential skill requirements for the future workforce are still under investigation [73,74]. Critically, ethical questions arise regarding the expectations imposed on blue-collar workers to acquire new skills, including creativity and problem-solving abilities, to secure their jobs [72]. While there is an expectation for employers to support their workforce in acquiring new skills to meet the emerging H-R demands, at a societal level, the labour market is urged to re-structure its educational and training programs [75]. *Which ethical considerations related to aspects such as the acquisition of new skills specifically pertain to H-R collaboration in manufacturing?*

Another ethical concern pertains to worker safety in H-R collaboration, encompassing both physical and psychological aspects. In the context of ensuring the physical safety of workers engaged in collaborative activities with collaborative robots, scholarly research is actively exploring various avenues to address this pressing issue [76]. ISO standards and regulations [6,7], as well as research into legal and regulatory frameworks [77], are also continuously evolving to enhance the safety requirements for manufacturing robots. Technical standardisation presents a primary force driving safety compliance, not only in automation but also in broader contexts. The standardisation landscape encompasses at least three distinct domains that are critical within the scope of our study. Domestic standards, which are either binding or factually enforced (e.g. China's GB/T 'recommended'' standards [78]); EU-level standards, such as the Harmonised Standards promoted by the European Commission; and international standards, which are negotiated under either the World Trade Organisation framework, or standard-setting agencies such as the ISO or the IEEE Standards Association (IEEE SA), mostly known as "standard developing organisations (SDOs). While the EU has pioneered standardisation initiatives in other areas (e.g. space missions, product safety, or privacy-preserving biometric systems), this is not the case in (smart) collaborative robotics. Therefore, we focus on international standards, such as the IEEE SA and the ISO. Critically, these standards are increasingly gaining regulatory, geopolitical, and geo-economic significance [79,80]. While international standards have shaped automation (including robotics) for over a century, at least in terms of interoperability and standard-essential patents, there are entire regulatory areas today that are covered either temporarily or permanently by technical standards, which are later transposed into legislation, executive and administrative decisions, arbitral awards, and court judgements in the absence of more appropriate traditional legal sources. Collaborative robotic safety is indeed a case in point [77]. This is mostly due to legislators being ill-equipped to navigate the unrelenting pace, uncertainty, and complexity of current technological developments, especially at the intersection between machines and AI [81] and solutions 'by design'. However, lawyers and ethicists are seldom involved in standard-setting exercises. Such legal and regulatory frameworks are being shaped alongside the development of theoretical and methodological human factors (HF) models aimed at creating safer workspaces [10,82]. This includes collision avoidance technologies and real-time H-R detection cameras/sensors to support effective and efficient H-R collaboration and task performance [10]. To facilitate this transition and enhance safety protocols on the factory floor, manufacturing industries adopting H-R collaboration will heavily rely on perception systems like cameras and smart sensors [11,12]. These systems will document the interactions between humans and collaborative robots throughout the production process, with the objective of improving workspace safety by monitoring H-R behaviours. In doing so, they can trigger a safety stop in risky situations (such as, if the collaborative robot approaches too closely to a human operator) [83,84]. With future factories relying on the continuous exchange of data gathered from smart systems, as well as sensors during the H-R dynamics on the shop floor, questions emerge regarding which data are collected, processed, and shared, and for what purposes [85]. Secondly, concerns arise regarding the extent to which these data can provide accurate information about human, robot, H-R interaction and performance. *Which ethical considerations related to aspects such as ensuring safety in the workplace (with safety-related monitoring) specifically pertain to H-R collaboration in manufacturing?*

While studies concerning physical safety are proliferating in the literature, there remains a dearth of research exploring the psychological impact that H-R dynamics may have on workers. The attention heightened when the Amazon case was published [86], revealing the extent to which surveillance practices and H-R performance demands can significantly affect worker wellbeing, leading to elevated levels of stress and anxiety. In line with these concerns, Fletcher and Webb [40] have also raised the issue of psychological harm that H-R collaboration may pose to workers. Research is required to explore the role of emotions and embodied experiences associated with the subjective perception of H-R collaboration [72]. *Which ethical considerations related to aspects such as psychological hazards in the workplace specifically pertain to H-R collaboration in manufacturing?*

Finally, the increasing integration of robots into manufacturing settings prompts inquiries into responsibility and liability, urging ethical and legal considerations to address the challenging interaction and future collaboration between human and robots on the shop floor. Occupational health and safety laws are also facing the challenge of defining the appropriate application of their norms to robots with varying degrees of autonomy, across numerous jurisdictions [77] This gives rise to questions regarding the allocation of liability for damages caused by malfunctioning robots [56,77]. *Which ethical considerations related to aspects such as responsibility and liability specifically pertain to H-R collaboration in manufacturing?*

3.2. Questionnaire 1: Uncovering (overlooked) ethical themes

3.2.1. Questionnaire purpose

The primary objective of the first questionnaire was to elicit ethical issues in H-R collaboration within the manufacturing sector that may have been overlooked or remained unexplored. Building on the outcome of the literature review presented in the previous section, we employed a prospective scenario-based approach [87–89], leveraging participants' expertise to reveal novel insights. Prospective scenarios help experts envision and explore possible future outcomes and situations [90], providing a framework to anticipate how ethical challenges in H-R collaboration may develop over time. The selected prospected scenarios included the following topics.

- **Prospective Scenario 1 - Workforce skilling.** This scenario presented potential challenges surrounding the training and retraining of human workers. It explored the ethical dimensions of skill development and adaptation within the dynamic landscape of H-R collaboration.
- **Prospective Scenario 2 - H-R interaction monitoring beyond safety purposes.** This scenario outlined a prospective shop floor environment in which perception systems, such as cameras and intelligent sensors, are extensively integrated to ensure a safe and reliable H-R collaboration setting. This integration facilitates the collection and processing of data concerning H-R performance, generating insights that can enhance production planning and operations.
- **Prospective Scenario 3 - Psychological safety in H-R collaboration.** This scenario suggested the challenge of future dynamic task allocation in H-R collaboration, where traditional allocation methods may not efficiently adapt to emergent changes, leading to inflexible collaboration that could impact worker mental and psychological wellbeing.
- **Prospective Scenario 4 - Responsibility/Accountability.** This scenario highlighted potential concerns about who should be held accountable for incidents occurring during H-R collaborations on the shop floor.

3.2.2. Data collection and analysis

The data collection occurred from February to March 2023. All scenarios posed the same question to the experts: "*What ethical themes and/or considerations does this scenario pose? Please, elaborate your responses, and the reason why you have suggested them.*" The experts had the freedom and flexibility to comment on one or all scenarios, without any word count limitations, and they could choose their preferred response style.

In the analytical phase, we considered the responses collected for each questionnaire as a single unit of analysis. The empirical data was interpreted through the Reflexive Thematic Analysis method, as outlined by Braun and Clarke [63,91,92], to inductively derive and construct candidate ethical themes relevant for the H-R collaboration future configuration. The long list of codes identified (approximately 500 codes)—was then transferred into a concept map in NVivo, and clustered based on their shared meanings. The clustered codes were further reviewed, going back to each coded text to check whether the codes were correctly grouped according to their similarity with other codes. This involved assessing whether the forming themes were able to share 'a story' about their specific meaning/value [63]. This process terminated when the generated themes conveyed multifaceted, conceptually rich meanings. Finally, the consolidated themes (with initial names) were organised into a provisional structure as showed in Fig. 1, in which each theme included 'sub-themes' (i.e. the theme's *attributes* with specifications of the core idea conveyed by the related theme). The analysis of the data coded within each theme's specific sub-theme/attribute supported the definition of operational statements, contributing to the definition of initial ethical considerations for H-R collaboration.

3.2.3. Strategies for the design of questionnaire 2

Table 3 presents the outcome of the analysis of the first questionnaire, and the resulting information that guided the design of the

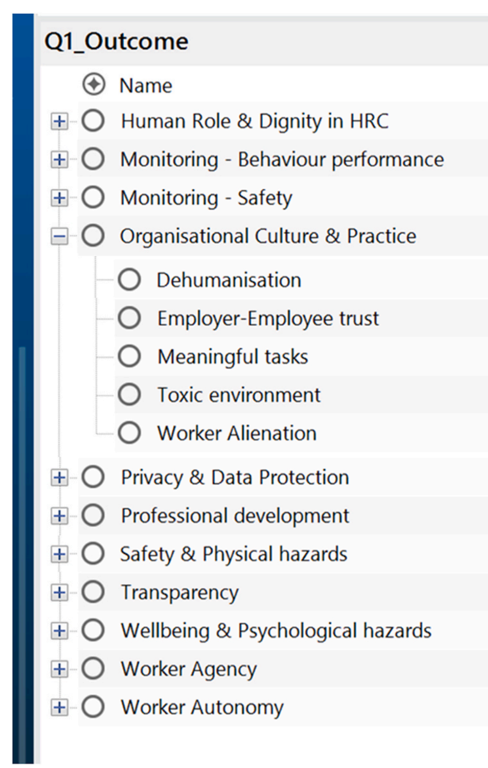

Fig. 1. Questionnaire 1's outcome, i.e. a first draft of the ethical framework.

subsequent questionnaire. Specifically, for each selected high-level theme, we provided a list of operational statements each specifying an attribute/sub-theme of the related theme.

3.3. Questionnaire 2: defining and refining the proposed ethical themes

3.3.1. Questionnaire purpose

With this questionnaire, our aim was to achieve a deeper understanding of the developed ethical themes, by exploring their associated attributes. To achieve this, the participating experts were asked to review each operational statement based on two key criteria. The first criterion referred to *completeness*, i.e. the accuracy of the description provided by each statement for the associated high-level ethical theme; the second criterion regarded *relevance*, i.e. the extent to which each statement was judged appropriate, through a 4-point Likert scale ranging from *'Irrelevant'* to *'Essential'*.

3.3.2. Data collection and analysis

The data collection spanned from June to July 2023. Table 4 below presents an excerpt from questionnaire 2, detailing how expert responses were collected for each proposed operational statement within the overarching high-level ethical themes.

In the analytical process, three sequential strategies were employed. Initially, colour-coding was used to guide the assessment of '*Relevance*'. The principles rated above 3.50 points were coded in green, suggesting a need for refinement but approaching essential status. Conversely, those below 3 were coded in red, indicating either insignificance or a requirement for more substantial adjustments. Following this, the Qualitative Content Analysis (QCA) method [93] was employed to thoroughly highlight the qualitative and quantitative aspects of proposed changes, encompassing word choices, syntax, and general comments. Finally, attention was turned to each expert's perspective to determine their level of supporting or expertise regarding the comment provided for each principle.

3.3.3. Strategies for the design of questionnaire 3

In addition to the analysis of each operational statement, a series of recommendations was considered for the design of the subsequent questionnaire 3. First, certain operational statements were found to be redundant or overlapping, highlighting the need for further refinement. Other comments cautioned against a simplistic list of ethical themes, suggesting instead the recognition of the complex and dynamic influences occurring at various levels within and outside an organisation. Therefore, three operational levels were introduced through which the revised themes were reorganised (Fig. 2).

- micro-level, i.e. the direct interaction between human workers and collaborative robots on the shop floor),
- meso-level, i.e. the organisational governance to encompass processes, procedures, and practices for organising the H-R labour safely and effectively),
- macro-level, for societal governance.

Table 4
An example of how we structured the second questionnaire.

Provisional high-level theme:	Organisational Culture and H-R practices
Statement(s): [example of 1 of 6]	To maximise the benefits of human-robot collaboration, the employer-employee relationship should be built upon a foundation of trust. *Irrelevant – Moderately important – Very important – Essential* *Is this statement accurate and complete?* *Yes* *Other* ____________
Are there missing considerations within this ethical theme?	

Furthermore, based on specific suggestions, statements were reformulated to better characterise the specific H-R collaboration configuration, as this participant sharply commented:

> **The high-level ethical themes formulated above represent many, if not most or perhaps almost all, of the most ethical themes related to the ethics of human-robot interaction in the workplace. One thing one could reflect on is whether employers have a duty to try to not only avoid harming, degrading etc employees, but whether they also have a duty to try to provide them with opportunities for meaningful work. However, many of the themes covered above (exercising skills, having autonomy, etc) correspond to the components of meaningful work. What perhaps was missing were two things: employees should have opportunities to excel in their work and to get recognition for any excellence they show and, second, the workplace should be designed and organised in a way that fosters good working relations and the values of collegiality among the people in the workplace.** *[Prof Sven Nyholm, Professor of the Ethics of Artificial Intelligence, LMU Munich, Germany]*

3.4. Questionnaire 3: revising the ethical themes within a novel ethical framework structure

3.4.1. Questionnaire purpose

In the third and final questionnaire, the experts were tasked with evaluating all themes and associated statements in the proposed novel structure. This was done in terms of quality by (1) selecting the thr ee most important statements from the list within each ethical theme, and (2) providing the reasons for their choices, along with any additional comments they may have. Table 5 below exemplified what the experts were prompted with—at the micro-level, the theme of '*human autonomy*' and the list of proposed revised statements, with the questionnaire questions.

3.4.2. Data collection and analysis

The data collection for this third and last questionnaire spanned from September to November 2023. For the analytical activity, a *preference orientation* was adopted, with participants choosing the statements they

Table 3
Analysis of first-level themes, and associated second-level themes with derived operational statements.

High-level theme	Associated sub-theme	Operational statement
Organisational culture and H-R practices	Dehumanisation	Employers should be mindful of the risks of work dehumanisation, which can lead to psychological hazards for workers, if they are solely driven by the pressure to increase productivity and reduce costs.
	Employer-Employee trust	To maximise the benefits of human-robot collaboration, the employer-employee relationship should be built upon a foundation of trust.
	Etc.	…

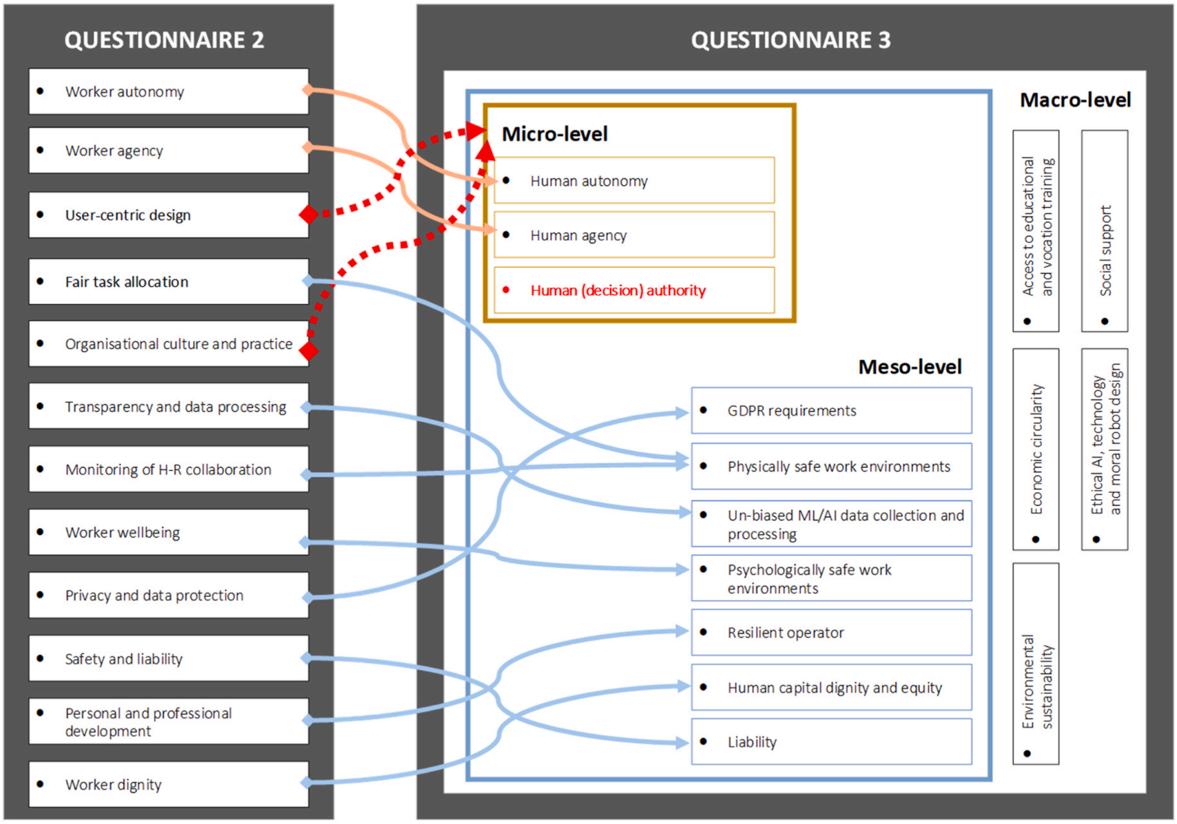

Fig. 2. Rationale behind the design of the next questionnaire.

Table 5
An excerpt of questionnaire 3.

Operational level	Theme	Statements to select
Micro	*Respect for human worker autonomy* in H-R collaboration* (*Human autonomy is defined as an individual's effective capacity for self-governance, i.e. the freedom to make choices and take actions guided by own beliefs, values, motivations, and reasons.)	☐ Human workers should have the capacity to exercise a level of autonomy in the performance of collaborative tasks with collaborative robots – these could encompass either physical or cognitive functions. ☐ Human workers should have the capacity to exercise a level of autonomy (e.g. freedom) to choose their own method of working collaboratively with collaborative robots. ☐ Collaborative tasks allocated to H-R agents should allow a level of autonomy for human workers for potential adaptation to the unpredictability of manufacturing process-related factors, such as temporal task demand variations, or for a change in the state of the operator (e.g. fatigue). ☐ Human worker autonomy should not be threatened by any monitoring of collaborative tasks performed with collaborative robots. ☐ Human worker autonomy should be sustained to flexibly allocate tasks between H-R configurations
	Please, provide your comments here:	
Does the micro-level encompass the relevant ethical themes? Is there anything missing, or that you would like to add/change?		

considered important, and providing rationale for their choices. The qualitative comments served to provide further insight into the theme, and the ethical considerations the theme would entail. Specifically, the statements that received an agreement score above 80 %[2] were selected, and the expert comments were used to refine them. The comments associated with those statements that were not selected by our experts helped further clarify the overall theme.

[2] The cut-off at 80 % was a discretionary decision based on the rationale of reliability scores used in qualitative data analysis, where an agreement greater than 75 % indicates a very good level of agreement.

3.4.3. *Strategies for the consolidation of the ethical framework*

This questionnaire's outcome includes the experts' final comments regarding the proposed structure, and the content/revision of the framework's content (i.e. themes and associated statements)

Overall, the interconnected structure of the ethical framework into three operational levels was positively received by the participating experts. They agreed that organisations are required to establish ethical governance (at the meso-level) to encompass processes, procedures, and practices for organising the H-R labour safely and effectively, by also supporting the development of skills in their human resources. At the micro-level, however, the experts considered the focus to be on the H-R dynamics triggered throughout the collaboration between one or more human workers and one or more collaborative robot systems during the manufacturing process on the shop floor. The organisation's micro and meso-levels are interdependent with each other and their external environments. The responses collected at the macro level were analysed, but further refinement was deemed necessary beyond the scope of this Delphi study.

Finally, in response to feedback suggesting that the 'themes' category should denote the principles informing and guiding ethical values, we have renamed the initial term 'ethical theme' to ***ethical principle***. Conversely, the 'statements' are now defined as ***ethical design goals***, which express specific objectives or targets aimed at achieving desired outcomes in ethical design considerations.

4. Results and discussion: the consolidated ethical framework for H-R collaboration in manufacturing

The consolidated ethical framework is presented in Fig. 3. The 4.1 and 4.2 sections delineate the micro and meso-levels separately. Within each micro and meso-level, the consolidated ethical principles alongside the selected associated ethical design goals are first presented (Tables 6 and 7), and further commented on and discussed in the subsequent subsections.

4.1. H-R collaboration on the shop floor (micro-level)

Within this level, the underlying key principles revolve around respect for human worker *autonomy*, *authority*, and *agency* within H-R collaboration (Table 6). These principles are related yet distinct concepts, as is highlighted within the following subsections. Core topics that will be discussed in this section in relation to the three principles include 'margins of freedom and autonomy in task execution,' 'meaningful work and task significance,' 'distribution of power', and 'trust'.

4.1.1. Human autonomy in H-R collaboration

The concept of *'autonomy'* can be defined as a person's effective capacity for self-governance, i.e. the freedom to make choices and act guided by own beliefs, values, motivations, and reasons [94]. At a theoretical level, the concept of autonomy has been subject to various conceptualisations, with the Self-Determination Theory being particularly noteworthy [95]. This theory defines autonomy as a sense of willingness and volition in taking actions aligned with one's goals and values [96]. Importantly, research has demonstrated that individuals who possess the ability to act autonomously tend to exhibit increased engagement, productivity, and enhanced psychological wellbeing [96]. Within work contexts, autonomy can be broadly referred to as the capability of a system, whether it is a human or non-human agent, to make independent decisions and accomplish objectives without relying on any external assistance or support [97]. Legally, too, human autonomy is defined as the ability to make informed decisions, provide consent, and preside over one's private sphere of action, thought, and decision-making; in the context of automation, control over one's data and mental integrity, alongside self-determination and agency, are indicative of human autonomy in dealing with collaborative robotic systems, and should be preserved [98,99]. To do so, the most fitting balance should be sought with *machine* autonomy—i.e. the ability of the collaborative robot to take decisions influencing the environment (that is, the overall outcome of the collaborative pursuit), independently from human influence or oversight [77,100–102]. This balancing act has long been referred to as the "autonomy vs safety" regulatory paradox [103].

AUT-1: *Collaborative tasks allocated to H-R agents should allow (a degree of) autonomy for human workers for potential adaptation to the unpredictability of manufacturing process-related factors, such as temporal task demand variations, or for a change in the state of the operator (e.g. fatigue).*

Within this ethical and legal principle of autonomy, the ethicists prioritised design goals that facilitate human autonomy, allowing workers to maintain control over their tasks and decisions while leveraging the capabilities of robots to enhance collaboration. Although in contemporary industrial processes the notion of allowing margins of freedom and autonomy in task execution may seem challenging to envision, the acknowledgement of the potential constraints within organisational settings has been suggested, especially when

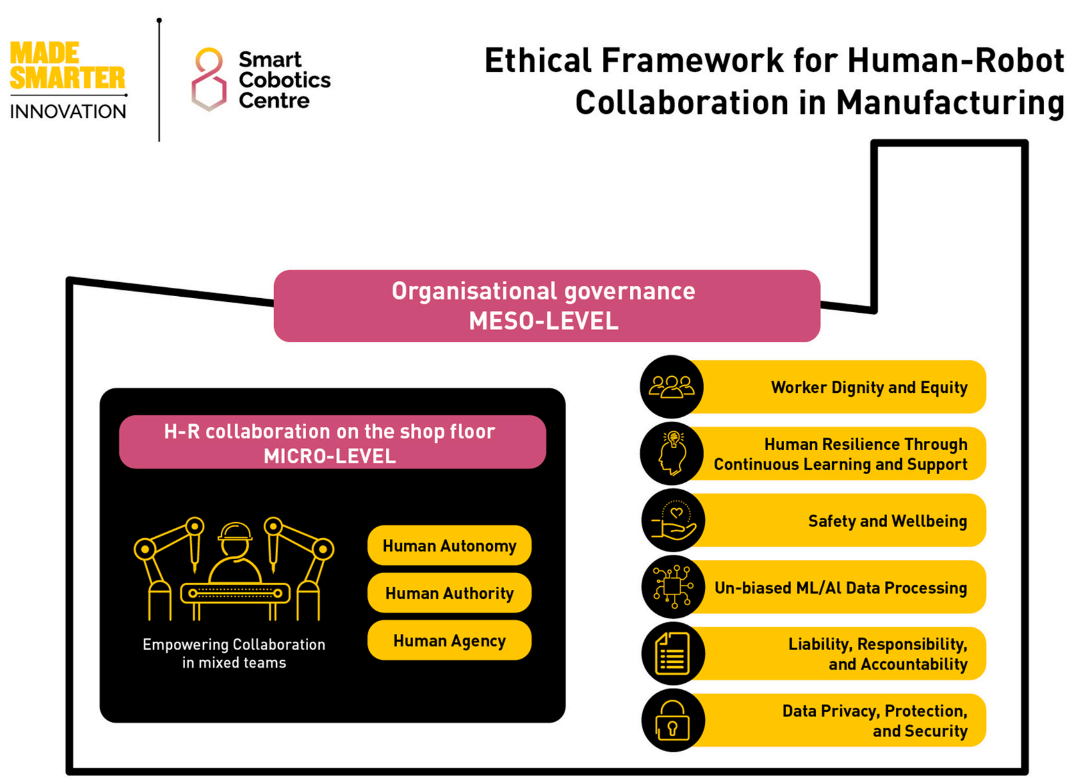

Fig. 3. Infographic of consolidated ethical framework in H-R collaboration in manufacturing.

Table 6
The micro-level, i.e. ethical principles for H-R collaboration on the shop floor.

#Ethical principle	#Associated ethical design goal(s)	% of convergence
Human autonomy	**AUT-1**: Collaborative tasks allocated to H-R agents should allow (a degree of) autonomy for human workers for potential adaptation to the unpredictability of manufacturing process-related factors, such as temporal task demand variations, or for a change in the state of the operator (e.g. fatigue).	93.8 %
	AUT-2: Human worker autonomy should not be threatened by any monitoring of collaborative tasks performed with collaborative robots.	87.6 %
Human (decision) authority	**ATH-1**: Human workers should be empowered to influence changes and improvements in the work environment, and H-R collaboration, fostering meaningful work.	88.1 %
	ATH-2: Human workers should be given a degree of control over their work processes, allowing them to make decisions, fostering a sense of ownership.	87.7 %
Human agency	**AGN-1**: Human workers should retain the ability to exert influence over the tasks allocated to collaborative robots, fostering a sense of agency (and trust) in their interactions with collaborative robots.	94.2 %

collaborative robots may lack the flexibility to fully accommodate human autonomy. As pointed out by the following ethicist:

> **Human autonomy is limited in the context of a workplace. Employers are usually permitted to decide some parameters in which in their workers can act. The real question is what degree of freedom workers have within those parameters, especially the freedom to organise their workload or the performance of their tasks.** *[Dr Alex Leveringhaus, Lecturer in Political Theory, University of Surrey, UK]*

Central to this is the notion of adaptability and flexibility. Critically, the design of work organisation should enable the enhancement of human capabilities, without infringing upon individual autonomy. In this regard, collaborative robots should be conceptualised as an 'assistant/aid' to the human worker. Notably, perceived loss of autonomy represents a significant challenge within H-R collaboration, as it can directly impact the level of engagement and participation of human workers in the collaborative process.

AUT-2*: Human worker autonomy should not be threatened by any monitoring of collaborative tasks performed with collaborative robots.*

Addressing factors that impact workers' autonomy, including monitoring and automation bias, is crucial for fostering a positive and empowering work environment. Monitoring systems, while valuable for tracking H-R interaction for safety purposes, can infringe upon workers' autonomy if implemented without proper consideration for privacy and individual autonomy. Excessive monitoring can influence how human workers perceive their surroundings, potentially leading to feelings of surveillance, oppression and alienation when collaborating with collaborative robots. These concepts will be further developed and addressed within the ethical framework's meso-level (section 4.2).

> **Workers' autonomy should be sustained as much as possible, depending on the type of task and in accordance with safety**

Table 7
The meso-level, i.e. ethical principles of organisational governance for H-R collaboration.

#Ethical principle	#Associated ethical design goal(s)	% of convergence
Liability, responsibility, and accountability	**LRA-1**: Adequate responsibilities, roles, and procedures should be in place for addressing any potential liability, responsibility, and accountability involved in H-R interaction and collaboration.	82.4 %
	LRA-2: Guidelines and mechanisms should be established to ensure accountability in cases where H-R interaction and collaboration lead to errors, damages, or accidents.	81.9 %
Data privacy, protection, and security requirements in H-R collaboration	**PPS-1**: Collect and retain only the necessary data for the intended purpose of improving safety and productivity, minimising the risk of privacy infringement.	95.2 %
	PPS-2: Informed and voluntary consent must be obtained for data collection and processing related to work activities in H-R interaction and collaboration.	91.4 %
Un-biased ML/AI data processing	**AI-1:** Ensure transparency in data collection and processing methods of ML/AI data, and continuously monitor and audit ML/AI systems for algorithm biases, making adjustments as necessary to ensure fairness.	98.9 %
	AI-2: Limit ML/AI data processing to the specific purposes disclosed to workers, ensuring that the collection and processing of ML/AI data align with relevant privacy laws and ethical standards.	80.4 %
Safety and wellbeing	**SAW-1**: Provide safety training to human workers collaborating with collaborative robots, focusing on understanding, and mitigating the risks associated with H-R interaction and collaboration.	96.9 %
	SAW-2: Incorporate safety measures into the design and development of collaborative robots to ensure the physical safety and wellbeing of human workers, thereby minimising the risk of accidents on the shop floor.	92.5 %
	SAW-3: Avoid intrusive or constant surveillance practices, as they can contribute to developing a surveillance culture and have adverse effects on the psychological wellbeing of human workers.	88.2 %
	SAW-4: Encourage a (just) culture of reporting safety incidents, even if they do not result in harm, to improve safety practices continuously.	84.7 %
	SAW-5: Establish a confidential feedback system that allows human workers to report psychological concerns, ensuring anonymity and protection against retaliation.	82.9 %
	SAW-6: Ensure that all safety practices and equipment meet or exceed relevant regulatory requirements and industry standards.	82.9 %
Human resilience through continuous learning and support	**RES-1**: Support worker personal development to empower them to navigate their roles safely and effectively, fostering a culture of continuous learning and skill enhancement.	90.4 %
	RES-2: Support leadership practices that prioritise the wellbeing and psychological safety in H-R interaction and collaboration.	83.0 %
Human dignity and equity	**HDE-1**: Implement policies and procedures to respect, retain and enhance human worker expertise and value.	96.7 %
	HDE-2: Ensure that human workers receive fair and competitive compensation for their roles in H-R interaction and collaboration on the shop floor, taking into account their skills, contributions, and the changing nature of work.	91.8 %

procedures. I wouldn't be too specific on the purposes (i.e. the possibility to choose the method of working collaboratively with robots, to flexibly allocate tasks, to adapt to manufacturing process, etc.), because there might be many, based on the type of H-R collaboration. Attention should be paid to all those factors impacting on workers' autonomy, such as monitoring, automation bias, etc. Establishing an open communication channel between workers and managers will ensure identification and mitigation of factors negatively affecting workers' autonomy. *[Dr Filippo Santoni de Sio, Associate Professor Ethics and Philosophy of Technology, Delft University of Technology, The Netherlands]*

When workers perceive that their performance is continuously assessed against the standards set for robotic systems, this can evoke feelings of job insecurity or stress. This is especially true when the data on the performance of the specific H-R collaboration suggests that they might be performing less effectively than their peers in comparable H-R collaborations. As one ethicist argued:

Workers may feel that their autonomy is threatened if they are constantly monitored and analysed. It is essential to respect workers' autonomy and ensure that they have a say in how their data is collected and used. *[Prof Matthew Studley, Professor of Ethics and Technology, Bristol Robotics Lab, UWE Bristol, UK]*

4.1.2. Human (decision) authority

Human '*authority*' refers to the amount of decisions an agent is allowed (and required) to make by itself [104]. This encompasses the ability of humans to guide, supervise, and make critical choices within the collaborative context, ensuring that human values, intentions, and behaviours are key during the interaction and collaboration with collaborative robots [105]. Within H-R collaboration, human authority can refer to the extent of control, decision-making, and influence that human worker wields over robotic systems or collaborative tasks in terms of determining how tasks are executed [104,106]. These concepts are predicated on the belief that human workers possess competence, sensibility, and the capacity for sound judgment. That includes instances of how authority in work situations is identified, allocated and exercised throughout the manufacturing processes.

As suggested by one ethicist, while it is generally advisable for humans to retain authority over robots, the specific distribution of authority may vary depending on the nature of the tasks, such as those involving complex decision-making or safety-critical operations. However, overly restrictive control measures may hinder the effectiveness of robotic systems, leading to inefficiencies or missed opportunities for optimisation.

In general, humans should be given authority over the robot. However, authority may depend on the type of task and should be designed in accordance with safety procedures. The possibility to take back control or override the robot decision can be important in specific circumstances but can become a danger in other. However, authority should go hand in hand with understanding the robot's working/decision/plan (transparency, explainability, and understandability). There should always be the possibility to ask a robot "why did you do that?" and "what-if questions". Only coupled with this kind of knowledge human authority can make sense. *[Dr Pericle Salvini, Senior Researcher, University of Oxford, UK]*

ATH-1*: Human workers should be empowered to influence changes and improvements in the work environment, and H-R collaboration, fostering meaningful work.*

Within the manufacturing context, the work design theory [107,108] and related job characteristic model have studied the cognitive and physical tasks in a job [109,110]. Task significance (through the lenses of individuals' opportunities for meaningful work) plays a critical role in determining the influence of human authority, as it pertains to the perception of how employees' work may affect their own work and the work of others [111]. This can influence worker cognitive engagement, the subjective preferences of the tasks involved, and a sense of purpose and satisfaction [112].

[…] human workers should have autonomy over their work condition; therefore, they should be given the authority and opportunity to contribute towards dictating H-R collaborative tasks to improve the process as well as their work environment. Also, supervision should remain as a human worker's responsibility as this also supports the sense the ownership over H-R collaborative tasks, though this may increase the mental workload on human workers so it will need to be handled appropriately. *[Dr Natalie Leesakul, Assistant Professor in Law and Autonomous Systems, University of Nottingham, UK]*

These factors also raise questions about how the distribution of power affects general working conditions in Industry 5.0, including the design and redesign of work that involves H-R collaboration. Notably, the concept of '*power imbalance*' in H-R collaboration has emerged as a critical ethical issue. This issue highlights the unequal distribution of power and authority/agency between human workers and robots within collaborative settings [23,113,114]. Since human workers are more adaptable than robots in terms of modifying their behaviours, there is a risk that human workers may conform their behaviour to that of robots'. As a result, there is a risk of creating an unequal distribution of power that gives the robot an advantage over humans. While the initial vision may have been for robots to integrate into existing workflows to assist and augment human capabilities, there exists a risk of inadvertently tipping the balance of power in favour of the technology, potentially marginalising human workers. This shift in focus, where humans are required to adjust their behaviours and workflows to align with the constraints of robotic systems, can present challenges and frustrations for workers accustomed to more flexible and adaptable human collaborators.

ATH-2*: Human workers should be given a degree of control over their work processes, allowing them to make decisions, fostering a sense of ownership.*

The emergence of this ethical dilemma underscores the importance of designing collaborative systems with careful consideration for the dynamics of power and human authority. It prompts a critical examination of how decisions can be exercised, tasks allocated, and responsibilities distributed within H-R collaborative environments [115]. Failure to address these power differentials can lead to various ethical and psychological implications. This includes feelings of diminished sense of ownership in the tasks at hand and disempowerment among human workers, reduced sense of belonging and satisfaction, as well as and erosion of trust in the organisation.

The idea itself that workers should adapt to the new technology and work configurations and not the other way around is problematic. *[Dr Filippo Santoni de Sio, Associate Professor Ethics and Philosophy of Technology, Delft University of Technology, The Netherlands]*

The human might become the extension/a tool of the robot. It has happened with warehouse robots supposed to accompany and relieve warehouse workers that were however running too fast, causing the humans to have to run behind them all the time to catch up, eventually getting very exhausted. [As I have noted above], we have higher expectations towards the flexibility/adaptability of humans in a situation than that of robots. Therefore, the burden is on humans. Reprogramming a robot, adjusting its settings etc, takes time. Telling a human to work faster or slower, takes one command. It's a very uneven distribution of power and responsibility: the robot has power, the

human is responsible. *[Dr Dorothea Baur, Consultant and Lecturer, Baur Consulting AG, Switzerland]*

Moreover, the problem lies in those workers who are unable to adapt, as well as the potential loss of knowledge and cooperative behaviour that may occur if employees are laid off. A lack of active involvement in the collaborative process can hinder innovation and problem-solving abilities, limiting the potential benefits that H-R collaboration can offer. Conversely, human workers who feel empowered by their organisation are more likely to be actively engaged in the collaborative process, thereby promoting innovation, and problem-solving [114].

Typically the employer/employee relationship must be built upon a certain amount of trust, a worker trusts that their employer will value their time and recompense them appropriately (or else they are forced into work, but that is a very different situation), and an employer trusts their worker will conduct the task at hand, but trust is tricky when implemented in a system where one side is not capable of trusting the other. *[Dr William Knight, Research Ethics and Integrity Officer, De Montfort University, Leicester, UK]*

One aspect of allocating human authority within tasks performed by H-R teams regards the impacts on the human worker's perceived workload, and therefore the overall H-R team performance. Indeed, in H-R collaboration tasks, human workers may encounter challenges such as cognitive or informational overload. This aspect is further discussed within the 'Safety and Wellbeing' principle, in the 4.2.4 section.

4.1.3. Human agency

Human '*agency*' refers to people's capacity to act intentionally, purposefully, and meaningfully, leading to both intended and unintended consequences of their actions [116]. Theoretically, Bandura's Social Cognitive Theory [117–119] posits that humans possess the capability to self-organise, regulate, and reflect upon their life circumstances. Through these processes, individuals can exert influence over the trajectory of their actions, rather than being solely passive products of their circumstances.

AGN-1*: Human workers should retain the ability to exert influence over the tasks allocated to collaborative robots, fostering a sense of agency (and trust) in their interactions with collaborative robots.*

The impact of attributing agency to agents includes concepts such as trust. Trust enhances and fosters user acceptance in H-R interactions [120,121]. Additionally, the level of trust in H-R interaction can further be influenced by the transparency and usability of the system [122]. Additionally, trust can significantly impact the acceptance and, consequently, the willingness of human operators to depend on the input/output provided by automated systems, particularly in hazardous and uncertain work environments [123,124]. Failures in the pursuit of exerting control in the external world can erode trust in a robot [125]. Negative emotions may be triggered when human workers do not have sufficient trust in the capabilities and reliability of the robots they are working with. These emotions can stem from concerns about the robot making errors, misinterpreting commands, or not being able to adapt to unexpected situations. It can hinder effective collaboration and performance, and therefore negatively influence productivity and task efficiency.

One aspect of human-robot collaboration is human-robot interaction. Even though there are less ethical concerns in industrial settings compared to for example the healthcare setting, it should still be ensured that the way the robot interacts with the human reflects its actual abilities. Workers will rely on these robots and trust them to work in a certain way. However, if the interaction style of the robot does not match its abilities, then this may lead to dangerous situations. *[Dr Anouk Van Maris, Senior Projectlead, Zuyderland, The Netherlands]*

However, trust can be re-established if the robot, despite initial failure, demonstrates a capacity for learning and improvement. Insufficient trust in robotic systems may prompt the operator to intervene and take control (Charalambous & Fletcher, 2019). OYet, feelings of overtrust can similarly be problematic. In performing H-R collaborative tasks, human workers may over-rely on the collaborative robots' tasks and actions, which can prevent them to act in error-driven situations.

4.2. Organisational governance (meso-level)

At this level, ethical principles encompass the critical role played by organisations in defining and implementing mechanisms to ensure the responsible and effective labour design and organisation of H-R collaboration. This involves the development and implementation of ethical principles, including with legally binding effect (as law is a form of applied ethics), providing guidelines, processes, procedures, and practices to ensure safety and wellbeing of workers, the respect of human dignity and equity, the respect of data privacy, protection and security, the prioritisation of un-biased data collected from the manufacturing floor, and considerations of liability, responsibility, and accountability in H-R collaboration (Table 7). Core topics that will be discussed in this section include: a clear distinction between the concepts of 'liability, responsibility, and accountability', the issue of the 'illusion of informed consent', the role of 'transparency' in ML/AI data processing, workspace design and human resource management considerations underlying the 'safety and wellbeing' principle in which, for example, considerations of monitoring practices may lead to both detrimental and unbeneficial psychological effects, the meaning of 'resilient operator', and knowledge management in H-R collaboration.

4.2.1. Liability, responsibility, and accountability

Within this domain, the idea is that only agents, who possess the ability to comprehend, predict, and influence work processes, can be held accountable for harms [39].

LRA-1*: Adequate responsibilities, roles, and procedures should be in place for addressing any potential liability, responsibility, and accountability involved in H-R interaction and collaboration.*

While attributing responsibility is a social function of being in control, i.e. a sense of agency [126], questions of '*responsibility gaps*' are raised in the context of autonomous robots [112,127]. Here, the term has been explored to address the issue of who bears moral or legal responsibility (i.e. accountability) in situations where these systems cause harm or, in more severe cases, result in the injury or death of human beings [110,128]. This has been further explained by one ethicist:

There are a number of issues here. There are, for instance, questions about the relationship between accountability and responsibility. There is a further question as to whether either term is understood in ethical or legal terms. Either way, some have argued that the prospect of unanticipated robot behaviour gives rise to so-called responsibility gaps, that is, situations in which no one can be held responsible for an outcome produced by a robot. My view is that such gaps do not exist. The point is that whoever deployed the robot must have known that the robot could engage in unforeseen behaviour due to the use of machine learning and AI. In that sense, the party that deploys a robot incurs certain risks for itself and imposes certain risks on third parties. Surely, risks need to be balanced against potential benefits. The fact that the use of a robot is risky does not automatically count against its deployment. However, from the perspective of responsibility (and accountability), I think the party that deployed the robot is responsible for things the robot does, even if, at the time of deploying the robot, it could not reasonably be expected to entirely predict all of the robots' potential behaviours. *[Dr Alex Leveringhaus, Lecturer in Political Theory, University of Surrey, UK]*

Within the context of H-R collaboration, the participating experts raised dual ethical considerations. First, in situations affecting the physical and psychological safety of workers. This can occur when robot inefficiencies, errors, or incidents may create a sense of insecurity and fear among workers, who may worry about their job safety, or the potential consequences of mistakes made by the robotic system. Secondly, there may be issues of blame when working with robots, especially if the system fails. It is unclear who would be held responsible, and workers may be blamed even if they are not at fault.

> **My personal view is that accidents and incidents do not have single causes - they are the result of a complex chain of events. As H-R collaboration becomes more common and more complex, we may want to attribute *shared* responsibilities for damaging a tool, for example. The aviation industry has a long tradition of thorough investigation on a no-blame basis and, while this process may be considered too expensive for everyday delays in production, it still provides a template for how to handle accountability. This problem may be an area of debate for decades to come.** *[Dr Blay Whitby, Technology Ethicist, Imperial College, London, UK]*

***LRA-2**: Guidelines and mechanisms should be established to ensure accountability in cases where H-R interaction and collaboration lead to errors, damages, or accidents.*

These risks make it imperative to design regulatory frameworks where (broader) ethical concepts are not detached from their (narrower) legal counterparts. When it comes to risk regulation, discussions revolve around distinguishing liability as a legal device from the somewhat broader appeals to responsibility and accountability. However, marrying legal and ethical concepts, or even defining and applying them, in this field is not straightforward. The application of legal-ethical notions such as *accountability*, *responsibility*, and *liability* to autonomous and semiautonomous systems, including collaborative robots powered by AI, has long been a terrain for contestation [129]. When algorithmic and human forms of intelligence cooperate, one fundamental challenge is that of foreseeing and addressing the so-called ‘emergent properties’ that might cause emergent behaviours – those attitudes and actions that are neither encoded as such, nor ordinarily humans [130]. Of course, there are also several intensities and combinations of human-machine interaction [131]– with liability, responsibility, and accountability adapting accordingly.

On liability, generally, an agent is liable insofar as the law provides for it to be so. An agent can be held liable for performing an action or for failing at performing it, which presupposes the ability to *refrain from performing* or to *perform* mentioned action. Each jurisdiction has its own liability regime, which are grounded in roughly similar principles and assumptions about, for instance, intelligence, causality, autonomy, free will, and agency. What makes collaborative systems so difficult to regulate is the impossibility of automatically applying any known category which was crafted for human or artificial forms of intelligence alone. The liability conundrums of concern for, say, an autonomous vehicle relate to the ultimate liability of a human back in the chain of action: given that the law finds it irrelevant, or unserviceable, to make an algorithm itself ultimately liable, then it tends to assign liability to a human that coded that algorithm, audited (or label-certified) it, embedded it into a robotic system, marketed it, deployed it, or failed to switch it off. The law generally refrains from identifying the algorithm with the ultimately liable agent because algorithms cannot explain their actions, nor can they pay damages for errors: as Richards and Smart [132] emphasised, the agency of algorithmically powered robotic systems is merely *apparent*. Various degrees (strict, joint, vicarious, …) of liability for damage, or criminal liability, liability for negligence, product liability, and numerous other legal terms of art, will thus apply to a human for the ‘decision’ of that algorithm. What makes collaborative settings so controversial liability-wise is the potential overlap between the liability of the co-worker as such, and the liability of the co-worker for the ‘share of decision’ (or of outcome?) which in fact pertains to the collaborative robot. Many more ‘geometries’ can be expected (liability of groups of co-workers; liability of groups of collaborative robots to be apportioned among one or more co-workers; liability of one co-worker in relation to several collaborative robots it works with; and so forth), but the fundamental challenge is that of apportioning the degree of liability associated with an outcome, between the algorithm (powering the collaborative robot) and humans. The next step is to decide how (or even whether) to redistribute the collaborative robotic degree of liability among some humans – in that case, the identification of the ‘chain of command’ is yet another challenge. Liability is based on fault and falls within the scope of tort law and in the H-R setting it revolves around compliance not merely with legislation, but with technical standards as well [133]. Legislation ‘apportions’ liability among agents, but whether the latter will be found, for instance, negligent frequently depends on their alignment with the relevant industry standards, as well as on the way contracts had defined the distribution of roles and functional prerogatives within the organisation.

What is, then, the distinction between liability and responsibility? Responsibility *can be* broader, or narrower, than liability, because it is assigned according to principles of ethics, moral stances, and philosophical categories [134], as opposed to the law per se. If we accept this distinction, then we can also accept that an algorithm, while not meaningfully standing as an assignee of liability, can be considered ‘responsible’ for the outcome of a decision-making process, depending on whether we address it ‘in its own right’. In a collaborative setting, this translates into identifying a ‘degree of responsibility’ that the collaborative robot can be endowed with, irrespective of the human. That part of responsibility will not be reassigned to humans: it will stay with the autonomous system. Here, ethical co [135]nsiderations play a substantive role: to exemplify, that of pondering, *to what extent would it be desirable to consider collaborative robotic intelligence responsible in lieu of an ultimate human intelligence?* When is it worth, or meaningful, or feasible, or desirable, to consider a non-human decision-maker ‘responsible’? As de Sio and Mecacci (2021) probe, when smart collaborative robots themselves enforce occupational safety rules, and an incident occurs, those collaborative robots might be attributed with responsibility due to the lack of human supervision [77].

Accountability displays an additional connotation, which is *relational* and mainly of relevance for public policy. One can *be*, or *feel*, accountable. In the second case, no immediate consequences arise for betraying social and public expectations; in the former case, instead, expectations are somewhat ‘enforced’ by those holding them – generally public opinion, lobbying and interest groups, nongovernmental organisations, or the citizenship at large. A human co-worker can be, or feel, accountable towards oneself, the collaborative system they are teaming up with, or both, but also to third parties within the organisation (e.g. colleagues) and other industry or societal stakeholders, including investors, consumers, and so on. Accountability may also be felt towards more abstract entities such as the broader society, the environment, or the global village – virtually anything and anyone. It is thus necessary to examine whether there is anything ‘special’ about accountability *in collaborative settings*. Just like most of other modes of being, feeling, and behaving, the special trait rests with the distributional expectations: a co-worker might want to be accountable, or might be held accountable without necessarily wanting to, for actions that are in fact the joint product of human-algorithmic co-agency.

Instead, liability in the EU is to be addressed in two further legislative proposals: a revision of the Product Liability Directive (which has now been finalised) and a new AI Liability Directive (which is still under discussion). Harmonising liability rules at least across the EU will be welcome, although the substance of these proposals may fall short – Hacker [136] considers that while they make ‘steps in the right direction, they ultimately represent a half-hearted approach’. This shortfall may impact upon the extent to which the legislative proposals address the aforementioned issues identified in the Delphi study for H-R relations.

4.2.2. Data privacy, protection, and security requirements in H-R collaboration

In the context of H-R collaboration, how the organisation uses data gathered on the shop floor is indeed paramount. Critically, in many jurisdictions, data privacy, protection, and security frameworks such as the EU's GDPR (General Data Protection Regulation), impose requirements impacting various stages of the collaboration lifecycle, including data collection, processing, storage, and sharing.

PPS-2: *Informed and voluntary consent must be obtained for data collection and processing related to work activities in H-R interaction and collaboration.*

In the EU, organisations processing personal data must adhere to GDPR principles such as data minimisation, purpose limitation, and data subject rights to ensure that personal data collected from human workers or generated by robots is handled in a lawful, fair, and transparent manner. However, given the novel work paradigm involving H-R collaboration (e.g. data collected from the implementation of monitoring systems), the participating experts argued that adjustments and refinements are required to effectively address the challenges associated with the collection and use of data from the shop floor (which may include ML/AI data from the actual H-R collaboration).

The ethical design goals outlined within this principle focusing on data privacy, protection, and security requirements may not encompass all potential challenges that could arise in H-R collaboration. Addressing every scenario affecting the implementation of these goals would require specific studies. Thus, we have chosen to highlight only the aspects that were most debated among our participating experts.

The issue of worker consent regarding the extensive data recorded via various devices in the context of H-R collaboration presents a multifaceted challenge. While it is imperative that workers are fully informed about the data collection and its intended usage, obtaining their consent in a transparent and non-coercive manner is crucial. The participating experts underscored the importance of informed consent in the context of data processing, particularly within employment relationships. However, they noted that while informed consent holds ethical weight, it may not typically serve as the legal basis for processing personal data under the GDPR, especially within the framework of employment dynamics due to the inherent power imbalance that characterises employer-employee relationships [137,138]. Yet, achieving a genuine understanding of the data being recorded and its implications may prove challenging, owing to the complex nature of big data analytics. In line with this, one ethicist raised the issue of '*an illusion of informed consent*', particularly in contexts where individuals are asked to consent to various interactions with technology.

> **Another worry is that if people working in this setting are asked to consent to the recording of data about their work with the robots, they may not fully understand what they are consenting to - so there might be an illusion of well-informed and free consent, rather than truly informed consent. A typical issue when it comes to consenting to different forms of interactions with technologies.** *[Prof Sven Nyholm, Professor of the Ethics of Artificial Intelligence, LMU Munich, Germany]*

PPS-1: *Collect and retain only the necessary data for the intended purpose of improving safety and productivity, minimising the risk of privacy infringement.*

Within H-R collaboration, this issue becomes increasingly salient, as workers may not possess a comprehensive understanding of the implications of consenting to the recording of data pertaining to their interactions with robots. The complexity of these technologies and data collection processes may obscure the true nature and consequences of such consent, potentially leading to a superficial or illusory understanding among workers. While consent may not be the correct legal basis for processing employees' personal data under the GDPR, at least, from a broader ethical perspective, genuine understanding and consent from employees is desirable. In any case, from both ethical and legal perspectives, employers must establish clear policies and procedures for data collection, storage, and retention, as well as implementing technical safeguards such as encryption and access controls to protect sensitive information.

4.2.3. Un-biased ML/AI data processing

One pressing issue concerns the potential bias that ML/AI algorithms may introduce, particularly impacting on gender diversity, minorities, or other individual factors such as illness, personal problems or disability [139]. Critically, bias in algorithms can arise from various sources, including biased training data, flawed algorithms, and human decision-making processes that influence algorithm design and implementation.

AI-1: *Ensure transparency in data collection and processing methods of ML/AI data, and continuously monitor and audit ML/AI systems for algorithm biases, making adjustments as necessary to ensure fairness.*

The processing of data by 'intelligent' algorithms in the workplace raises crucial questions regarding transparency and worker awareness of the purposes for which their data is being collected and how it is being processed. Indeed, transparency plays a crucial role in the design and implementation of systems involving data processing and self-learning algorithms. Additionally, the practice of big data analytics raises concerns about the ethical implications of implying 'correlations' without considering causality or context, as this can lead to unjust actions and reinforce biases within H-R collaboration settings. Opacity and lack of accountability in decision systems pose significant challenges regarding understanding and explaining algorithmically-derived intelligence.

> **Transparency is key - all critical decisions made by system should be explainable afterwards and verifiable beforehand.** *[Dr Bianca Himenez, Chief Data Officer, Ethical Machine Learning, PUR, France]*

AI-2: *Limit ML/AI data processing to the specific purposes disclosed to workers, ensuring that the collection and processing of ML/AI data align with relevant privacy laws and ethical standards.*

The extensive gathering of data on the shop floor to fulfil the safety-related requirements, poses additional question to the use that can be made outside of the intended purposed.

> **There's a lot of data being gathered, especially of the employees. This system will not only record human-robot interactions, but literally anything happening on the shop floor. I understand that this is a requirement for such a system to work, but still it would be wise to delete any data that is not necessary for the goal of developing and enhancing a safe and reliable working environment for the employees' privacy.** *[Prof Michael Fisher, Professor of Computer Science, University of Manchester, UK]*

4.2.4. Safety and wellbeing

This theme, originally proposed as two distinct ones in questionnaire 3 of the Delphi study (one for physically safe workspace environments, and the other for psychologically safe workspace environments), has been merged following comments suggesting to 'thread it together into circular physical-mental issues'. This principle encompasses both the physical and psychological safety considerations –which can include mental wellbeing strategies.

SAW-6: *Ensure that all safety practices and equipment meet or exceed relevant regulatory requirements and industry standards.*

Physical safety is indeed paramount in the design and implementation of robots used in H-R collaboration [140]. Ensuring health and safety is a legal requirement, and is implemented through, among other measures, safety standards governing the robotic system design and control operation (e.g. Refs. [6,7,77,141]. As in I.5.0 the aim is to eliminate all physical barriers between workers and industrial robots to facilitate mixed team collaboration, the challenge lies in ensuring safety measures are in place to mitigate or eliminate the risk of physical harm to humans. From one side, safety features are embedded into the design of collaborative robots. In addition, perception systems installed on the manufacturing floor are meant to monitor and detect the physical working environment, and therefore support the performance of effective and safe H-R collaboration tasks [142,143]. From the other side, the challenge is to research and establish ML/AI safety-critical systems capable of learning from data without currently having a formal method for verification and validation.

SAW-2*: Incorporate safety measures into the design and development of collaborative robots to ensure the physical safety and wellbeing of human workers, thereby minimising the risk of accidents on the shop floor.*

The integration of advanced technologies into work processes can potentially overwhelm human workers, increasing the likelihood of safety incidents and operational disruptions. This phenomenon arises from the need for workers to balance their own cognitive abilities and creative capacities with the sophistication of robots. A recent study conducted by Karakikes and Nathanael [106] highlighted that workers are more effective in assisting collaborative robots when decision authority is assigned in alignment with an appropriate cognitive workload. Importantly, in scenarios where workers perceive a high workload, H-R performance decreases [106]. However, the amount of influence over task allocation can also increase the perceived workload in human workers [144]. In line with the above, the ethicists participating in the Delphi study suggested that the design and allocation of H-R collaborative tasks should not solely prioritise performance and economic criteria but should also consider the dispositional aspects of human workers, their cognitive abilities, and their overall sense-making process, with the ultimate goal of fostering a more ethical and human-centric environment.

> **People have very different ways of demonstrating overwork, stress, mental fatigue etc from person to person. A robotic system may not have the appropriate programming or indeed sensors to detect when a person is reaching their limit, and so might either suggest inappropriate workloads for workers or, indeed, push a worker beyond their limit to their detriment.** *[Dr Murray Sinclair, Visiting Scholar, Loughborough University, UK]*

To this end, concerns regard the adequacy of experimental validation, as it may not fully encompass all potential real-world scenarios—this issue becomes apparent when considering tests and accidents involving autonomous driving cars. This raises a further question: *is it sufficient to demonstrate that autonomous systems make fewer errors or cause fewer fatalities than human drivers to justify their deployment?* And if applying this concern to the manufacturing issue: *What constitutes a reasonable threshold for safety testing before permitting human interaction in these emerging H-R collaboration environments?* Finally, ensuring a physically safe work environment in H-R collaboration settings requires a comprehensive approach that integrates safety and ethical behaviour into organisational safety culture [145,146].

> **Safety and ethical behaviour have to be embedded in organisations. Although regulatory requirements help, it has to be a cultural change.** *[Dr Mona Ashok, Associate Professor of Digital Transformation, Henley Business School, UK]*

> **I LOVE the tenet of encouraging a culture of accident reporting, because this is something that makes it clear that technology alone doesn't solve; there's a need for ongoing process and management that keeps this CULTURE alive in the company. Ethics is directly linked to one's culture, so this is very powerful.** *[Dr Bianca Himenez, Chief Data Officer, Ethical Machine Learning, PUR, France]*

SAW-4*: Encourage a (just) culture of reporting safety incidents, even if they do not result in harm, to improve safety practices continuously.*

Encouraging an ethical culture of reporting safety incidents, even those that do not result in harm, is crucial for continuously improving safety practices in H-R collaboration settings. However, it may not be sufficient to address concerns about reporting and accountability, particularly if workers feel pressured by the fear of replacement. In such cases, it was suggested that additional structural measures may be necessary, such as safety inspections conducted by third-party services, and/or mandatory data recorders (commonly known as black boxes) in all collaborative robots (as confirmed in the specific expert input reported below). The aviation industry has a longstanding tradition of conducting thorough investigations on a no-blame basis. While this approach may be deemed too costly for everyday delays in production, it nonetheless offers a model for addressing safety-related incidents [147].

> **"Encourage an ethical culture … " may not be enough if people are under the pressure of being fired or replaced. They may simply not dare. More structural measures may be needed, like for instance external controls.** *[Dr Pericle Salvini, Senior Researcher, University of Oxford, UK]*

> **Two additional items could be included as requirements: 1) the mandatory use of data recorders (aka black boxes) to be incorporated in all collaborative robots in order to improve safety and the reconstruction of accidents; 2) safety inspections by third party services.** *[Dr Filippo Santoni de Sio, Associate Professor Ethics and Philosophy of Technology, Delft University of Technology, The Netherlands]*

SAW-1*: Provide safety training to human workers collaborating with collaborative robots, focusing on understanding, and mitigating the risks associated with H-R interaction and collaboration.*

Human workers should be encouraged to report even minor incidents, fostering an organisational 'just culture' [148,149] and continuous improvement mechanisms. This includes adequate training and periodic updates for worker to equip them with the necessary skills and knowledge to prevent accidents and respond effectively to incidents. Training programs should cover a range of topics, including operating procedures, safety protocols, emergency response protocols, and troubleshooting techniques. Additionally, periodic updates and refresher courses will help reinforce key-concepts.

> **All necessary safeguards must be in place to prevent accidents and problems, and operators should be trained and updated periodically. Operators should be encouraged not to minimise even minor incidents.** *[Dr Sarah Fletcher, Principal Research Fellow in Industrial Robotics Psychology, Cranfield University, UK]*

SAW-3*: Avoid intrusive or constant surveillance practices, as they can contribute to developing a surveillance culture and have adverse effects on the psychological wellbeing of human workers.*

As H-R collaboration requires constant monitoring for safety purposes, there might be the risk that monitoring data can also be used for performance purposes. This can raise concerns of trust in the employer-employee relationship and dynamics; therefore, the psychological impact of H-R collaboration needs careful consideration, as only some issues which may be triggered by this interaction can be now predicted. The psychological effect of this practice has been discussed already, after the Amazon report has been published [86], with some initial regulatory countermeasures and sanctions taken.[3] In our study, all participants argued that working in an 'always-observed' environment can lead to both detrimental and unbeneficial psychological effects. As our study ethicists noted, constant observation may induce feelings of stress, anxiety, and a sense of invasion of privacy among workers.

[3] On 23 December 2023, the French Data Protection Authority has fined Amazon France Logistique of eur 32 million for setting up an excessively intrusive system for monitoring employee activity and performance. The company has also been fined for video surveillance without information nor sufficient security. *Source: REUTERS.*

Information about human performance, robot performance, and human-robot-collaboration performance will allow for an extremely detailed evaluation of the efficiency of each human operator like a machine. This, under the pressure of increased productivity and decreased costs, could lead to a further dehumanisation of work in factories (e.g. Foxconn) or logistics hubs (e.g. Amazon) with inevitable psychological problems. In particular we can identify the following classes of problems: 1. Violation of privacy by spying on any gesture or behaviour, even private, of the worker. 2. Risk of discrimination if workers may be penalised for individual factors such as illness, personal problems, or disability. 3. Mental health endangered by a stressful and oppressive work environment, where workers feel constantly watched and judged. 4. Reduction of autonomy and responsibility and therefore of the possibility of developing one's own skills and competences. 5. Lack of worker confidence in the organisation and therefore of personal motivation. *[Dr Marco Veruggio, Robotics Scientist, Scuola di Robotica, Genoa, Italy]*

And additionally:

While it is equally essential to derive performance data for humans, the pressures of being constantly measured and evaluated can quickly lead to a toxic work environment. While we can always increase the capabilities of a machine by continuing to develop technologies, a human is not meant to function at 100% efficiency. They should thus not be held to that standard. If a job can be done faster and more reliably, they will inevitably be replaced by a machine (which is a valid evolution). However, putting them in a digital panopticon prison of data collection is unethical since it will inevitably harm their mental health. It is essential to keep a level of abstraction and create a looser framework to measure human performance, not micromanage every move of their muscles (as can be ethically done with a machine). *[Dr Filippo Santoni de Sio, Associate Professor Ethics and Philosophy of Technology, Delft University of Technology, The Netherlands]*

Psychological hazards in the framework of H-R collaboration may emerge when workers constantly feel the need to compare their performance against the expectations set by the robotic system. Unlike robots, human workers thrive on a rhythm of intensive work interspersed with periods of rest. Currently, human workers may have the flexibility to manage their energy levels and productivity accordingly taking breaks when needed, or balance this with their human co-worker, yet still meeting the expectations of their employers. In a potentially dystopian scenario characterised by constant surveillance, human workers would be under constant pressure to maintain high levels of output. Workers may worry that the feedback from the system reveals them to be less effective than their robotic peers, fearing replacement if they fail to do so. This fear of falling short of expectations can create a sense of pressure and anxiety, impacting the mental wellbeing of employees.

Working together with an automated or semi-automated system for human workers will present challenges in terms of performance expectations. The very nature of robotic / algorithmic entities, being very structured, and time- and movement-precise, will exert psychological pressure upon human workers to perform in a similar precise and efficient mechanical fashion, which is not natural for humans. *[Prof Matthew Studley, Professor of Ethics and Technology, Bristol Robotics Lab, UWE Bristol, UK]*

On the other hand, if the robot performs at a slow pace—slower than what is expected by the human worker—although can be provide this may provide opportunities for humans to take breaks, it may also create a sense of frustration. When workers experience delays or inefficiencies in tasks performed by robots, it can lead to feelings of impatience, irritation, and stress.

Even though the topic is called human-robot collaboration, the robot is always there to support the human. Either by fully taking over certain tasks, or supporting the human where they need the support. From my ethical perspective, it is far more dangerous for a robot to push a person than it is for a robot to be too slow and needing to be pushed by the human. Frustration due to simple/slow actions is unwanted and should be avoided, if possible, but it indirectly also gives people the opportunity to take a step back and unwind, even for just a second, as it's not possible to work at peak performance at all times. It is far more dangerous for a robot to push a human during collaboration. True, sometimes a person may need a push, but there are many other reasons why a person is not working as fast as they can. They may be fatigued or stressed out or simple have aching joints from a workout the day before. In these cases, it is not recommended to push the worker as this may lead to physical injuries. But besides physical injuries, it may also hurt them mentally if they are slower for psychological reasons. *[Dr Anouk Van Maris, Senior Projectlead, Zuyderland, The Netherlands]*

***SAW-5**: Establish a confidential feedback system that allows human workers to report psychological concerns, ensuring anonymity and protection against retaliation.*

As the nature of work is reshaped, with H-H teams being progressively replaced by H-R teams, concerns such as work dehumanisation and alienation require deeper examination [150]. The changing nature of work with collaborative robots may exacerbate feelings of social isolation among workers, particularly if they are physically isolated from human co-workers. This sense of isolation can negatively impact morale, communication, and teamwork within the organisation. In this context, the feeling of being stripped of opportunities for meaningful work interaction can manifest as a sense of powerlessness and alienation, as individuals feel disconnected from their work and their own sense of identity and existential distress.

Social interaction and interacting with other humans (as opposed to intelligent systems) seems, for most people, a considerable source of well-being and mental good-health. There are perhaps risks of an excessively automated environment being 'de-humanised', i.e. lacking opportunities for human-human interaction, which can both lower mental well-being and be de-humanising. *[Dr Neil Sinclair, Professor of Philosophy, University of Nottingham, UK]*

The psychological impact of H-R collaboration on workers will encompass the specific dynamics of how collaborative robots interact with human counterparts and the resulting psychological ramifications. Future research should also focus on the communicative capabilities of robots and their potential effects on human psychology, with verbal and non-verbal clues employed by robots during collaboration. This also includes the physical movements of the robots and perceived safety by human worker. Certain movements deemed optimal for robots may appear hazardous or intimidating to workers, leading to heightened stress and discomfort.

Promoting open and transparent communication between human workers and management to address concerns, questions, and feedback related to H-R interaction and collaboration on the shop floor can help identify countermeasures to tackle these issues. To implement such a feedback system, organisations should prioritise confidentiality and anonymity to protect the privacy and identity of workers who come forward with concerns. Furthermore, it is important for organisations to establish protocols for responding to feedback and addressing psychological concerns in a sensitive and compassionate manner. This may involve providing access to counselling services, implementing workplace wellness programs, or offering Supporting and resources to workers experiencing stress or anxiety related to H-R collaboration.

4.2.5. Human resilience through continuous learning and support

While the adoption of robots by employers may lead to increased efficiency and higher profits, there is a growing recognition of the need to reinvest some of these gains into improving workers' skills and competencies [151]. These skills should complement rather than compete with robotic advancements. When workers lack the technical expertise required to operate or troubleshoot robots effectively, they may perceive being subordinate to robots (i.e. '*power imbalance*' as also discussed in section 4.1.2). In this context, organisations play a crucial role in fostering a culture of continuous learning and skill development among workers.

RES-1*: Support worker personal development to empower them to navigate their roles safely and effectively, fostering a culture of continuous learning and skill enhancement.*

Critically, providing opportunities for workers to bridge any skills gaps to engage meaningfully with robotic technology can help foster a sense of agency and empowerment in the H-R collaboration settings [114]. Additionally, a bottom-up approach, such as actively involving employees, for example, in soliciting their input on how collaborative systems can be designed and improved, can ultimately lead to more effective and human-centred solutions[114].

Supporting worker personal development can provide human workers with expert knowledge and adaptability skills to effectively tackle and resolve challenging or unanticipated situations in the workplace. Additionally, reskilling efforts should aim to ensure that the skills acquired are transferable to other workplaces, recognising that technical skills tailored to a specific machine or process may not necessarily translate to other organisational contexts.

> **Does this training encompass a broad range of skills, or is it limited to the application of the particular system/set of robots employed by that particular workplace, are those skills transferable? A joiner or a mechanic can take those skills to another workplace, but a worker trained on one specific machine might find those skills are not applicable elsewhere. Effectively tying them down to that particular place or a small set of places.** *[Dr William Knight, Research Ethics and Integrity Officer, De Montfort University, Leicester, UK]*

The concept of resiliency and resilient operators have been primarily been explored in the HF literature, particularly in reference to those (human) systems that display autonomy (i.e. adaptive resources and capacity) to ensure safety by effectively managing and mitigating ambiguity and uncertainty [152]—this concept is also referred to as "*discretionary space*" or "*margins of maneuver*" [153,154]. Within the context of H-R collaboration, providing operators with discretionary space could imply that skilled workforce have the knowledge and corrective capacity to intervene in the event of a systems failure [155]. This defines the need for encouraging the development of 'Resilient operators', an extension of the notions of 'Operator 4.0' [114,156,157].

> **There's a lot of stuff about the 'Resilient Operator' that comes into play in H-R collaboration, including organisational culture issues impinging on the relationships between the operator, the equipment, the management. It is evident that it is not easy to do that leads to better, more reliable outcomes. There is a reason why people are the ultimate source of resilience when all else fails. There is a reason why we still have pilots in highly sophisticated aircraft. And to be resilient in these contexts requires considerable training, and excellent support so that then can maintain their 'margins of manoeuvre' over extended periods, decades, even. There is a big issue about operators who cannot adjust, also about loss of knowledge and of co-operative behaviour if people are fired. There's a lot of ethical considerations lurking in all of this.** *[Dr Murray Sinclair, Visiting Scholar, Loughborough University, UK]*

RES-2*: Support leadership practices that prioritise the wellbeing and psychological safety in H-R interaction and collaboration.*

Specific training and coaching sessions focused on topics such as communication, conflict resolution, and emotional intelligence can sustain human workers to become capable agents able to interpret situational nuances, weighing trade-offs, and making informed decisions that align with organisational goals and values. The expectation of worker competence extends beyond technical proficiency to encompass ethical considerations and social responsibility. Moreover, middle and senior managers are challenged to engage in upskilling and reskilling initiatives, to acquire the needed skill set for effectively both supporting the integration of technological advancements within the factories, and the workers on the shop floor. This raises questions about the potential capabilities managers should acquire to effectively lead in an automated workplace.

> **If we take reskilling seriously then the profession most in need of change is management.** *[Dr Blay Whitby, Technology Ethicist, Imperial College, London, UK]*

Furthermore, organisations should recognise and reward ethical leadership behaviours, reinforcing the importance of these practices. This may involve highlighting examples of ethical leadership in action, and incorporating ethical leadership criteria into performance evaluations and promotions.

4.2.6. Human dignity and equity

In line with the core principles of I.5.0, promoting human dignity and equity is essential for creating inclusive human-centric, sustainable, and resilient workplaces.

HDE-1*: Implement policies and procedures to respect, retain and enhance human worker expertise and value.*

The knowledge, skills, and abilities of individuals must be considered a valuable asset for the future industry to be respected, valued, and protected in all aspects of work. The reliance on automation has the potential to reduce the need for human workers, leading to a potential loss of valuable expertise and knowledge accumulated over years of experience. This issue has been a longstanding challenge in many domains, with the digitalisation of expert knowledge posing significant challenges for industries that heavily depend on human expertise for their operations [158–160].

In addition to preserving and retaining expert knowledge [159,161], the I.5.0 organisation of work should prioritise the growth of workers' knowledge and expertise, all while respecting their dignity. This entails designing task allocation and job roles in a way that promotes continuous learning, skills development, and career advancement opportunities for all employees. The question also regards which skills and tasks should the future worker perform. Although robots will increasingly perform tasks that are repetitive or injury-prone, we must consider that many workers can find fulfilment in performing routine tasks that do not require significant cognitive demands [72]. Therefore, organisations should provide their workers with opportunities that align with their unique interests and talents, rather than channelling them into predetermined paths—an embodiment of the ethical value of freedom of choice.

> **Whereas relieving workers from physically demanding and dangerous tasks clearly has a positive impact on their physical integrity and helps them to remain healthy and able to work for a long time and into old age. In terms of workers' wellbeing, the situation is more complex and requires a more in-depth analysis due to people's heterogeneity. Certainly, not every worker will feel relieved when he/she is taken away repetitive tasks. Some**

may feel effective in doing these tasks and might enjoy not having to think about complex challenges (i.e. their individualistic need for cognition is low and they have a strong adversity against high cognitive load). Hence, they might feel overburdened and less valued if these tasks are taken away suddenly. Others may indeed see opportunities for upgrading their skills, which could also lead to an increase in material wellbeing. *[Dr Tobias Kopp, Postdoctoral researcher, Karlsruhe University of Applied Sciences, & Business consultant, esentri AG, Ettlingen, Germany]*

New forms of work collaboration with robotic system pose fundamental concerns about the dignity and rights of workers in the digital age. As previously discussed, if robots set the standards for human performance and behaviour, there is a risk of '*instrumentalisation'* of humans. This not only undermines the autonomy and agency of individuals but also fails to recognise the inherent value and dignity of human beings.

One worry I would have about this is that people are expected to be/behave like - and live up to standards set by - robots. Having people work with new technologies is always a social form of experiment, since it is hard to predict what the effects on the humans will be - so a worry here is that there might be psychological risks that are hard to predict. And again, the worry might be that the people working in this context might be treated like machines - like robots - rather than as human beings. This can be seen as failing to fully respect human dignity. *[Prof Sven Nyholm, Professor of the Ethics of Artificial Intelligence, LMU Munich, Germany]*

HDE-2*: Ensure that human workers receive fair and competitive compensation for their roles in H-R interaction and collaboration on the shop floor, taking into account their skills, contributions, and the changing nature of work.*

Fair compensation acknowledges the value that human workers bring to the collaborative effort and recognises their contributions to achieving organisational goals. It should reflect not only the technical tasks performed by workers but also non-technical skills, such as their ability to adapt, problem-solve, and innovate in the context of H-R collaboration. Critically, organisations should be mindful that while automation can enhance efficiency and productivity, there is a risk that it may inadvertently stifle innovation by restricting the scope for human creativity. Competitive compensation ensures that human workers are adequately rewarded for their efforts and remain motivated and engaged in their work. It also helps attract and retain talent in a competitive labour market, ensuring that organisations have access to the skills and expertise needed to drive innovation and success in H-R collaboration.

The opportunities for creativity and to improve production methods and approaches may be limited by the robot capabilities or constraints. This could lead to a loss of innovation. The focus is often on the human having to adapt to the robot, whereas the expectation was that robots would assist humans and improve productivity, the reality is, that as robots are less flexible than humans, it is the human that has to adapt to the robot needs. This may mean that humans working in such a scenario may consider themselves (and their employers too) as less valuable than the robot, and to concede to the robots needs ahead of their own. This subservience to the mechanical may have a negative impact on self-esteem, job satisfaction. It may also lead to frustrations and dissatisfaction with their work. *[Sara Wilford, Associate Professor, De Montfort University, Centre for Computing and Social Responsibility, UK]*

5. Conclusions

This article has presented an ethical framework addressing emerging ethical considerations within the context of H-R collaboration in future I.5.0 manufacturing. Notably, this framework is the result of a collaborative effort developed through one-year long engagement with 30 experts in ethics that has involved iterative cycles of data collection, analysis, and interpretation. Therefore, the proposed framework encompasses a co-constructed approach that integrates diverse perspectives and expertise.

This ethical framework is adapted to manufacturing context, with ethical principles that underpin either the specific H-R dynamics present throughout the manufacturing floor, or at organisational level to ensure that human workers receive the appropriate resources for their competencies and skills to be effectively utilised and expressed [162]. The organisational governance at meso-level underscores the role of employers and organisations to organise H-R collaboration labour in a way that promotes the growth of workers' knowledge and expertise while respecting their dignity [163].

The ethical considerations of H-R collaboration in the field of manufacturing are multifaceted and have far-reaching consequences that extend across various levels of operation. Critically, these ethical issues can propagate through the complex network of interactions within manufacturing ecosystems, much like ripples in a pond. This ripple effect can be particularly pronounced in open complex systems, where multiple elements and agents interact, inter-depend, and influence one another. Indeed, these systems are characterised by numerous inter-connected elements, both human and non-human (e.g. robots, supply chains, regulatory bodies), which interact and adapt to changes. In such an environment, the ethical implications of H-R collaboration may become deeply entangled with the wider ecosystem. Ensuring ethical behaviour for sustainable manufacturing in I.5.0 involves managing these complex relationships and their ethical implications, which can be challenging due to their dynamic and evolving nature.

In our framework, human workers will continue to play a crucial role in overseeing and optimising these automated processes, ensuring their efficiency and reliability. As Coeckelbergh [23] argued "*we should ask not only what can humans still do in the new automation economy and society but also what we want them to do*" (p.45). Notably, even the choice of terminology used to describe the future organisation of labour (i.e., H-R collaboration) can significantly influence our perceptions of the relationship between humans and robots in the workplace. While terms like 'collaboration' or 'co-working' may imply a sense of partnership between humans and robots within industrial settings, it is important to acknowledge the distinct roles and contributions of each party in the collaborative process. Being mindful of the language we use to describe this type of work can help underscore the human-centric approach in designing the future organisation of labour, with the responsible integration of automation while prioritising the wellbeing and growth of human workers.

While we retained the terminology of 'H-R collaboration' as used in the literature and shared by scholars in the field, we suggested a hierarchical relationship between humans and robots, with robots serving as 'aids' to human workers. In manufacturing industries, well-defined roles and responsibilities specifications are typically established. Within the context of H-R collaboration, it is recognised that retaining some degree of flexibility with the human workers is important, to accommodate varying situations occurring during the interaction with collaborative robots.

5.1. Implication for research and practice

Ethical frameworks, such as the one presented in this work, should primarily serve as ethical design guidance for anybody who designs H-R systems, specifically for the manufacturing sector. Additionally, it may guide and support all involved stakeholders within industry (including managers, employees, unions) to help prioritise human safety, wellbeing and dignity in environments where future collaborative robots will operate. Transparency in communication about the capabilities and limitations of robotic systems, as well as the decision-making processes, is vital to maintaining trust in H-R collaboration. Embracing ethics within the design of these systems should be seen as a positive action which can increase technology adoption. Incorporating this into cutting-edge technical research and embedding it into the thought process of future engineers and designers is crucial.

Academia plays a critical role in advancing ethical frameworks for H-R collaboration in manufacturing by engaging in interdisciplinary research to ensure a comprehensive approach to robot ethics. Actions could include the integration of ethical concerns and human-centric considerations into university curricula (across a range of subjects), raising awareness and sensitivity towards students' "*ethical capability*"—i.e. the cognitive and emotional resources needed to apply ethical decision-making approaches and robot design requirements [164]. Additionally, engaging with broader societal stakeholders can enhance the implementation of these frameworks. For instance, involving industry stakeholders in conducting and publishing case studies on various H-R interactions can provide valuable insights into practical challenges and requirements, bridging the gap between theoretical frameworks and real-world applications. Furthermore, academics can contribute to the development of regulatory standards and provide evidence-based recommendations to policymakers, helping to establish robust legal frameworks governing robotic systems.

5.2. Considerations for future research

Recognising the need for further refinement and research to address the ethical and legal issues surrounding H-R collaboration in future manufacturing that are presented in the current work, several areas warrant additional consideration. First, methodologically, since we employed feedback and insights mainly from European academic ethicists to lay the foundations of the current ethical framework, a more systematic external validation process is essential. Indeed, expanding our understanding by incorporating the perspectives of a wider stakeholder group, such as industrialists and worker representatives, is particularly crucial to ensure its relevance and effectiveness in real-world manufacturing environments. These further studies may employ quantitative research designs aimed at refining and ranking ethical principles. Secondly, the use of prospective scenarios informed by the literature served as *elicitation prompts* for the participating experts to provide their insights on the study phenomena. While the additional iterations of data collection and analysis helped reduce the potential biases in the responses—and, consequently, in the designed framework—that could have been influenced by the themes identified in the literature review, we acknowledge that using different scenarios might have led to the generation and co-construction of different insights.

The meaning and role that ethics should play in the design of future industry can differ among I.5.0 stakeholders— *What strategies should be employed to convey a shared understanding of the challenges in H-R collaboration, ensuring that all stakeholders can effectively leverage the role of ethics in this evolving configuration?* This process, akin to the '*socialisation*' of the terms, involves fostering a common language and framework around ethical considerations that all stakeholders can reference and adhere to. Indeed, there is a need for inclusive dialogue involving various stakeholders to design ethical frameworks that reflect shared values and societal goals. This will not only enhance the credibility and acceptance of the ethical framework but also ensure its effective application across various industrial contexts, ultimately contributing to the sustainable and responsible development of future industry. Furthermore, the involvement of engineering and safety professionals (both academic and practitioners) can add insights on the trade-offs between economic and ethical considerations: *How do organisations balance the trade-offs between economic efficiency and ethical considerations in the implementation of H-R collaboration in manufacturing environments?* Although we are aware that in the future collaborative robots may achieve increasing levels of autonomy, we recognise that eventually humans and collaborative robots may attain divergent interests and objectives, which may not always align with conventional notions of ethics or productivity. Their respective pursuit of interests and optimisation of outcomes can therefore vary, highlighting the complexity of ethical decision-making in such interactions. It becomes essential to understand the nuanced patterns of interest pursuit, particularly within groups and amidst competing interests from all involved stakeholders.

Finally, we recognise that different societies (such as, for example, those in East Asia) might have distinct approaches to ethics in H-R collaboration. This recognition necessitates specific investigation in future studies. We emphasise, however, that this diversification must be based on prior epistemic diversification within the techno-policy lexicon and doctrines themselves, as essential resources for all experts in smart manufacturing, regardless of their nationality or institutional affiliation.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: EPSRC (Engineering and Physical Sciences Research Council) and ISCF (Industry Strategy Challenge Fund) under the Made Smarter scheme No EP/V062158/1 Made Smarter Innovation - Research Centre for Smart Collaborative Industrial Robotics (UK) are gratefully acknowledged.

CRediT authorship contribution statement

Tiziana C. Callari: Writing – original draft, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Riccardo Vecellio Segate:** Writing – review & editing. **Ella-Mae Hubbard:** Writing – review & editing, Visualization. **Angela Daly:** Writing – review & editing. **Niels Lohse:** Writing – review & editing, Funding acquisition.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors are deeply grateful to the subject-matter experts participating in this research, in all rounds of the data collection, for trusting us and sharing their knowledge and insights. We also are thankful for the inputs received from the Made Smarter Innovation project partners for their insights throughout the development of the study.

References

[1] C. Bai, P. Dallasega, G. Orzes, J. Sarkis, Industry 4.0 technologies assessment: a sustainability perspective, Int. J. Prod. Econ. 229 (2020) 107776, https://doi.org/10.1016/j.ijpe.2020.107776.
[2] C.O. Klingenberg, M.A.V. Borges, J.A.d.V. Antunes, Industry 4.0: what makes it a revolution? A historical framework to understand the phenomenon, Technol. Soc. 70 (2022) 102009, https://doi.org/10.1016/j.techsoc.2022.102009.
[3] A. Hentout, M. Aouache, A. Maoudj, I. Akli, Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017, Adv. Robot. 33 (15–16) (2019) 764–799, https://doi.org/10.1080/01691864.2019.1636714.
[4] A. Weiss, A.-K. Wortmeier, B. Kubicek, Cobots in industry 4.0: a Roadmap for future practice studies on human–robot collaboration, IEEE Transactions on Human-Machine Systems 51 (4) (2021) 335–345, https://doi.org/10.1109/THMS.2021.3092684.
[5] L. Liu, F. Guo, Z. Zou, V.G. Duffy, Application, development and future opportunities of collaborative robots (Cobots) in manufacturing: a literature review, Int. J. Hum. Comput. Interact. (2022) 1–18, https://doi.org/10.1080/10447318.2022.2041907.

[6] ISO, 10218-1: Robots and robotic devices — Safety requirements for industrial robots — Part 1: Robots (2011).
[7] ISO, 1, 0218-2: Robots and robotic devices — Safety requirements for industrial robots — Part 2: Robot systems and integration (2011).
[8] ISO/TS, 15066: Robots and robotic devices — Collaborative robots (2016).
[9] A. Reiman, J. Kaivo-oja, E. Parviainen, E.-P. Takala, T. Lauraeus, Human factors and ergonomics in manufacturing in the industry 4.0 context – a scoping review, Technol. Soc. 65 (2021) 101572, https://doi.org/10.1016/j.techsoc.2021.101572.
[10] Z.M. Bi, C. Luo, Z. Miao, B. Zhang, W.J. Zhang, L. Wang, Safety assurance mechanisms of collaborative robotic systems in manufacturing, Robot. Comput. Integrated Manuf. 67 (2021) 102022, https://doi.org/10.1016/j.rcim.2020.102022.
[11] A. Bicchi, M.A. Peshkin, J.E. Colgate, Safety for physical human–robot interaction, in: B. Siciliano, O. Khatib (Eds.), Springer Handbook of Robotics, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 1335–1348.
[12] V. Villani, F. Pini, F. Leali, C. Secchi, Survey on human–robot collaboration in industrial settings: safety, intuitive interfaces and applications, Mechatronics 55 (2018) 248–266, https://doi.org/10.1016/j.mechatronics.2018.02.009.
[13] A. Nair, S.V. Pillai, S.A. Senthil Kumar, Towards emerging Industry 5.0 – a review-based framework, Journal of Strategy and Management (2024), https://doi.org/10.1108/JSMA-04-2023-0067 *ahead-of-print*(ahead-of-print).
[14] H. Özköse, G. Güney, The effects of industry 4.0 on productivity: a scientific mapping study, Technol. Soc. 75 (2023) 102368, https://doi.org/10.1016/j.techsoc.2023.102368.
[15] X. Xu, Y. Lu, B. Vogel-Heuser, L. Wang, Industry 4.0 and industry 5.0—snception, conception and perception, J. Manuf. Syst. 61 (2021) 530–535, https://doi.org/10.1016/j.jmsy.2021.10.006.
[16] EC, *Industry 5.0 – Human-Centric, Sustainable and Resilient*: European Commission, Directorate-General for Research Innovation, 2020.
[17] C.J. Conti, A.S. Varde, L. Wang, Human-robot collaboration with Commonsense reasoning in smart manufacturing contexts, IEEE Trans. Autom. Sci. Eng. 19 (3) (2022) 1784–1797, https://doi.org/10.1109/TASE.2022.3159595.
[18] S. Huang, B. Wang, X. Li, P. Zheng, D. Mourtzis, L. Wang, Industry 5.0 and society 5.0—Comparison, complementation and co-evolution, J. Manuf. Syst. 64 (2022) 424–428, https://doi.org/10.1016/j.jmsy.2022.07.010.
[19] B. Wagner, Ethics as an tscape from regulation. From "Ethics-washing" to ethics-whopping? in: E. Bayamlioglu, I. Baraliuc, L.A.W. Janssens, M. Hildebrandt (Eds.), COGITAS ERGO SUM: 10 Years of Profiling the European Citizen Amsterdam University Press, 2018, pp. 84–89.
[20] E. Bietti, From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy, Paper presented at the Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain (2020), https://doi.org/10.1145/3351095.3372860.
[21] A. Daly, K. Devitt, M. Mann, AI ethics needs good data, in: P. Verdegem (Ed.), AI for Everyone?, University of Westminster Press, London: London, 2021.
[22] G.V. Köktürk, Ö.Z. Akbaş, Ö.Ö. Özcan, Sociological dynamics of technological change in the context of society 5.0, Comp. Sociol. 21 (6) (2022) 700–732, https://doi.org/10.1163/15691330-bja10068.
[23] M. Coeckelbergh, Robot Ethics, MIT Press, 2022.
[24] S. Nyholm, Humans and Robots: Ethics, Agency, and Anthropomorphism, Rowman & Littlefield Publishers, 2020.
[25] F. Operto, Elements of Roboethics. Paper Presented at the Makers at School, Educational Robotics and Innovative Learning Environments, Cham, 2021, 2021//.
[26] S. Steinert, The five robots—a laxonomy for Roboethics, International Journal of Social Robotics 6 (2) (2014) 249–260, https://doi.org/10.1007/s12369-013-0221-z.
[27] G. Veruggio, Roboethics, IEEE Robot. Autom. Mag. 17 (2) (2010) 105–109.
[28] G. Veruggio, F. Operto, Roboethics: a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics, in: W. Wallach, P. Asaro (Eds.), Machine Ethics and Robot Ethics, Routledge, London, 2017, pp. 2–8.
[29] S.G. Tzafestas (Ed.), Roboethics: A Navigating Overview, Springer Cham, New York, 2016.
[30] S. Nyholm, Ethical and legal issues related to autonomous vehicles, in: Future Law, Ethics, and Smart Technologies, Brill, Leiden, The Netherlands, 2023, pp. 190–204.
[31] P.M. Asaro, What should ce want from a robot ethic? Int. Rev. Inf. Ethics 6 (12) (2006) 9–16, https://doi.org/10.29173/irie134.
[32] B. Bennett, A. Daly, Recognising rights for robots: can we? Will we? Should we? Law, Innovation and Technology 12 (1) (2020) 60–80, https://doi.org/10.1080/17579961.2020.1727063.
[33] G. Veruggio, The EURON Roboethics Roadmap, in: Paper presented at the 6th IEEE-RAS International Conference on Humanoid Robots, 2006, pp. 612–617. https://www3.nd.edu/~rbarger/ethics-roadmap.pdf, 4-6 Dec. 2006.
[34] G. Veruggio, F. Operto, Roboethics: a bottom-up interdisciplinary discourse in the field of applied ethics in robotics, in: W. Wallach, P. Asaro (Eds.), Machine Ethics and Robot Ethics, Routledge, London, 2017.
[35] M. Boden, J. Bryson, D. Caldwell, K. Dautenhahn, L. Edwards, S. Kember, A. Winfield, Principles of robotics: regulating robots in the real world, Connect. Sci. 29 (2) (2017) 124–129, https://doi.org/10.1080/09540091.2016.1271400.
[36] J.J. Bryson, The meaning of the EPSRC principles of robotics, Connect. Sci. 29 (2) (2017) 130–136, https://doi.org/10.1080/09540091.2017.1313817.
[37] V.C. Müller, Legal vs. ethical obligations – a comment on the EPSRC's principles for robotics, Connect. Sci. 29 (2) (2017) 137–141, https://doi.org/10.1080/09540091.2016.1276516.
[38] T.C. Callari, A.-M. Oostveen, E.-M. Hubbard, S. Fletcher, N. Lohse, Where Are We at? A Review of the State-Of-The-Art of the Ethical Considerations in Human-Robot Collaboration. Paper Presented at the AHFE 2023 International Conference, Applied Human Factors and Ergonomics), San Francisco, CA, USA July 20-24, 2023. https://openaccess.cms-conferences.org/publications/book/978-1-958651-56-8/article/978-1-958651-56-8_3.
[39] S. Nyholm, Attributing agency to automated systems: reflections on human–robot collaborations and responsibility-Loci, Sci. Eng. Ethics 24 (4) (2018) 1201–1219, https://doi.org/10.1007/s11948-017-9943-x.
[40] S.R. Fletcher, P. Webb, Industrial robot ethics: the challenges of closer human collaboration in future manufacturing systems, in: M.I. Aldinhas Ferreira, J. Silva Sequeira, M.O. Tokhi, E.E. Kadar, G.S. Virk (Eds.), A World with Robots: International Conference on Robot Ethics: ICRE 2015, Springer International Publishing, Cham, 2017, pp. 159–169.
[41] A. van Wynsberghe, S. Li, A paradigm shift for robot ethics: from HRI to human–robot–system interaction (HRSI), Medicolegal Bioeth. 9 (2019) 11–21, https://doi.org/10.2147/MB.S160348.
[42] L. Floridi, J. Cowls, A Unified Framework of Five Principles for AI in Society (1) (2019), https://doi.org/10.1162/99608f92.8cd550d1.
[43] H. Roberts, J. Cowls, E. Hine, F. Mazzi, A. Tsamados, M. Taddeo, L. Floridi, Achieving a 'good AI society': uomparing the aims and progress of the EU and the US, Sci. Eng. Ethics 27 (6) (2021) 68, https://doi.org/10.1007/s11948-021-00340-7.
[44] 'United Nations GeneralAssembly', The Universal Declaration of Human Rights (UDHR), United Nations General Assembly, New York, 1948.
[45] EC, Charter of Fundamental Rights of the European Union, Retrieved from, Brussels, 2000.
[46] S.G. Tzafestas, Ethics: fundamental elements, in: S.G. Tzafestas (Ed.), Roboethics: A Navigating Overview, Springer International Publishing, Cham, 2016, pp. 13–24.
[47] G. Veruggio, F. Operto, G. Bekey, Roboethics: social and ethical implications, in: B. Siciliano, O. Khatib (Eds.), Springer Handbook of Robotics, Springer International Publishing, Cham, 2016, pp. 2135–2160.
[48] J. Dreier, Contemporary Debates in Moral Theory, Wiley, 2009.
[49] A. Gustafson, Consequentialism and non-consequentialism, in: E. Heath, B. Kaldis, A. Marcoux (Eds.), The Routledge Companion to Business Ethics, Routledge, London, 2018.
[50] S.G. Tzafestas, Roboethics: fundamental concepts and future prospects, Information 9 (6) (2018), https://doi.org/10.3390/info9060148.
[51] M. Ashok, R. Madan, A. Joha, U. Sivarajah, Ethical framework for artificial intelligence and digital technologies, Int. J. Inf. Manag. 62 (2022) 102433, https://doi.org/10.1016/j.ijinfomgt.2021.102433.
[52] N. Bostrom, E. Yudkowsky, The ethics of artificial intelligence, in: K. Frankish, W. M. Ramsey (Eds.), The Cambridge Handbook of Artificial Intelligence, Cambridge University Press, Cambridge, 2014, pp. 316–334.
[53] J. Chamberlain, The risk-based approach of the European inion's proposed artificial intelligence regulation: some comments from a tort law perspective, European Journal of Risk Regulation 14 (1) (2023) 1–13, https://doi.org/10.1017/err.2022.38.
[54] G. Borges, Liability for AI systems under current and future law, An overview of the key changes envisioned by the proposal of an EU-directive on liability for AI 24 (1) (2023) 1–8, https://doi.org/10.9785/cri-2023-240102.
[55] P.M. Asaro, Robots and responsibility from a legal perspective, Proc. IEEE 4 (14) (2007) 20–24.
[56] A. Guerra, F. Parisi, D. Pi, Liability for robots I: legal challenges, J. Inst. Econ. 18 (3) (2022) 331–343, https://doi.org/10.1017/S1744137421000825.
[57] S.R. Brady, Utilizing and adapting the Delphi method for Use in qualitative research, Int. J. Qual. Methods 14 (5) (2015) 1609406915621381, https://doi.org/10.1177/1609406915621381.
[58] F. Hasson, S. Keeney, H. McKenna, Research guidelines for the Delphi survey technique, J. Adv. Nurs. 32 (4) (2000) 1008–1015, https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x.
[59] F. Hirschhorn, Reflections on the application of the Delphi method: lessons from a case in public transport research, Int. J. Soc. Res. Methodol. 22 (3) (2019) 309–322, https://doi.org/10.1080/13645579.2018.1543841.
[60] C. Okoli, S.D. Pawlowski, The Delphi method as a research tool: an example, design considerations and applications, Inf. Manag. 42 (1) (2004) 15–29, https://doi.org/10.1016/j.im.2003.11.002.
[61] J.W. Zartha Sossa, W. Halal, R. Hernandez Zarta, Delphi method: analysis of rounds, stakeholder and statistical indicators, foresight 21 (5) (2019) 525–544, https://doi.org/10.1108/FS-11-2018-0095.
[62] M. Bloor, H. Sampson, S. Baker, K. Dahlgren, Useful but no Oracle: reflections on the use of a Delphi Group in a multi-methods policy research study, Qual. Res. 15 (1) (2013) 57–70, https://doi.org/10.1177/1468794113504103.
[63] V. Braun, V. Clarke, Thematic Analysis: a Practical Guide, SAGE Publications Ltd, London, 2022.
[64] K. Malterud, V.D. Siersma, A.D. Guassora, Sample size in qualitative Interview studies: guided by information power, Qual. Health Res. 26 (13) (2016) 1753–1760, https://doi.org/10.1177/1049732315617444.
[65] V. Braun, V. Clarke, To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales, Qualitative research in sport, exercise and health 13 (2) (2021) 201–216, https://doi.org/10.1080/2159676X.2019.1704846.
[66] A. Cherubini, R. Passama, A. Crosnier, A. Lasnier, P. Fraisse, Collaborative manufacturing with physical human–robot interaction, Robot. Comput. Integrated Manuf. 40 (2016) 1–13, https://doi.org/10.1016/j.rcim.2015.12.007.

[67] E. Matheson, R. Minto, E.G.G. Zampieri, M. Faccio, G. Rosati, Human–robot collaboration in manufacturing applications: a review, Robotics 8 (4) (2019), https://doi.org/10.3390/robotics8040100.
[68] A.C. Simões, A. Pinto, J. Santos, S. Pinheiro, D. Romero, Designing human-robot collaboration (HRC) workspaces in industrial settings: a systematic literature review, J. Manuf. Syst. 62 (2022) 28–43, https://doi.org/10.1016/j.jmsy.2021.11.007.
[69] A. Ranavolo, G. Chini, F. Draicchio, A. Silvetti, T. Varrecchia, L. Fiori, A. Ajoudani, Human-robot collaboration (HRC) technologies for reducing work-related musculoskeletal Diseases in industry 4.0, 223, Association (IEA, 2021, pp. 335–342.
[70] A. Grybauskas, A. Stefanini, M. Ghobakhloo, Social sustainability in the age of digitalization: a systematic literature Review on the social implications of industry 4.0, Technol. Soc. 70 (2022) 101997, https://doi.org/10.1016/j.techsoc.2022.101997.
[71] W.D. Holford, The future of human creative knowledge work within the digital economy, Futures 105 (2019) 143–154, https://doi.org/10.1016/j.futures.2018.10.002.
[72] A. van Wynsberghe, M. Ley, S. Roeser, Ethical aspects of human–robot collaboration in industrial work settings, in: M.I. Aldinhas Ferreira, S.R. Fletcher (Eds.), The 21st Century Industrial Robot: when Tools Become Collaborators, Springer International Publishing, Cham, 2022, pp. 255–266.
[73] N. Leesakul, A.-M. Oostveen, I. Eimontaite, M.L. Wilson, R. Hyde, Workplace 4.0: exploring the implications of technology adoption in digital manufacturing on a sustainable workforce, Sustainability 14 (6) (2022), https://doi.org/10.3390/su14063311.
[74] P. Leitão, C.A.S. Geraldes, F.P. Fernandes, H. Badikyan, Analysis of the workforce skills for the factories of the future, in: Paper presented at the 2020 IEEE Conference on Industrial Cyberphysical Systems (ICPS) (2020, 10-12 June), 2020, pp. 353–358. https://ieeexplore.ieee.org/document/9274757.
[75] L. Schlogl, E. Weiss, B. Prainsack, Constructing the 'Future of Work': an analysis of the policy discourse, New Technol. Work. Employ. 36 (3) (2021) 307–326, https://doi.org/10.1111/ntwe.12202.
[76] M. Faccio, I. Granata, A. Menini, M. Milanese, C. Rossato, M. Bottin, G. Rosati, Human factors in cobot era: a review of modern production systems features, J. Intell. Manuf. 34 (1) (2023) 85–106, https://doi.org/10.1007/s10845-022-01953-w.
[77] R. Vecellio Segate, A. Daly, 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, Eur. J. Risk Regul. (2023) 1–40, https://doi.org/10.1017/err.2023.72.
[78] R. Vecellio Segate, Litigating trade secrets in China: an imminent pivot to cybersecurity? J. Intellect. Property Law Pract. 15 (8) (2020) 649–659, https://doi.org/10.1093/jiplp/jpaa092.
[79] A. Grigorescu, The Ebb and Flow of Global Governance: Intergovernmentalism versus Nongovernmentalism in World Politics, Cambridge University Press, Cambridge, 2020.
[80] T. Liang, H. Mao, Y. Tan, B. Sun, How does socio-technical landscape affect the formation of technical standards? Sage Open 13 (4) (2023) 21582440231216009 https://doi.org/10.1177/21582440231216009.
[81] J.-M. Bello y Villarino, Global standard-setting for artificial intelligence: Para-regulating international law for AI? The Australian Year Book of International Law Online 41 (1) (2023) 157–181, https://doi.org/10.1163/26660229-04101018.
[82] A. Cardoso, A. Colim, E. Bicho, A.C. Braga, M. Menozzi, P. Arezes, Ergonomics and human factors as a requirement to implement safer collaborative robotic torkstations: a literature review, Saf. Now. 7 (4) (2021), https://doi.org/10.3390/safety7040071.
[83] P. Chemweno, L. Pintelon, W. Decre, Orienting safety assurance with outcomes of hazard analysis and risk assessment: a review of the ISO 15066 standard for collaborative robot systems, Saf. Sci. 129 (2020) 104832, https://doi.org/10.1016/j.ssci.2020.104832.
[84] A. Zacharaki, I. Kostavelis, A. Gasteratos, I. Dokas, Safety bounds in human robot interaction: a survey, Saf. Sci. 127 (2020) 104667, https://doi.org/10.1016/j.ssci.2020.104667.
[85] J.F. Arinez, Q. Chang, R.X. Gao, C. Xu, J. Zhang, Artificial intelligence in advanced manufacturing: current status and future tutlook, J. Manuf. Sci. Eng. 142 (11) (2020), https://doi.org/10.1115/1.4047855.
[86] Human Impact Partners, The Public Health Crisis Hidden in Amazon Warehouses, Retrieved from, Oakland, CA, 2021.
[87] J. Nelson, S. Buisine, A. Aoussat, Anticipating the use of future things: towards a framework for prospective use analysis in innovation design projects, Appl. Ergon. 44 (6) (2013) 948–956, https://doi.org/10.1016/j.apergo.2013.01.002.
[88] J. Nelson, S. Buisine, A. Aoussat, C. Gazo, Generating prospective scenarios of use in innovation projects, Trav. Hum. Le. 77 (1) (2014) 21–38, https://doi.org/10.3917/th.771.0021.
[89] M. Sinclair, Ergonomics issues in future systems, Ergonomics 50 (12) (2007) 1957–1986, https://doi.org/10.1080/00140130701639474.
[90] A.S. Oliveira, M.D. de Barros, F. de Carvalho Pereira, C.F.S. Gomes, H.G. da Costa, Prospective scenarios: a literature review on the Scopus database, Futures 100 (2018) 20–33, https://doi.org/10.1016/j.futures.2018.03.005.
[91] V. Braun, V. Clarke, One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual. Res. Psychol. 18 (3) (2021) 328–352, https://doi.org/10.1080/14780887.2020.1769238.
[92] V. Braun, V. Clarke, E. Boulton, L. Davey, C. McEvoy, The online survey as a qualitative research tool, Int. J. Soc. Res. Methodol. 24 (6) (2021) 641–654, https://doi.org/10.1080/13645579.2020.1805550.
[93] M. Schreier, Qualitative Content Analysis in Practice, SAGE Publications, London, 2012.
[94] S. Darwall, The value of autonomy and autonomy of the will, Ethics 116 (2) (2006) 263–284.
[95] R.M. Ryan, E.L. Deci, Self-determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness, Guilford publications, 2017.
[96] R.M. Ryan, E.L. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, Am. Psychol. 55 (1) (2000) 68.
[97] T. Gamer, M. Hoernicke, B. Kloepper, R. Bauer, A.J. Isaksson, The autonomous industrial plant – future of process engineering, operations and maintenance, J. Process Control 88 (2020) 101–110, https://doi.org/10.1016/j.jprocont.2020.01.012.
[98] EC, Ethics Guidelines for Trustworthy AI, Retrieved from Directorate-General for Communications Networks, Content Technology, 2019.
[99] M. Gartner, Regulatory Acknowledgment of individual autonomy in European digital legislation: from eeta-principle to explicit protection in the data act, European Data Protection Law Review 8 (4) (2022) 462–473, https://doi.org/10.21552/edpl/2022/4/6.
[100] A. Bertolini, Artificial Intelligence and Civil Liability, Retrieved from, Brussels, 2020.
[101] EC, Artificial Intelligence, Robotics and 'Autonomous' Systems, 2018. Retrieved from Brussels.
[102] N. Nevejans, European Civil Law Rules in Robotics, 2016. Retrieved from Brussels.
[103] H. Matsuzaki, G. Lindemann, The autonomy-safety-paradox of service robotics in Europe and Japan: a comparative analysis, AI Soc. 31 (4) (2016) 501–517, https://doi.org/10.1007/s00146-015-0630-7.
[104] Z. Kemény, J. Váncza, L. Wang, X.V. Wang, Human–robot collaboration in manufacturing: a multi-agent view, in: L. Wang, X.V. Wang, J. Váncza, Z. Kemény (Eds.), Advanced Human-Robot Collaboration in Manufacturing, Springer International Publishing, Cham, 2021, pp. 3–41.
[105] A. Kolbeinsson, E. Lagerstedt, J. Lindblom, Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing, Production & Manufacturing Research 7 (1) (2019) 448–471, https://doi.org/10.1080/21693277.2019.1645628.
[106] M. Karakikes, D. Nathanael, The effect of cognitive workload on decision authority assignment in human–robot collaboration, Cognit. Technol. Work 25 (1) (2023) 31–43, https://doi.org/10.1007/s10111-022-00719-x.
[107] A.M. Grant, Y. Fried, T. Juillerat, Work matters: job design in classic and contemporary perspectives, in: APA Handbook of Industrial and Organizational Psychology vol. 1, American Psychological Association, Washington, DC, US, 2011, pp. 417–453. *Building and developing the organization*.
[108] R.J. Torraco, Work design theory: a review and critique with implications for human resource development, Hum. Resour. Dev. Q. 16 (1) (2005) 85–109, https://doi.org/10.1002/hrdq.1125.
[109] S. Parker, Beyond motivation: job and work design for development, health, ambidexterity, and more, Annu. Rev. Psychol. 65 (2014) 661–691. Retrieved from, https://www.annualreviews.org/doi/pdf/10.1146/annurev-psych-010213-115208.
[110] S. Parker, T.D. Wall, Job and Work Design: Organizing Work to Promote Well-Being and Effectiveness, vol. 4, Sage, 1998.
[111] J. Smids, H. Berkers, P. Le Blanc, S. Rispens, S. Nyholm, Employers have a duty of aeneficence to design for meaningful work: a general argument and logistics warehouses as a case study, J. Ethics (2023), https://doi.org/10.1007/s10892-023-09442-9.
[112] A. Pollak, M. Paliga, B. Kożusznik, The Impact of New Technologies on Work Design – Case Study of the Industrial Robot Controllers from One Organization, in: Paper presented at theIEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA) (14-22 May), 2021, pp. 156–160, doi:10.1109/CogSIMA51574.2021.9475924.
[113] M.A. Daniels, G.J. Greguras, Exploring the nature of power distance: implications for micro- and macro-level gheories, processes, and outcomes, J. Manag. 40 (5) (2014) 1202–1229, https://doi.org/10.1177/0149206314527131.
[114] E. Kaasinen, F. Schmalfuß, C. Özturk, S. Aromaa, M. Boubekeur, J. Heilala, T. Walter, Empowering and engaging industrial workers with Operator 4.0 solutions, Comput. Ind. Eng. 139 (2020) 105678, https://doi.org/10.1016/j.cie.2019.01.052.
[115] G. Brown, J.L. Pierce, C. Crossley, Toward an understanding of the development of ownership feelings, J. Organ. Behav. 35 (3) (2014) 318–338, https://doi.org/10.1002/job.1869.
[116] K.A. Shogren, T.D. Little, M.L. Wehmeyer, Human agentic Theories and the development of self-determination, in: M.L. Wehmeyer, K.A. Shogren, T.D. Little, S.J. Lopez (Eds.), Development of Self-Determination through the Life-Course, Springer Netherlands, Dordrecht, 2017, pp. 17–26.
[117] A. Bandura, Exercise of human agency through collective efficacy, Curr. Dir. Psychol. Sci. 9 (3) (2000) 75–78.
[118] A. Bandura, Social cognitive theory: an agentic perspective, Annu. Rev. Psychol. 52 (1) (2001) 1–26.
[119] A. Bandura, Toward a psychology of human agency, Perspect. Psychol. Sci. 1 (2) (2006) 164–180, https://doi.org/10.1111/j.1745-6916.2006.00011.x.
[120] M. Akiyoshi, Trust in things: a review of social science perspectives on autonomous human-machine-team systems and systemic interdependence,

Frontiers in Physics 10 (2022). Retrieved from, https://www.frontiersin.org/articles/10.3389/fphy.2022.951296.
[121] M. Lewis, K. Sycara, P. Walker, The role of trust in human-robot interaction, in: H. A. Abbass, J. Scholz, D.J. Reid (Eds.), Foundations of Trusted Autonomy, vol. 117, Springer International Publishing, 2018, pp. 135–159.
[122] A. Ezenyilimba, M. Wong, A. Hehr, M. Demir, A. Wolff, E. Chiou, N. Cooke, Impact of transparency and Explanations on trust and situation awareness in human–robot teams, J. Cogn. Eng. Decis. Mak. 17 (1) (2022) 75–93, https://doi.org/10.1177/15553434221136358.
[123] G. Charalambous, S.R. Fletcher, P. Webb, The development of a scale to evaluate trust in industrial human-robot collaboration, International Journal of Social Robotics 8 (2016) 193–209.
[124] C.A. Miller (Ed.), Trust in Human-Robot Interaction, Academic Press, 2021.
[125] T.L. Sanders, T. Wixon, K.E. Schafer, J.Y.C. Chen, P.A. Hancock, The influence of modality and transparency on trust in human-robot interaction, in: Paper presented at the 2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA) (3-6 March), 2014, pp. 156–159, doi:10.1109/CogSIMA.2014.6816556.
[126] C.D. Frith, Action, agency and responsibility, Neuropsychologia 55 (2014) 137–142, https://doi.org/10.1016/j.neuropsychologia.2013.09.007.
[127] A. Matthias, The responsibility gap: uscribing responsibility for the actions of learning automata, Ethics Inf. Technol. 6 (3) (2004) 175–183, https://doi.org/10.1007/s10676-004-3422-1.
[128] M.C. Gombolay, R.A. Gutierrez, S.G. Clarke, G.F. Sturla, J.A. Shah, Decision-making authority, team efficiency and human worker satisfaction in mixed human-robot teams, Aut. Robots 39 (3) (2015) 293–312, https://doi.org/10.1007/s10514-015-9457-9.
[129] E. Fosch-Villaronga, Robots, Healthcare, and the Law: Regulating Automation in Personal Care, Routledge, 2020.
[130] R. Vecellio Segate, Shifting privacy rights from the individual to the group: a Re-adaptation of algorithms regulation to address the Gestaltian configuration of groups, Loyola University Chicago J. Regul. Compliance 8 (1) (2022) 55–114. https://www.compliancelawjournal.com/compliancelawjournal/issue_viii/MobilePagedArticle.action?articleId=1793260#articleId1793260.
[131] U. Othman, E. Yang, Human–robot collaborations in smart manufacturing environments: review and Outlook, Sensors 23 (12) (2023), https://doi.org/10.3390/s23125663.
[132] N.M. Richards, W.D. Smart, How should the law think about robots? in: R. Calo, A.M. Froomkin, I. Kerr (Eds.), Robot Law Elgar, 2016, pp. 3–22.
[133] N. Berx, A. Brescia, R. Aqamarina, E.M. Curcio, L. Pintelon, G. Carbone, Stakeholders' perspectives on safety-related human–robot collaborative scenarios, Int. J. Adv. Rob. Syst. 20 (5) (2023) 17298806231200095, https://doi.org/10.1177/17298806231200095.
[134] J. Loh, Responsibility and robot ethics: a critical overview, Philosophies 4 (4) (2019), https://doi.org/10.3390/philosophies4040058.
[135] F. Santoni de Sio, G. Mecacci, Four responsibility gaps with artificial intelligence: why they matter and how to address them, Philosophy & Technology 34 (4) (2021) 1057–1084, https://doi.org/10.1007/s13347-021-00450-x.
[136] P. Hacker, The European AI liability directives – critique of a half-hearted approach and lessons for the future, Computer Law & Security Review 51 (2023) 105871, https://doi.org/10.1016/j.clsr.2023.105871.
[137] C. Brassart Olsen, To track or not to track? Employees' data privacy in the age of corporate wellness, mobile health, and GDPR†, International Data Privacy Law 10 (3) (2020) 236–252, https://doi.org/10.1093/idpl/ipaa004.
[138] M. Otto, Workforce analytics v fundamental rights protection in the EU in the age of big data, Comp. Labor Law Pol. J. 40 (40) (2018) 389–404.
[139] R.R. Fletcher, A. Nakeshimana, O. Olubeko, Addressing fairness, bias, and appropriate Use of artificial intelligence and machine learning in global health, Front. Artif. Intell. 3 (2021) 561802, https://doi.org/10.3389/frai.2020.561802.
[140] N. Akalin, A. Kiselev, A. Kristoffersson, A. Loutfi, A Taxonomy of factors influencing perceived safety in human–robot interaction, International Journal of Social Robotics 15 (12) (2023) 1993–2004, https://doi.org/10.1007/s12369-023-01027-8.
[141] ISO, 9241-220: Ergonomics of Human-System Interaction. Processes for Enabling, Executing and Assessing Human-Centred Design within Organizations, 2019.
[142] A. Mohammed, B. Schmidt, L. Wang, Active collision avoidance for human–robot collaboration driven by vision sensors, Int. J. Comput. Integrated Manuf. 30 (9) (2017) 970–980, https://doi.org/10.1080/0951192X.2016.1268269.
[143] N. Pedrocchi, F. Vicentini, M. Matteo, L.M. Tosatti, Safe human-robot cooperation in an industrial environment, Int. J. Adv. Rob. Syst. 10 (1) (2013) 27, https://doi.org/10.5772/53939.
[144] A. Tausch, C. Peifer, B.M. Kirchhoff, A. Kluge, Human–robot interaction: how worker influence in task allocation improves autonomy, Ergonomics 65 (9) (2022) 1230–1244, https://doi.org/10.1080/00140139.2022.2025912.
[145] N. Berx, W. Decré, I. Morag, P. Chemweno, L. Pintelon, Identification and classification of risk factors for human-robot collaboration from a system-wide perspective, Comput. Ind. Eng. 163 (2022) 107827, https://doi.org/10.1016/j.cie.2021.107827.
[146] M.J. Tear, T.W. Reader, S. Shorrock, B. Kirwan, Safety culture and power: interactions between perceptions of safety culture, organisational hierarchy, and national culture, Saf. Sci. 121 (2020) 550–561, https://doi.org/10.1016/j.ssci.2018.10.014.
[147] E.A. Okine, E. Zarei, B.J. Roggow, Exploring the intellectual insights in aviation safety research: a systematic literature and bibliometric review, Saf. Sci. 170 (2024) 106354, https://doi.org/10.1016/j.ssci.2023.106354.
[148] S.W.A. Dekker, Just Culture: Balancing Safety and Accountability, second ed., CRC Press, 2012.
[149] EUROCONTROL, Just culture. http://www.eurocontrol.int/articles/just-culture, 2016.
[150] Y. Luo, S.A. Zahra, Industry 4.0 in international business research, J. Int. Bus. Stud. 54 (3) (2023) 403–417, https://doi.org/10.1057/s41267-022-00577-9.
[151] O. Kolade, A. Owoseni, Employment 5.0: the work of the future and the future of work, Technol. Soc. 71 (2022) 102086, https://doi.org/10.1016/j.techsoc.2022.102086.
[152] J.-C. Le Coze, Safety Science Research: Evolution, Challenges and New Directions, CRC Press, 2019.
[153] S.W.A. Dekker, Just culture: who gets to draw the line? Cognit. Technol. Work 11 (3) (2009) 177–185, https://doi.org/10.1007/s10111-008-0110-7.
[154] D.J. Provan, D.D. Woods, S.W.A. Dekker, A.J. Rae, Safety II professionals: how resilience engineering can transform safety practice, Reliab. Eng. Syst. Saf. 195 (2020) 106740, https://doi.org/10.1016/j.ress.2019.106740.
[155] M. Ciccarelli, A. Papetti, M. Germani, Exploring how new industrial paradigms affect the workforce: a literature review of Operator 4.0, J. Manuf. Syst. 70 (2023) 464–483, https://doi.org/10.1016/j.jmsy.2023.08.016.
[156] S.W.A. Dekker, J.M. Nyce, There is safety in power, or power in safety, Saf. Sci. 67 (2014) 44–49, https://doi.org/10.1016/j.ssci.2013.10.013.
[157] D. Romero, J. Stahre, M. Taisch, The Operator 4.0: towards socially sustainable factories of the future, Comput. Ind. Eng. 139 (2020) 106128, https://doi.org/10.1016/j.cie.2019.106128.
[158] G. Beier, M. Matthess, L. Shuttleworth, T. Guan, D.I. de Oliveira Pereira Grudzien, B. Xue, L. Chen, Implications of Industry 4.0 on industrial employment: a comparative survey from Brazilian, Chinese, and German practitioners, Technol. Soc. 70 (2022) 102028, https://doi.org/10.1016/j.techsoc.2022.102028.
[159] T. Haipeter, Digitalisation, unions and participation: the German case of 'industry 4.0', Ind. Relat. J. 51 (3) (2020) 242–260, https://doi.org/10.1111/irj.12291.
[160] M. Wilkesmann, U. Wilkesmann, Industry 4.0 – organizing routines or innovations? VINE Journal of Information and Knowledge Management Systems 48 (2) (2018) 238–254, https://doi.org/10.1108/VJIKMS-04-2017-0019.
[161] D.G. Schniederjans, C. Curado, M. Khalajhedayati, Supply chain digitisation trends: an integration of knowledge management, Int. J. Prod. Econ. 220 (2020) 107439, https://doi.org/10.1016/j.ijpe.2019.07.012.
[162] B. Johansson, S. Engberg, I. Lindgren, T. Sordal, Impacts of RPA on employees skill variety, in: Paper presented at the 26th Americas Conference on Information-Systems (AMCIS 2020) (August 15-17), 2020. https://aisel.aisnet.org/amcis2020/cognitive_in_is/cognitive_in_is/9/.
[163] P.H. RosenS, Wischniewski, Task design in human-robot-interaction scenarios – challenges from a human factors perspective, in: Paper presented at the Advances in Human Factors and Systems Interaction, ChamPaper presented at the AHFE 2017 Conference - Advances in Human Factors and Systems Interaction Vol. 2018592, 2018, pp. 71–82. https://doi.org/10.1007/9%2078-3-319-60366-7_8.
[164] V.H. Agisilaou, H. Harris, Learning to be ethical: the role of ethical capability in community development education, Community Dev. J. 58 (1) (2023) 154–172, https://doi.org/10.1093/cdj/bsac041.