

POLITECNICO DI TORINO

Master's Degree in Engineering and Management



**Politecnico
di Torino**

MASTER's Degree Thesis

AI-Assisted Assembly Systems for Inclusive Manufacturing: A Comparative Techno-Legal Analysis of Data, Cognitive Support, and Governance

Supervisor

Prof. Alessandro SIMEONE

Co-Supervisor

Prof. Maria Samantha ESPOSITO

Candidate

Andrea SCARPINO

333245

July 2026

Abstract

This thesis explores AI-assisted assembly systems for inclusive manufacturing through a comparative techno-legal analysis of data, cognitive support, and governance. Within the Industry 5.0 paradigm, manufacturing systems are increasingly expected not only to improve productivity and quality but also to promote human well-being, accommodate workforce diversity, and support more sustainable forms of human-machine collaboration. The research compares three AI-assisted assembly configurations selected according to two analytical variables: the type of data processed and the type of AI employed. The first system combines non-physiological task-related data with generative AI to provide accessible corrective instructions. The second relies on physiological and behavioural data with non-generative AI to estimate cognitive load and support reciprocal learning in human-robot collaboration. The third combines physiological and multimodal data with generative AI to infer human intentions and support robot execution. Each system is reconstructed through a common analytical framework. The comparison shows that AI-assisted assembly systems can contribute to more inclusive and human-centred manufacturing by reducing cognitive friction, supporting error recovery, and improving interaction with the technology. At the same time, as systems become more adaptive, multimodal, and personalised, they also require more intensive data processing, more complex architectures, and stronger governance safeguards. This increases not only technical validation requirements, but also legal and ethical exposure. The thesis argues that the long-term legitimacy of these systems depends on adopting the least intrusive technical configuration capable of delivering a demonstrable assistive benefit. Accordingly, GDPR and AI Act requirements should be treated as design drivers shaping data minimisation, purpose limitation, transparency, retention, human oversight, and impact assessment. The thesis, therefore, contributes a structured framework for evaluating when AI-assisted assembly systems are technically useful, proportionate, governable, and consistent with the objectives of inclusive Industry 5.0.

Acknowledgements

“Arriving at one goal is the starting point to another.”
John Dewey

This thesis represents not only the conclusion of an academic journey, but also the starting point of a new chapter. It is the result of years of dedication, sacrifices, and perseverance, made possible by the guidance, support, and encouragement of those who accompanied me along the way.

Table of Contents

List of Tables	IX
List of Figures	X
Nomenclature	XIII
1 Chapter 1	
Introduction	1
1.1 From Industry 4.0 to Industry 5.0	1
1.2 Human-centred and inclusive manufacturing	2
1.3 Inclusion and cognitive diversity in manufacturing environments . .	4
1.4 Assembly as cognitive work: cognitive load, assembly instructions, and human-technology interactions	4
1.5 The technological opportunity: AI for cognitive and decision support in assembly	5
1.6 Assistance versus workplace surveillance and algorithmic management	7
1.7 The importance of the GDPR and the AI Act as design drivers . . .	8
1.8 Thesis aim and objectives	10
2 Chapter 2	
State of the art	12
2.1 Inclusive manufacturing and neurodiversity in industrial work . . .	12
2.2 Human-machine collaboration and autonomy levels	15
2.3 Cognitive load and workload in assembly environments: concepts and estimation methods	17
2.3.1 Physiological proxies	18
2.3.2 Behavioural and performance proxies	19
2.4 Assistance systems in assembly: modalities, AI-enabled perception, robotics, and GenAI	20
2.4.1 Digital work instructions and multimodal guidance	20
2.4.2 Error detection and procedural mistake analysis	21
2.4.3 Human-technology collaboration with cobot-based assistance	21

2.4.4	Operator profiling and personalised assistance	22
2.4.5	GenAI as instruction generator and decision-support component	22
2.5	Recurrent failure modes in AI-assisted assembly and inclusive support systems	24
2.5.1	False positives, false negatives, and error cascades	25
2.5.2	Environmental variability and model degradation	25
2.5.3	Alert fatigue, alarm overload and over-reliance	26
2.5.4	Explainability and the limits of transparency	26
2.6	Approaches and designs for assembly assistance and inclusive manufacturing systems	27
2.6.1	Performance metrics: time, errors, and recovery	28
2.6.2	Workload and CLT: NASA-TLX and beyond	28
2.6.3	Acceptance, trust, and organisational fit	29
2.6.4	Inclusive evaluation across user diversity and long-term effects	29
2.7	Research gaps and thesis positioning	30
3	Chapter 3	
	Data and data governance in AI-assisted production systems	32
3.1	The role of data in AI-assisted systems	32
3.2	Types of data involved in industrial assistive and manufacturing systems	33
3.3	Personal data, special categories of data, inferred data, and behavioural data	34
3.4	Aggregated, anonymised, identifiable, and pseudonymised data	36
3.5	Core data protection principles	37
3.6	Data protection by design and by default	39
3.7	DPIA as an ex-ante design and governance tool	41
3.8	Data governance issues for workplace AI	42
3.9	Chapter synthesis	45
4	Chapter 4	
	Research design and analytical framework	47
4.1	Thesis rationale and comparative logic of case selection	47
4.2	The comparative matrix: data type and AI paradigm	48
4.3	Structured case-study methodology	52
4.4	Engineering analysis dimensions	53
4.5	Information-flow and decision-flow mapping	53
4.6	Comparative analysis criteria	54
4.7	Scope, assumptions, and limitations	55

5 Chapter 5

System-level case-study analysis	56
5.1 System A: non-physiological data and GenAI	56
5.1.1 System purpose and target users	56
5.1.2 System architecture	57
5.1.3 Informational and decision flow	58
5.1.4 Reported performance and evaluation evidence	60
5.1.5 Robustness and technical limitations	62
5.1.6 Human-machine interface and interaction design	64
5.1.7 Technical complexity and industrial implementability	64
5.1.8 Data and governance profile	66
5.1.9 Preliminary assessment	67
5.2 System B: physiological data and non-GenAI	67
5.2.1 System purpose and target users	67
5.2.2 System architecture	68
5.2.3 Informational and decision flow	70
5.2.4 Reported performance and evaluation evidence	72
5.2.5 Robustness and technical limitations	73
5.2.6 Human-machine interface and interaction design	75
5.2.7 Technical complexity and industrial implementability	76
5.2.8 Data and governance profile	78
5.2.9 Preliminary assessment	79
5.3 System C: physiological data and GenAI	80
5.3.1 System purpose and target users	80
5.3.2 System architecture	80
5.3.3 Informational and decision flow	82
5.3.4 Reported performance and evaluation evidence	83
5.3.5 Robustness and technical limitations	85
5.3.6 Human-machine interface and interaction design	86
5.3.7 Technical complexity and industrial implementability	88
5.3.8 Data and governance profile	89
5.3.9 Preliminary assessment	90

6 Chapter 6

Comparative engineering analysis	91
6.1 Comparative architecture patterns	91
6.2 Comparative sensing configurations and data intensity	93
6.3 Comparative information and decision logic	96
6.4 Comparative performance evidence	98
6.5 Comparative robustness and failure modes	100
6.6 Comparative human-robot interface models	102

6.7	Comparative technical complexity and industrial implementability	104
6.8	Engineering synthesis: main trade-offs across the three systems	106
7	Chapter 7	
	Legal and governance comparison	109
7.1	Differentiating the legal profiles of the three configurations	109
7.2	Risks arising from personal, inferred, and special category data	112
7.3	Purpose limitation, data minimisation, and storage issues	115
7.4	External services, APIs, and cross-border data-transfer implications	117
7.5	Transparency, human oversight, and contestability requirements	119
7.6	Function creep, workplace monitoring, and proportionality safeguards	121
7.7	GDPR and AI Act implications across the three systems	124
7.8	Design and governance requirements derived from the comparison	129
8	Chapter 8	
	Discussion of findings	134
8.1	Inclusion versus data intrusiveness	134
8.2	Personalisation versus proportionality	135
8.3	GenAI benefits versus reliability constraints	137
8.4	Physiological monitoring benefits versus data sensitivity	139
8.5	Technical usefulness versus legal legitimacy	140
8.6	Main contributions of the thesis	142
9	Chapter 9	
	Conclusions and future research	143
	References	146
A	Appendix	152

List of Tables

2.1	Cognitive demand, typical shop floor barrier and example of supportive modality. Adapted from [3] and [4].	15
2.2	Failure modes related to their typical symptoms and their likely causes. Adapted from [27] and [30].	27
4.1	The comparative matrix.	49
6.1	Comparative sensing architectures and engineering data intensity. .	95
6.2	Comparative strength of performance evidence.	100
7.1	Comparative data and inference risk profile.	114
7.2	Purpose limitation, minimisation, and storage controls by system. .	117
7.3	DPIA focus by system.	123
7.4	Design and governance requirements derived from the comparison. .	131
8.1	Staged implementation logic for proportionate support.	136
8.2	Company checklist for AI-assisted inclusive assembly systems. . . .	141
A.1	Reported evaluation evidence and methodological limitations of System A.	152
A.2	Data and governance profile of System A.	153
A.3	Reported evaluation evidence and methodological limitations of System B.	155
A.4	Data and governance profile of System B.	156
A.5	Reported evaluation evidence and methodological limitations of System C.	159
A.6	Data and governance profile of System C.	160

List of Figures

1.1	Industry 5.0 pillars, adapted from [1].	2
1.2	Human-machine evolution towards human-centric manufacturing, adapted from [2].	3
1.3	Example of methodological framework structured across interconnected layers, integrating hardware, software, digital systems, and human-machine interaction. Adapted from [9].	6
2.1	Operator Generations evolution, adapted from [26].	16
2.2	Taxonomy of cognitive workload assessment methods in industrial settings.	18
2.3	LLMs and GenAI applications in manufacturing, adapted from [35].	24
3.1	Operationalisation of core data protection principles in AI-assisted assembly systems.	40
4.1	Methodological framework adopted for the comparative analysis of the three AI-assisted assembly systems.	52
4.2	Generic information and decision flow in AI-assisted assembly systems.	54
5.1	Reconstructed architecture of System A.	59
5.2	Validated pipeline of the case study of System A.	60
5.3	Time comparison for error scenario E1. Adapted from [10].	62
5.4	Time comparison for error scenario E2. Adapted from [10].	62
5.5	Reconstructed overall architecture of System B.	70
5.6	Validated pipeline of the case study of System B, highlighting the MITL and HITL RL loops.	71
5.7	NMCL and the three scenarios' evidence. Adapted from [8].	74
5.8	Reconstructed overall architecture of System C.	82
5.9	Validated intention-to-action pipeline of the case study of System C, from synchronised multimodal input to robot execution.	84
5.10	Diagram of the described setup. Adapted from [49].	87
6.1	Comparative architecture patterns of the three systems.	93

6.2	Qualitative engineering trade-off map of the three systems.	107
7.1	Conceptual positioning of the three systems according to data sensitivity, inferential intrusiveness, AI complexity, and operational actionability.	111
7.2	GDPR and AI Act assessment flow across the three system configurations.	125
8.1	Progressive support architecture for proportionate implementation.	137

Nomenclature

AGV

Automated Guided Vehicle

AI

Artificial Intelligence

API

Application Programming Interface

AR

Augmented Reality

ASR

Automatic Speech Recognition

CLT

Cognitive Load Theory

CPS

Cyber-Physical Systems

DPIA

Data Protection Impact Assessment

EDPB

European Data Protection Board

EDPS

European Data Protection Supervisor

EEG

Electroencephalogram

EMG

Electromyography

fNIRS

functional Near-Infrared Spectroscopy

FRIA

Fundamental Rights Impact Assessment

GDPR

General Data Protection Regulation

GenAI

Generative Artificial Intelligence

H-CPS

Human Cyber-Physical Systems

HDT

Human Digital Twin

HITL

Human-in-the-loop

HMC

Human-Machine Collaboration

HMI

Human-Machine Interaction

HRC

Human-Robot Collaboration

HRV

Heart Rate Variability

LLM

Large Language Model

MITL

Machine-in-the-loop

ML

Machine Learning

NASA-TLX

NASA Task Load Index

NLP

Natural Language Processing

NMCL

Normalised Memory Cognitive Load

PPE

Personal Protective Equipment

RL

Reciprocal Learning

ROS

Robot Operating System

SVM

Support Vector Machine

T2S

Text-to-speech

TAM

Technology Acceptance Model

UI

User Interface

XAI

Explainable Artificial Intelligence

Chapter 1

Introduction

This chapter introduces the research problem addressed in the thesis and situates it within the transition from Industry 4.0 to Industry 5.0. It explains why AI-assisted assembly systems should be analysed not only as technical tools for improving task execution, but also as socio-technical systems that affect inclusion, cognitive workload, data governance, and workers' rights. The chapter then defines the aim, objectives, and overall positioning of the thesis.

1.1 From Industry 4.0 to Industry 5.0

Over the past decade, CPS, widespread connectivity, and advanced data analytics have significantly changed the manufacturing sector. Industry 4.0 has largely been associated with gains in productivity, product quality, and operational flexibility, all of which remain central to industrial competitiveness.

More recent academic research and policy discussions, however, have questioned whether an efficiency-centred view is sufficient on its own. Manufacturing systems are also expected to cope with disruption, reduce their environmental impact, and remain socially sustainable over time. The effects of technological change therefore extend beyond immediate operational performance and increasingly need to be considered as part of long-term industrial strategy.

Within the European policy debate, Industry 5.0 has emerged as a broader framework for industrial development, extending the focus beyond productivity alone. Rather than replacing Industry 4.0, it adds greater emphasis on human well-being, sustainability, and resilience, as shown in Figure 1.1. The European Commission's 2021 report [1] describes this shift as a move from a shareholder-value model toward a stakeholder-value perspective, in which worker well-being and wider societal goals become more central to industrial development.

These policy objectives also have direct implications for engineering practice. In

production environments, ‘human-centricity’ cannot remain only a general aspiration; it needs to influence concrete design choices and operational requirements. This may involve reducing unnecessary physical and cognitive effort, preserving meaningful human control, supporting worker development, and designing systems that can accommodate a wider range of needs.

These considerations are especially relevant in areas where human expertise remains difficult to replace, including manual and semi-manual assembly, complex rework, and exception handling. Such activities often involve variability, contextual judgement, and situations that are not easily addressed through full automation.

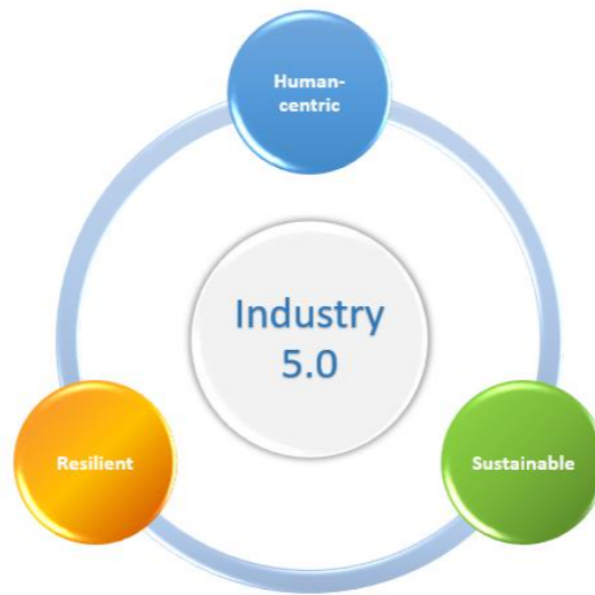


Figure 1.1: Industry 5.0 pillars, adapted from [1].

1.2 Human-centred and inclusive manufacturing

Academic research has increasingly formalised the ‘human-centric’ turn. Lu et al. (2022) [2] note that the concept still lacks a fully shared definition, but the literature converges on the need to support physical and cognitive well-being and to design systems in which humans remain central rather than residual. This perspective aligns with a broader stream of Industry 5.0 scholarship that treats the worker not simply as a component within an optimised production machine but as a stakeholder whose autonomy, safety and development shape both system

design and acceptable deployment practices, as depicted in Figure 1.2. From an engineering perspective, human-centric manufacturing indicates that productivity and quality are not the only optimisation targets: system performance must be evaluated against worker-centric outcomes such as workload, usability, learning curves, trust, and the capacity to safely recover from errors. In other words, in these systems, the operator cannot be viewed only as the end user of a tool. The worker is also an active participant in the task, a subject exposed to monitoring and inference, and a source of practical knowledge that can improve system performance over time.

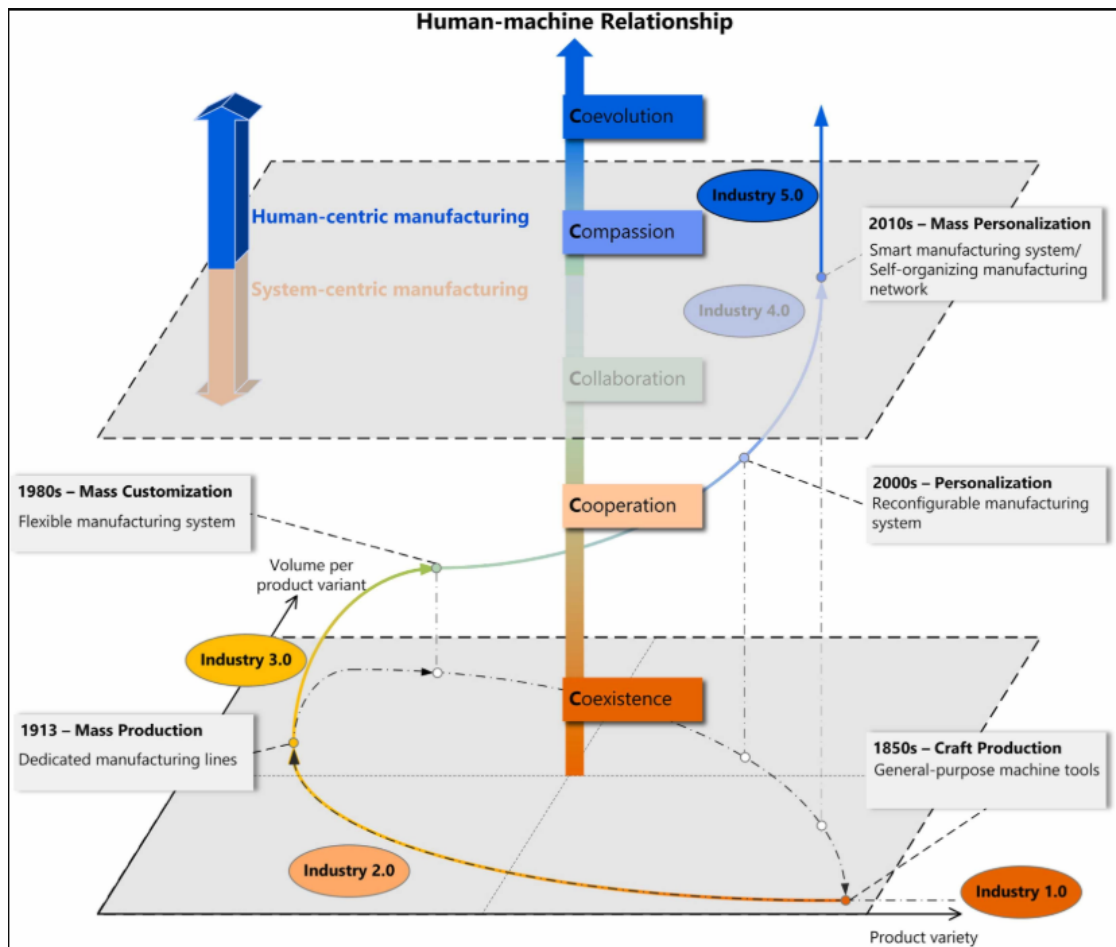


Figure 1.2: Human-machine evolution towards human-centric manufacturing, adapted from [2].

1.3 Inclusion and cognitive diversity in manufacturing environments

Workforce diversity is one of the main reasons why human-centric manufacturing has gained attention. As industrial organisations face demographic change, skill shortages, and pressure to broaden labour participation, inclusion has become both a social responsibility and a practical workforce challenge.

When we talk about inclusion in manufacturing, the focus is often on physical accommodations or special training programmes. But inclusion also presents a systems design challenge because a lot of the work in manufacturing, and especially assembly, requires you to think just as much as what you do with your hands. Operators, for example, have to be able to interpret complex work instructions, keep track of the process cycle, distinguish between similar parts and make quick decisions when things don't go as expected.

Consequently, this cognitive load exerts a disproportionate impact on individuals facing challenges related to reading comprehension, sustained attention, working memory capacity, logical sequencing, or quantitative reasoning. Subsequently, inclusive manufacturing increasingly motivates the development of adaptive instruction modalities and cognitive support mechanisms.

Scientific work on assembly training and instruction design supports this direction. Peltokorpi et al. (2023), for example, discuss manual assembly learning in relation to disability and instruction forms in an industrial experiment, positioning advanced work instructions as a path to enable workers with disabilities in cognitively demanding production tasks [3]. Similarly, Vanneste et al. (2020) investigate cognitive support for assembly operations via different instruction media (oral, paper or AR) in a population where most contributors have cognitive or motor disabilities, illustrating how instruction modality can influence perceived workload and task performance [4]. These studies underscore an important point for present and future development: inclusive design in manufacturing often requires not only 'accessibility' in the narrow sense but also careful management of cognitive demands embedded in task procedures and instruction interfaces.

1.4 Assembly as cognitive work: cognitive load, assembly instructions, and human-technology interactions

Assembly work involves more than physical execution. Operators must interpret information, remember intermediate steps, maintain sequence awareness, and recognise when the process deviates from what is expected. These demands make

CLT relevant to the analysis of assembly support. Originally developed in cognitive science and educational psychology, CLT is based on the idea that human information-processing capacity is limited. Sweller (1988) showed that demanding problem-solving and learning activities can consume cognitive resources that would otherwise support schema acquisition and effective performance [5]. Although the theory was not developed specifically for manufacturing, it offers a useful basis for examining how work instructions and UIs affect operators during assembly tasks. Cognitive workload is assessed in manufacturing and human-factors research through a combination of subjective and objective methods. One of the most widely used subjective tools is the NASA-TLX, which evaluates perceived workload across dimensions including mental, temporal, and physical demand, as well as performance, effort, and frustration [6].

Evidence from realistic work settings also shows that technological assistance does not always reduce workload. In AR-supported maintenance and assembly training, for example, shorter completion times have been observed alongside higher mental workload, particularly during more demanding tasks [7]. This suggests that the effect of assistance depends heavily on how information is presented and how the interface interacts with the task. For inclusive manufacturing, this is an important limitation: adding technology can support performance, but it can also introduce new cognitive demands.

On the contrary, poorly designed assistance can create additional attentional switching, UI complexity, and alert fatigue, thereby undermining the very human-centric goals it aims to serve. This motivates a careful, system-level analysis of assistance technologies and the conditions under which they support rather than burden workers.

1.5 The technological opportunity: AI for cognitive and decision support in assembly

Within this context, AI-based assembly support systems can help operators manage cognitively demanding tasks by monitoring task execution and responding when support is needed.

A typical system may observe the workspace and the progress of the procedure through sensors, compare what is happening with the expected assembly state, and identify possible deviations using tools such as computer vision or rule-based logic. Depending on the application, the response may range from a simple warning or corrective instruction to direct physical assistance in a HRC setting.

The recent emergence of GenAI and LLMs adds a further capability: automatically transforming structured process knowledge and error descriptions into human-facing instructions that can be adaptive, multimodal, and potentially personalised.

Some systems combine computer vision with GenAI to produce stepwise corrective guidance or flowchart-like decision support, and others explore adaptive assistance driven by estimates of cognitive load or by user profiles [8], [9], [10], [11]. As shown in Figure 1.3, the considered system can be read as a layered pipeline from data acquisition to inference and worker-facing interventions (instructions, warnings, or robotic actions), which is useful to locate both the technical failure modes and the stages where legal safeguards should be implemented.

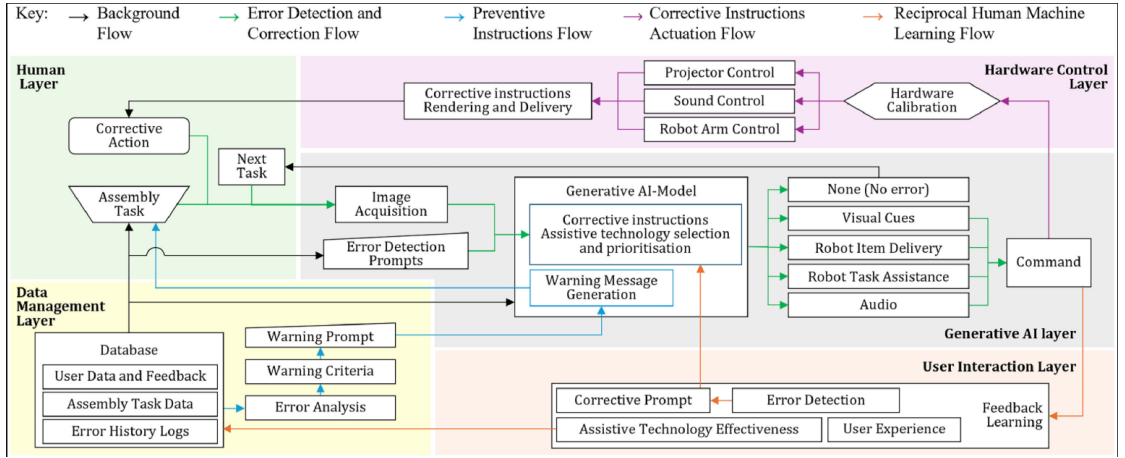


Figure 1.3: Example of methodological framework structured across interconnected layers, integrating hardware, software, digital systems, and human-machine interaction. Adapted from [9].

These prototypes also reflect an important methodological direction, where human-centric systems often incorporate feedback loops. In some approaches, humans correct the system (HITL), allowing models to improve over time, while in others the system provides real-time feedback to the operator (MITL) with the aim of reducing error propagation and shortening recovery time. At the same time, the technical move toward adaptive support increasingly relies on collecting richer data streams like video, audio, detailed logs, and, in some cases, physiological signals used as proxies for cognitive effort. Reviews of cognitive workload assessment in Industry 5.0 settings highlight this growing emphasis on wearable and sensor-based methods, reinforcing that the human-centric turn is associated with expanded sensing of the human at work [12], [13].

1.6 Assistance versus workplace surveillance and algorithmic management

The same sensing, inference, and logging mechanisms that enable cognitive support can also enable forms of workplace surveillance and algorithmic management, i.e., the use of algorithmic systems relying on tracked data to organise, assign, monitor, supervise and evaluate work [14].

This is not merely a hypothetical concern: systematic studies of workplace surveillance research emphasise tensions and contradictions between organisational objectives and employee interests, and they document risks involving privacy invasion, trust erosion, and power asymmetries [15]. Different discussions of employee monitoring further highlight that advanced monitoring technologies challenge existing regulatory and organisational safeguards, and they specifically note the difficulty of complying with data minimisation and transparency principles when digital technologies collect and process large volumes of data.

Manufacturing assistance systems can create significant surveillance risks because the same features that enable support may also make continuous worker monitoring possible. Camera-based sensing can produce detailed records of how work is performed over time. The issue becomes more sensitive when these observations are used to infer a worker's condition. Labels such as 'fatigued', 'high cognitive load', or 'unsuitable for the task' may carry consequences that go well beyond the original sensor data, especially when uncertain model outputs are interpreted as quasi-medical or work-related judgements.

The same concern applies to later reuse. Information initially collected for quality assurance or operator support may become attractive for individual performance assessment. Some prototypes in the selected literature retain operator-linked logs, create user profiles, or compare cognitive indicators with predefined thresholds [8], [9]. Once these practices are embedded in the architecture, the boundary between support and evaluation can become difficult to maintain unless the permitted uses of the data are clearly defined.

This issue goes to the heart of the Industry 5.0 claim to human-centricity. A system cannot be considered genuinely human-centred if technical sophistication comes at the cost of disproportionate monitoring, reduced operator autonomy, or excessive data collection. Privacy, fairness, and the possibility of questioning system outputs should therefore be addressed during design rather than added only after deployment [14], [15], [16], [17].

1.7 The importance of the GDPR and the AI Act as design drivers

Because the issue is intrinsically socio-technical, questions of governance and legal compliance cannot be separated from the design of the technical architecture. Within the European regulatory framework, two legal instruments are especially significant for AI-enabled workplace systems: the General Data Protection Regulation (EU) 2016/679 and the Artificial Intelligence Act, regulation (EU) 2024/1689 [18], [19].

Under the GDPR, the lawfulness and overall legitimacy of many AI-assisted assembly systems depend to a large extent on whether their architectures incorporate the Regulation's core principles and safeguards. Article 5 of the GDPR lays down essential principles, including purpose limitation, data minimisation, storage limitation, integrity and confidentiality [18].

These principles directly affect the design of systems based on continuous sensing and extensive data logging. Article 6 GDPR further requires that any processing activity rest on a valid lawful basis, an issue that becomes particularly delicate in the employment context, where consent is often problematic because of structural power imbalances and where the necessity and proportionality of processing must therefore be rigorously demonstrated [18].

Where the system processes physiological signals or otherwise generates information capable of revealing health-related conditions, Article 9 GDPR assumes particular importance, as it governs the processing of special categories of personal data, including health data, and permits such processing only under limited and specific conditions [18].

Transparency requirements are equally relevant from a design perspective. Articles 13 and 14 GDPR require that data subjects be provided with clear information regarding the purposes of processing, the general logic involved, the recipients of the data, the retention period, and the rights available to them [18]. Compliance with these obligations may prove especially challenging in systems that combine multiple sensors, ML components, and third-party services. Beyond transparency, the GDPR also requires that safeguards be embedded into the system architecture itself. Article 25 introduces the principles of data protection by design and by default, while Article 32 requires the implementation of appropriate technical and organisational security measures [18].

In addition, where monitoring activities or the deployment of novel technologies are likely to give rise to high risks for the rights and freedoms of individuals, Article 35 requires the carrying out of a DPIA. In practice, this requirement is particularly relevant in scenarios involving camera-based surveillance or other forms of systematic monitoring.

Finally, when data is processed by external providers, which is commonly the case for speech recognition, T2S, or some AI services, Article 44 and the GDPR transfer framework become relevant.

For camera-based processing specifically, interpretative guidance is crucial. The EDPB’s Guidelines 3/2019 on processing personal data through video devices [20] provide detailed guidance on how GDPR should be applied to video-based processing, including the importance of purpose specification and the risks of normalising excessive surveillance. This matters for assembly support systems because computer vision is often a core enabler for error detection and step verification.

The AI Act introduces a complementary layer of governance specifically directed at risks and obligations associated with AI [21]. Its risk-based framework requires an assessment of whether the intended use of a given system falls within one of the categories classified as high-risk and, where this is the case, compliance with specific requirements relating to risk management, transparency, human oversight, and robustness. In this respect, Article 6 AI Act, read together with the use cases set out in Annex III, establishes the general mechanism for identifying certain AI systems as high-risk. More specifically, it identifies sectors and intended uses that are presumed to fall within the high-risk regime, including, depending on the purpose pursued, systems used in the context of employment and worker management [19].

If a workplace system is classified as high-risk, the AI Act establishes a set of requirements that can be translated into engineering and organisational controls. Article 9 requires a risk management system across the lifecycle. Article 12 concerns, instead, record-keeping and logging, while Article 13 concerns transparency and information provision [19]. Article 14 requires human oversight, and Article 15 sets expectations for accuracy, robustness, and cybersecurity [19]. On the deployment side, Article 26 defines obligations of deployers, which is particularly relevant in workplaces because organisational measures (i.e. training, governance policies, monitoring of performance in real conditions, and restrictions on how outputs can be used) shape real-world impacts as much as technical properties do.

In certain contexts, Article 27 introduces a fundamental rights impact assessment requirement for deployers, and even where not strictly applicable, it provides a methodological model for structuring rights-orientated analysis in workplace deployments [19], [22].

Taken together, GDPR and the AI Act imply that compliance is not an ‘appendix’ to technical design. Compliance influences fundamental architectural decisions: what data is collected, where it is processed, how long it is retained, who can access it, whether outputs are explainable and contestable, and how the organisation prevents a supportive system from becoming a monitoring infrastructure.

1.8 Thesis aim and objectives

This thesis treats AI-assisted assembly systems as socio-technical systems, where technical performance cannot be considered separately from ethical and legal concerns. The literature reviewed in the previous sections supports this view. Industry 5.0 and inclusive manufacturing research increasingly places human well-being, sustainability, and resilience alongside efficiency, while studies on assembly assistance highlight the potential of multimodal guidance, adaptive instructions, and cognitive support to make demanding tasks easier to manage and to reduce the propagation of errors. Assembly work often places considerable demands on attention, sequencing, and error recovery. In this context, AI-based assistance can help operators manage difficult tasks, respond to mistakes more quickly, and reduce avoidable cognitive effort.

These benefits do not come without trade-offs. More responsive support often depends on continuous sensing, inference about the operator or task state, and detailed event logging. As a result, the system may also increase the extent of workplace monitoring and generate much richer information about individual workers. The main question addressed in this thesis is therefore not whether AI-assisted assembly is simply beneficial or harmful. It is under which technical and governance conditions such systems can provide meaningful support without becoming disproportionate or difficult to justify. A purely engineering-orientated pursuit of efficiency must be reconciled with the protection of the fundamental rights and freedoms of the individuals operating within such systems, and the real challenge lies in contextualising existing legal principles within concrete human-machine interaction scenarios.

These issues also influence how the thesis is structured. Worker profiling and surveillance are not treated as separate labour-law topics, but they cannot be ignored when examining the data architecture of assistive systems. The way data are collected, linked, and reused can affect transparency, trust, and the balance of information between workers and organisations. Even data gathered for a legitimate assistive purpose may become problematic when they are later used in a different context.

For this reason, the analysis considers technical performance, human-centred design, and governance within the same framework rather than as separate areas of assessment.

More specifically, the thesis proceeds from three objectives. First, it seeks to examine AI-assisted assembly support not merely as a set of enabling technologies, but as a socio-technical infrastructure that affects task execution, cognitive workload, usability, and operator experience. Secondly, it seeks to identify the legal and governance implications of these systems from the perspective of data protection and trustworthy AI, paying particular attention to issues of data minimisation,

proportionality, transparency, storage limitation, inferred data, and ex-ante risk assessment. Thirdly, it aims to show that the long-term viability of these systems depends on the adoption of concrete design choices and governance mechanisms capable of embedding safeguards directly into the system architecture, rather than adding them only ex-post as compliance measures.

This last point is particularly important considering the relationship between the GDPR and the AI Act. As Voigt and von dem Bussche (2017) [23] explain, controllers of AI systems must distinguish carefully between data processed during system operation and data used for model development or training. The same authors further observe that the AI Act is structured according to a risk-based approach, complements rather than substitutes the GDPR, and introduces additional obligations relating to transparency, data quality, and the provision of information relevant to DPIAs in the case of high-risk AI systems. For this reason, the present thesis assumes that the balance between technical effectiveness and legal-ethical legitimacy cannot remain confined to the level of abstract principles. Rather, it must be translated into concrete and verifiable safeguards, design constraints, governance procedures, and accountability mechanisms capable of ensuring that assistive functionalities do not evolve into disproportionate or opaque forms of data processing.

Overall, the thesis is situated at the intersection of human-centric manufacturing, cognitive assistance, and legal-regulatory governance. It is motivated by the Industry 5.0 agenda and by the growing relevance of inclusive and cognitively sustainable forms of work design. The thesis also acknowledges that AI-assisted support systems can become difficult to justify when their use of data is excessive, opaque, or poorly governed. Their long-term value therefore depends on more than performance gains. It also depends on whether the underlying architecture can support those gains without compromising legal compliance, ethical acceptability, or sustained worker trust.

Chapter 2

State of the art

This chapter sets out the scientific and technical background for the analysis of AI-assisted assembly systems in inclusive manufacturing. It brings together research on human-centred manufacturing, cognitive workload, digital and multimodal instructions, HRC, GenAI, and recurrent failure modes in assistive systems. The final section then draws out the main gaps in the literature that inform the comparative analysis developed later in the thesis.

2.1 Inclusive manufacturing and neurodiversity in industrial work

Inclusive manufacturing is increasingly discussed within the broader Industry 5.0 debate, especially where human-centricity is interpreted as a commitment to designing production systems around heterogeneous human capabilities rather than around an implicit ‘standard operator’.

In this perspective, inclusion is not only a matter of workforce policy or organisational values. It is also a question of how tasks, interfaces, and support systems are designed. The shift is significant because it moves the discussion away from the idea that workers must simply adapt to production systems as they are, and toward the idea that production systems themselves should be shaped in ways that reduce avoidable barriers and support a wider range of users [2].

This issue is particularly visible in assembly work. Even in technologically advanced manufacturing environments, assembly often remains an area in which humans are crucial because of production variability, small batches, frequent reconfiguration, exception handling, and the contextual judgements required when the process diverges from its nominal path. As a result, inclusion in assembly cannot be understood only in terms of access to employment. It also depends on whether

the work system can reduce unnecessary cognitive friction and provide forms of support that make task execution more manageable for operators with different profiles and needs.

For the purposes of this thesis, inclusion is considered through two interacting dimensions. The first concerns the cognitive demands built into the task itself, such as sequence tracking, rule following, part or tool discrimination, attention switching, language processing, and numerical reasoning. The second concerns the support provided by the work system. Features such as multimodal guidance, adaptive pacing, error prevention, collaborative robotics, and simplified interfaces can substantially change how demanding a task becomes for a given operator. Inclusion therefore depends not only on the task, but also on how the surrounding system is designed to support it.

In an industrial experiment in a sheltered workplace, Peltokorpi et al. (2023) investigated different instruction modalities (including projection-based approaches) for operators with different disability profiles and showed that instruction design can influence early cycle performance and learning [3]. A related exploratory study by Vanneste et al. (2020) compared oral, paper, and AR instructions among assembly operators, the majority of whom had cognitive or motor disabilities; the work demonstrates that instruction modality is not neutral and can shape both performance and workload in ways that matter for inclusion [4].

In parallel, the concept of neurodiversity has gained increasing prominence within both workplace research and policy debates. Recent scholarship has shown that neurodivergent individuals often encounter structural barriers in recruitment, workplace adaptation, and long-term retention, while at the same time underscoring the organisational and social benefits associated with more inclusive practices. At the same time, the literature highlights a persistent gap in connecting the discourse on neurodiversity with concrete human capital management strategies and operational workplace practices [24].

Neurodiversity is especially relevant when assembly work is considered in terms of actual task demands rather than diagnostic categories. The objective is not to describe workers through fixed lists of deficits, but to identify recurring situations in which the design of the task or interface may create avoidable difficulty. This shifts attention towards work systems that can accommodate different needs without requiring increasingly detailed classifications of the individual.

In assembly, the most relevant demands often concern sustained attention, working memory and sequencing, reading and language comprehension, logical reasoning, and numerical processing. These demands become visible in everyday situations such as keeping track of the current step, returning to a task after an interruption, interpreting dense instructions, distinguishing between similar components, or performing simple calculations under time pressure.

Table 2.1 links these cognitive demands to typical points of difficulty in assembly

tasks and to possible forms of support.

- Attention and attentional control are implicated in sustained monitoring of step progression, in switching between physical action and instruction reading, and in detecting intelligent mismatches (e.g., choosing between similar screws). Attention-related difficulties can interact with typical shop floor stressors such as time pressure, noise, interruptions, and multi-tasking.
- Working memory and sequencing are critical for tracking multi-step instructions, remembering tool changes, and maintaining state across partial completions. The risk here is not only forgetting a step, but also the cost of repeatedly re-orienting after interruptions, which is a common phenomenon in real production environments.
- Reading and language comprehension become central when instructions are text-heavy, when part names or codes are similar, or when procedures rely on symbolic notations. Dyslexia is widely characterised as a difficulty in decoding and spelling that can persist into adulthood and affect reading fluency. In assembly contexts, even moderate reductions in reading fluency can translate into slower task pacing and increased cognitive load under time constraints.
- Logic and rule-based reasoning are relevant in tasks that involve conditional decisions, such as ‘if A, then B’, as well as constraints related to component orientation or torque sequences. These skills become especially important when the process departs from its expected path and the operator needs to identify the cause of the problem and decide how to recover.
- Mathematical and numerical reasoning is relevant in measurement, marking, proportioning, and parameter selection tasks (e.g., length measurement, spacing, tolerance checks). Developmental dyscalculia is commonly described as a persistent difficulty in learning numerical and arithmetic concepts: Castaldi (2020) frames it as a learning disability affecting numeric and arithmetic comprehension and highlights heterogeneity and open questions in how dyscalculia relates to individual cognitive profiles too [25].

Ultimately, identifying these dimensions does not imply that industrial systems should attempt to ‘diagnose’ workers. Instead, it provides a functional justification for why instruction design and technological assistance must be highly flexible and adaptable.

Table 2.1: Cognitive demand, typical shop floor barrier and example of supportive modality. Adapted from [3] and [4].

Cognitive demand	Typical manufacturing barriers	Support modalities
Attention and attentional control	Distractions or interruptions; noise; missed cues; low focus on critical steps	Salient minimal visual cues (LED, projection); short audio prompts; step confirmation
Working memory and sequencing	Losing track of multi-step sequences; errors after interruptions; repeated re-checking; wrong order of operations	Chunk steps; progress indicator; just-in-time reminders; automatic step verification; recovery guidance
Reading and language comprehension	Slow or wrong reading; misreading part codes; difficulty parsing dense text under time pressure	Audio-first or multimodal guidance; simplified language; pictorial cues; short, numbered steps; confirm understanding
Logic and rule-based reasoning	Difficulty with conditional rules; error diagnosis; recovery decisions in non-nominal situations	Decision trees; flowcharts; rule-based checks; clear escalation paths; human confirmation
Mathematics and numerical reasoning	Measurement and marking errors; proportional division; parameter selection mistakes; stress under numeric tasks	Automated measurement; visual rulers; voice on-demand ‘math helper’; tolerance-aware checks

2.2 Human-machine collaboration and autonomy levels

Human-centric and inclusive manufacturing is frequently operationalised through HMC and, in many manufacturing cells, through HRC. In this context, collaboration does not simply mean the physical co-presence of humans and machines, but it denotes the distribution of functions across perception, decision-making, and the interactive mechanisms by which humans and machines coordinate in real time.

A widely used conceptual point of reference for describing this form of collaboration is the paradigm of the ‘Operator 4.0’, as represented in Figure 2.1.

Within the relevant literature, this concept is presented as a vision of the smart and skilled worker supported by H-CPS and adaptive automation, with the broader

purpose of fostering a form of human-automation symbiosis consistent with the objectives of social sustainability [26].

This framing is useful because it highlights a range of support models, including augmented, assisted, and collaborative work, instead of reducing the design space to a binary choice between manual work and full automation. It also provides a direct link to Industry 5.0, where automation is treated as a means to support resilient and human-centred production rather than as an end in itself.

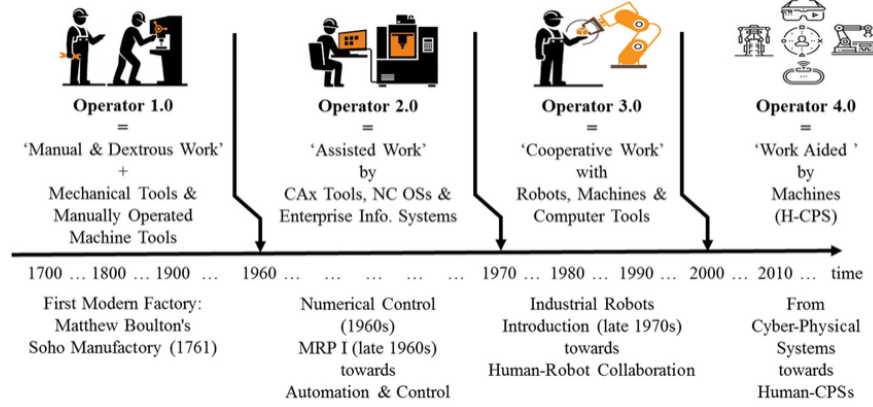


Figure 2.1: Operator Generations evolution, adapted from [26].

From a human factor engineering perspective, one of the central design variables in HMC and HRC concerns the level of system autonomy. A well-established body of literature has shown that automation may be introduced at different stages of information processing, decision-making, and action execution, and that its design should not be understood merely as a substitute for human work [27]. Rather, particular attention must be paid to the role of the human operator in supervision, coordination, and interaction with the automated system.

In this perspective, the levels of automation may also vary, at least to some extent, independently across different processing stages, thereby enabling more nuanced design solutions than those implied by a simple opposition between low and high automation.

This literature is directly applicable to assembly assistance systems, as these technologies inherently embed design choices regarding the system's autonomous perceptual and interpretive capabilities. Specifically, designers must determine whether a system should simply recommend actions or actively enforce them and establish the extent to which the human operator is required to confirm, correct, or override system outputs.

In practical terms, these autonomy levels operate on a spectrum: from passive information presentation (e.g., static digital work instructions) to active detection and warning systems (e.g., flagging procedural deviations), culminating in

semi-autonomous recovery (e.g., proposing corrective steps) and direct physical collaboration (e.g., a cooperative robot handing over parts or performing guided marking) [16]. The broader HRC research landscape increasingly focuses on collaborative assembly and related processes.

Advancements in human-robot collaborative technologies that enhance assembly and welding, frame HRC as a route to combine complementary human and robot strengths, especially in tasks that neither could achieve alone with the same efficiency or flexibility. Consequently, dynamic task allocation remains a central thematic challenge: determining optimal assignments of sub-tasks to the human, the robot, or a shared collaborative space [28].

Ultimately, the quality and inclusivity of this collaboration depend not only on static allocation strategies but also on the system’s capacity for real-time adaptation to the operational context and the cognitive and physical state of the human operator.

2.3 Cognitive load and workload in assembly environments: concepts and estimation methods

A recurring justification for adaptive assistance in assembly is the need to manage cognitive load or mental workload. Although these constructs are not identical across disciplines, they are widely used as design targets in human-centric manufacturing, because they capture the idea that the human operator has limited cognitive resources and that poor system design can consume these resources unnecessarily. From a cognitive science perspective, CLT provides a useful basis for understanding how limited information-processing capacity constrains task execution and learning. Several studies argue that demanding problem-solving activities can consume substantial cognitive resources, leaving less capacity for schema acquisition and effective learning [5]. Although assembly work is not simply a learning task, the same logic is relevant when operators must follow complex instructions or recover from errors under cognitive pressure.

In applied human factors and engineering, workload is often operationalised through instruments and proxies rather than derived from first principles. Among these, the NASA-TLX remains one of the most influential subjective measurement tools. Developed by Hart and Staveland (1988) [6], it provides a robust empirical basis for assessing perceived workload.

More recently, within Industry 5.0-orientated literature, the assessment of cognitive workload is increasingly framed as a complex, multi-method challenge. Recent scholarship on cognitive workload in this emerging industrial context has highlighted the need for integrative assessment approaches capable of combining subjective instruments, such as the NASA-TLX, with objective physiological measurements

[12]. Furthermore, their analysis identifies significant existing gaps in the literature regarding the practical implementation of real-time workload monitoring and the development of robust predictive modelling for adaptive shop floor systems, as it is shown in Figure 2.2, where the five main methodological families (physiological, subjective/self-report, performance-based, cognitive load models, and multimodal approaches) and some representative examples for each category are summarised.

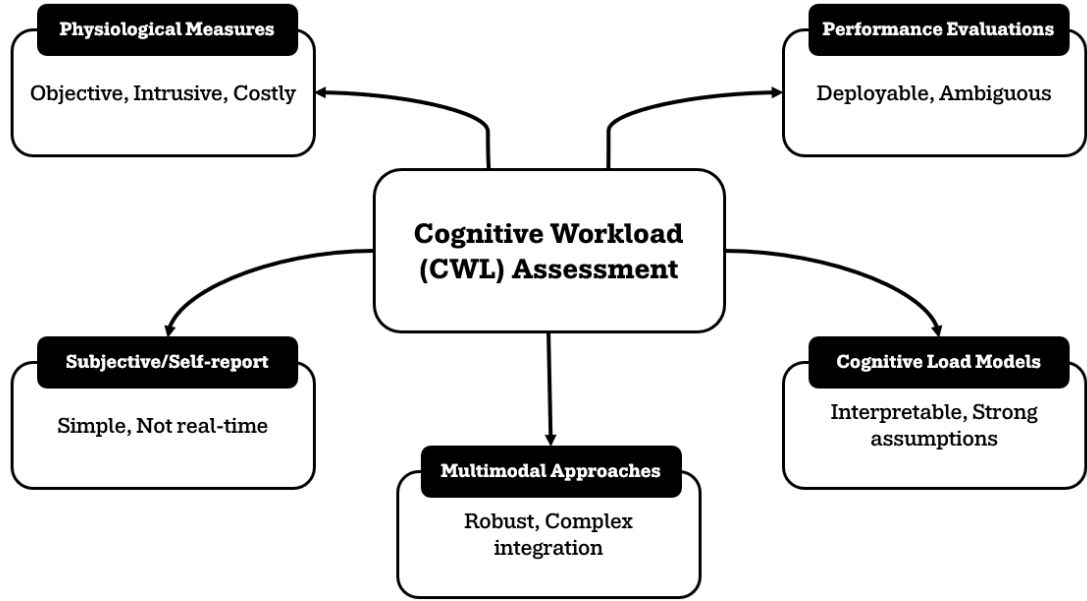


Figure 2.2: Taxonomy of cognitive workload assessment methods in industrial settings.

2.3.1 Physiological proxies

Physiological measures are attractive in this context because they can provide continuous signals that may be related to cognitive load. Commonly examined indicators include cardiovascular measures such as heart rate and HRV, respiration, ocular responses such as blink rate and pupil dilation, and neural signals including EEG and fNIRS. Studies often combine several of these modalities to improve workload estimation.

However, a physiological response associated with cognitive effort in a controlled setting does not automatically provide a reliable measure of workload in manufacturing. Shop-floor conditions introduce additional sources of variation, and the same signal may be influenced by factors unrelated to the task itself.

Physiological signals are influenced by confounders such as physical effort, heat,

stress unrelated to task complexity, and individual differences. Moreover, measurement feasibility is constrained by PPE requirements, operator comfort, and overall technology acceptance.

As a result, even when physiological measures correlate with workload in controlled settings, transferring them to noisy shop floor environments requires robust calibration and careful interpretation, issues repeatedly emphasised in applied workload assessment literature [12].

2.3.2 Behavioural and performance proxies

Compared to physiological sensors, behavioural proxies often present a more highly deployable alternative in industrial settings. Examples include hesitation, repeated checking of previous steps, error frequency, time spent on individual operations, recovery after interruptions, and gaze patterns captured through eye-tracking when such data are available. In some assembly-support prototypes, camera-based observation is also used as one input for estimating the operator's cognitive state [8].

These measures are attractive in industrial settings because they can often be collected with equipment already used for process monitoring or error detection. Their interpretation is less straightforward. A longer completion time, a hesitation, or a higher error rate does not necessarily indicate greater mental effort, since the same behaviour may be influenced by experience, individual strategy, task difficulty, or the surrounding environment.

Task completion time, error count, and rework frequency are likewise common evaluation measures, but they should not be treated as direct indicators of workload. Their values may change for reasons that are only partly related to cognitive demand. A system may, for example, reduce completion time while increasing perceived workload if it introduces frequent attention switching or an intrusive interface. Conversely, a supportive interface may slightly increase task time while reducing mental demand and frustration. This is why many industrial studies combine performance metrics with NASA-TLX and usability measures [6].

Beyond task-specific metrics, workload is also shaped by the surrounding work environment, as we can notice in Figure 2.2. Poor lighting can increase visual fatigue and make errors harder to detect, while high noise levels may interfere with speech-based interaction. Temperature and vibration can also affect operator comfort and sustained attention. These factors are particularly relevant to inclusive system design because they may intensify existing cognitive barriers.

Therefore, recent reviews examining cognitive workload within the Industry 5.0 paradigm emphasise the necessity of integrating both task-related and environmental factors into the analysis, rather than conceptualising workload exclusively as an isolated internal cognitive state [12].

2.4 Assistance systems in assembly: modalities, AI-enabled perception, robotics, and GenAI

Assembly assistance systems can be interpreted as attempts to reduce error likelihood, shorten recovery time, and improve learning under real constraints. In the literature, assistance approaches can be grouped by the layer they primarily target:

- Instruction delivery (modalities through which information is presented to the operator).
- Perception and diagnosis (mechanisms by which procedural deviations and anomalies are detected).
- Physical collaboration (the distribution and sharing of physical actions with robotic agents).
- Adaptation (the strategies through which the system personalises and dynamically adjusts support).

2.4.1 Digital work instructions and multimodal guidance

A foundational area of research concerns the transformation of work instructions from paper manuals to digital, multimodal, and context-aware guidance. AR has been particularly prominent as an instruction technology due to its capability to overlay spatially aligned guidance directly onto the physical workspace. Wang et al. (2022) [29] provide a comprehensive review of AR-based instruction in manual assembly, training, and repair, summarising decades of research and highlighting both the promise of AR instructions and the technical and human-factors constraints that limit adoption.

Empirical evidence often demonstrates that digital and AR instructions can improve completion time and reduce errors, but outcomes depend strongly on task type and interface design. AR-based digital work instructions report faster completion and lower perceived workload in the assisted condition, along with user preference for the digital system, against paper documents [29]. Nevertheless, AR systems can also increase mental workload if they introduce display distress, field-of-view constraints, or attention switching, which is an issue frequently highlighted in AR and industrial human-factors research.

With specific regard to inclusive manufacturing, the question of instructional modality becomes even more relevant. The literature shows that oral, paper-based, and augmented-reality instructions do not support all operators in the same way, particularly when workers present cognitive or motor deficiencies [4]. Rather than supporting a uniform instructional model, existing studies suggest that different

modalities may respond to different support needs, making a ‘one-size-fits-all’ approach inadequate [3]. A similar conclusion emerges from research on projection-based and adaptive projection instructions, which indicates that the form in which guidance is delivered can also affect early-cycle performance in contexts involving operators with heterogeneous disability profiles.

2.4.2 Error detection and procedural mistake analysis

A second major strand focuses on the real-time detection and classification of assembly errors. In complex manufacturing tasks, errors can be classified not only by outcome (i.e., defect present) but also by procedural deviation: wrong component selection, wrong tool, incorrect orientation, missing step, or incorrect sequence [10], [11]. In practice, many prototypes rely on computer vision models (e.g., object detection) to recognise components and infer step completion [10], [11]. A recent investigation by Fan et al. (2025) [10], [11] implements such pipelines using modern object detectors and structured prompts or logic to translate detection outputs into corrective guidance.

The broader literature increasingly frames this issue as part of mistake analysis in procedural activities, with applications ranging from industrial tasks to rehabilitation and training. This perspective is particularly relevant because it shows that error detection cannot be reduced to a simple object-recognition problem [30]. In many cases, errors are temporal or relational: the correct objects may all be present in the workspace, yet the task is still performed incorrectly because they are manipulated in the wrong sequence or under the wrong conditions.

A related applied trend is the use of ‘digital-physical fusion’ approaches and data-driven models of assembly errors, which aim to integrate physical process signals with digital representations to model error generation and testing. While these technologies are not purely inclusive by default, they serve as foundational enabling blocks for inclusive support architectures. By providing highly reliable, low-latency error detection, they can successfully drive the adaptive, personalised instructions required by a neurodiverse workforce [3], [9], [31].

2.4.3 Human-technology collaboration with cobot-based assistance

Collaborative robots (cobots) introduce a critical layer to assembly assistance. In manufacturing environments, HRC can support by handing over parts, holding components, executing precise marking, or assisting in error recovery steps, often with lower barriers than full automation, especially for flexible, multi-variant production. Recent studies synthesise progress in collaborative technologies for

assembly and welding and frame HRC to improve productivity and quality while enabling complex operations that exceed either human or robot capability alone [32].

For inclusive manufacturing, HRC is relevant in two main ways. First, robots can take over physically demanding or precision-sensitive subtasks, reducing the effort required to maintain alignment, hold tools, or perform fine measurements. Second, cobots can also act as interaction devices, for example by moving projectors, pointing lasers, or providing step cues, as shown in recent inclusive-assembly prototypes [8]. At the same time, HRC increases the importance of autonomy and safety design. Shared-workspace hazards must be managed through appropriate standards, risk assessment, human oversight, and emergency procedures. Demonstrating that these safeguards remain effective across real operating conditions is still a significant challenge.

2.4.4 Operator profiling and personalised assistance

System personalisation is often presented as an enabler of inclusion because adapting modality, pacing, level of detail, or interaction style may help address different operator needs. The difficulty is that effective personalisation requires either explicit user profiles or real-time estimates of the operator’s state [31], [33].

Manufacturing research explores this problem through several approaches. The Operator 4.0 and H-CPS perspectives describe workers supported by adaptive automation and advanced interfaces [26]. A related concept is the HDT, which represents aspects of the worker, such as skills, state, or ergonomics, to support adaptation and human-centred optimisation.

One approach estimates cognitive load across dimensions such as attention, memory, language, mathematics, logic, and reading, then uses these estimates to classify operator status and select support strategies, including RL loops and technology-based assistance [8]. Another builds cognitive or preference profiles to choose among instruction modalities such as audio, visual projection, or robotic support [9].

2.4.5 GenAI as instruction generator and decision-support component

The introduction of GenAI changes the logic of assembly assistance in an important way. Traditional assembly-support systems generally rely on predefined workflows. When a deviation is detected, the system typically retrieves an instruction, warning, or rule-based response that has already been specified in advance.

GenAI changes this mechanism by using task information, contextual data, and the detected error state to produce guidance for the situation at hand. The response is

therefore generated from the available context instead of being selected directly from a fixed library. This marks a relevant shift from retrieval-based support toward more context-sensitive forms of interaction.

This shift matters because the system is no longer limited to indicating that an error has occurred. Instead, it may also generate a more tailored explanation of what went wrong and propose a corrective action in a form that is easier for the operator to understand and pursue. Depending on the architecture, this output may be expressed through natural-language text, speech synthesis, simplified procedural prompts, decision-tree-like instructions, or multimodal combinations that include visual overlays and interface cues [9], [10], [11]. In inclusive settings, this flexibility is particularly valuable because the format, the pace, and the level of detail of the instruction may substantially affect usability for operators with different cognitive profiles.

A dynamically generated explanation may therefore function not only as a corrective tool, but also as a bridge between detection and actual human understanding. This architecture suggests that LLMs can effectively serve as a ‘glue layer’ between perception and interaction, which is especially valuable in inclusive systems where instruction format and tone can be tailored by adapting in different fields, which could also be interconnected, as illustrated in Figure 2.3.

At the same time, the use of GenAI introduces new reliability and governance challenges. Generated outputs may vary depending on prompt design, missing context, ambiguous inputs, or unusual edge cases, and this variability may create problems of consistency, traceability, and verification. For this reason, the most promising direction is not unrestricted free-form generation. Current approaches increasingly combine GenAI with structured output constraints, predefined vocabularies, domain rules, validator layers, and human supervision. [34]. In this sense, the role of GenAI is not to replace other forms of control, but to make assistance more responsive to context while ensuring that system behaviour remains bounded, reviewable, and consistent with its intended assistive purpose.

From the perspective of this thesis, this development is particularly important because it expands both the technical possibilities and the governance challenges associated with assistive systems. On the one hand, GenAI may help reduce unnecessary cognitive effort, support recovery from errors, and enable instructions to adapt more effectively to the needs of the user. On the other hand, it raises questions about output reliability, transparency, validation, and the boundaries of acceptable machine-generated support in workplace settings. It is therefore not sufficient to ask whether generative outputs are useful. It is also necessary to ask under which technical and governance conditions they remain understandable, contestable, and appropriately constrained in practice.

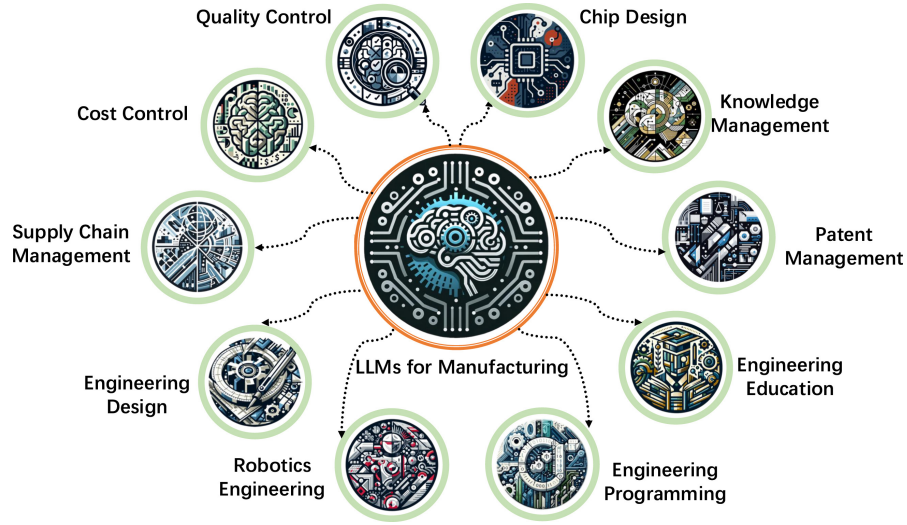


Figure 2.3: LLMs and GenAI applications in manufacturing, adapted from [35].

2.5 Recurrent failure modes in AI-assisted assembly and inclusive support systems

Failure modes should not be treated as rare or unforeseeable anomalies in AI-assisted assembly, since they are better understood as recurring risks that emerge from the interaction between variable environments, imperfect models, shifting data distributions, and limited human attention.

Failures become particularly relevant in inclusive support systems because their effects are not limited to production efficiency. An incorrect detection, poorly timed prompt, or unsuitable adaptation may increase cognitive effort, undermine trust, or make the interface harder to use. For operators who already face barriers in task execution, repeated failures can therefore weaken the value of the assistance itself. Failure behaviour should consequently be examined as part of the system's normal operating profile, not only as an exceptional deviation.

Table 2.2 summarises the main patterns discussed in the literature. False positives, false negatives, and downstream error cascades arise primarily from limitations in detection and interpretation. A different set of problems appears over time, when dataset shift or concept drift reduces performance as environmental conditions, process characteristics, or operator behaviour move away from those represented during validation. A further group concerns sociotechnical dynamics, including alert fatigue, over-reliance, disuse, and weak explainability.

Taken together, these failure modes show that the quality of an assistive system depends not only on whether it can generate outputs, but also on whether those

outputs remain accurate, interpretable, and usable under real production conditions.

2.5.1 False positives, false negatives, and error cascades

In perception-driven assistance, a core technical risk is the trade-off between false positives and false negatives. False positives (i.e., unwarranted alerts) frequently lead to unnecessary task interruptions and a rapid loss of user trust, whereas false negatives (i.e., missed detections) allow critical errors to propagate into severe product defects or unsafe operational states. Even under highly controlled laboratory conditions, object detectors routinely confuse visually similar components. Several recent prototypes targeting inclusive assembly explicitly report persistent confusion between geometrically identical fasteners, illustrating that this risk is structural and algorithmic rather than merely accidental [10], [11].

More broadly, the procedural mistake detection literature highlights that ‘mistake analysis’ is hard because it must separate permissible variation from true errors. This distinction is critical in sequential assembly, where a single misclassified step can generate profound downstream ambiguities. The system may misinterpret all subsequent physical states, triggering an ‘error cascade’ characterised by a continuous stream of incorrect prompts and misguided corrective actions.

2.5.2 Environmental variability and model degradation

Manufacturing environments are particularly prone to dataset shift [8], [10]. Dynamic variables such as fluctuating lighting conditions, camera repositioning, tool wear, component batch variations, and evolving operator behaviours can drastically alter data distributions relative to the original training set [36].

The general problem is that predictive models can fail when the joint distribution differs between training and deployment [37]. In digital or evolving environments, concept drift is becoming an important concern, as the relationship between input data and target outcomes may change over time [12].

These risks are particularly relevant for inclusive systems, which often depend on long-term use and continuous personalisation [31], [33]. When a system adapts continuously to individual operators or changes through feedback loops, it is fundamental to adopt careful measures of governance.

Without systematic testing, an algorithmic improvement optimised for one specific scenario or user profile might inadvertently degrade performance in another. This critical vulnerability is explicitly recognised in recent reviews of cognitive workload and human-centric manufacturing, which advocate for robust validation methodologies and continuous, lifecycle-orientated monitoring, explicitly moving away from one-off laboratory validations [1].

2.5.3 Alert fatigue, alarm overload and over-reliance

A common sociotechnical failure mode is alert fatigue: when the user is exposed to frequent or low-value alerts, attention and responsiveness can decrease, and the system can become background noise [27], [30]. Much of the formal ‘alert fatigue’ literature originates in clinical decision-support contexts, but the underlying phenomenon, desensitisation under frequent non-actionable alerts, is a general human-factors concern.

In industrial settings, the problem of excessive alarms has long been addressed through dedicated alarm management standards and guidelines. Alarm systems should support operator awareness and decision-making, rather than create additional overload. This idea is also relevant to assembly assistance: systems that generate warnings or corrective prompts can be seen as a form of procedural alarm system [27]. Their design therefore requires similar attention to issues such as alert prioritisation, control of notification frequency, contextual relevance, and ongoing management throughout the system lifecycle [30].

Furthermore, inappropriate trust calibration could be manifested in two dangerous extremes. Over-reliance occurs when an operator blindly follows system prompts, even when their own situational awareness or sensory evidence strongly suggests otherwise [30], [38]. Conversely, disuse or under-reliance occurs when an operator actively ignores the system following a history of repeated false positives. Both extremes are uniquely critical in inclusive contexts: if a system is deployed to support neurodivergent operators or those already facing cognitive barriers, their reliance patterns will be complexly shaped by both fluctuating system trust and their underlying perceived self-efficacy.

2.5.4 Explainability and the limits of transparency

While XAI [30] is frequently invoked as a strict prerequisite for trustworthy decision support in industrial settings, achieving meaningful explainability remains notoriously difficult. This is particularly true when opaque deep learning models are deployed for visual perception, or when LLMs are utilised to dynamically generate instructions.

In assembly assistance, ‘explainability’ can be interpreted pragmatically: the operator needs to understand why an alert was triggered, how confident the system is, and what evidence supports the proposed correction. For computer-vision-based systems, effective explainability may involve showing the object currently detected alongside the one expected at that stage of the task.

In systems based on procedural logic, it may instead mean clearly indicating which step or constraint has been violated. For GenAI-driven systems, explanations should make clear which procedural rule, section of the manual, or previous step

was used to generate the instruction. Without this foundational cognitive framework, system explanations risk becoming generic and opaque, and ultimately it can fail to support operator contestability or meaningful human oversight [28], [30], [38].

Table 2.2: Failure modes related to their typical symptoms and their likely causes. Adapted from [27] and [30].

Failure mode	Typical symptoms	Likely causes
False positives (FP)	Unnecessary alerts; interruptions; declining trust	Over-sensitive thresholds; noisy inputs or occlusions; weak class separation
False negatives (FN)	Missed errors; defect propagation; late detection; safety risk in critical steps	Insufficient training coverage; occlusion; ambiguous states; missing temporal logic
Dataset shift / Concept drift	Degradation over time; new variants; camera shifts	Changing distributions; tool wear; process changes; lighting variation
Alert fatigue / Overload	Ignored alerts; delayed response; workarounds; reduced situation awareness	High alert rate; low actionability; repeated prompts; poor prioritisation
Over-reliance / Disuse	Blind following or complete ignoring; reduced skill retention; unsafe overrides	Miscalibrated trust; opaque confidence; inconsistent performance; lack of training
Explainability gaps	User confusion about ‘why’; inability to contest; reduced adoption	Black-box perception; free-form generation; missing provenance to step rules

2.6 Approaches and designs for assembly assistance and inclusive manufacturing systems

Given the inherently socio-technical nature of inclusive manufacturing systems, comprehensive evaluation frameworks must simultaneously address both rigorous

technical performance and universal human outcomes [39], [40]. Across the contemporary literature, evaluation methodologies are typically structured around a united combination of objective performance metrics, cognitive workload assessments, usability testing, and technology acceptance modelling [6], [13], [29].

2.6.1 Performance metrics: time, errors, and recovery

The most common quantitative indicators in assembly assistance studies include task completion time, error rate, rework frequency, and error recovery time. These metrics are attractive because they map directly onto production outcomes. For example, numerous AR instruction studies benchmark time-to-completion and error counts across varying instruction [29], [39]. At the algorithmic level, error-detection models are conventionally evaluated using a confusion matrix and precision metrics [10], [11]. However, task-level evaluation must ultimately determine whether the system successfully prevents physical defects and minimises recovery time [39]. Furthermore, the procedural mistake detection literature increasingly emphasises the necessity of metrics that capture both error recognition quality and temporal segmentation quality. In fact, a system that accurately classifies an error but detects it too late in the operational sequence is functionally ineffective on the shop floor. However, relying exclusively on objective performance metrics can produce deeply misleading conclusions in the context of inclusive design: faster task completion does not inherently equate to superior human-centric outcomes [3], [13].

2.6.2 Workload and CLT: NASA-TLX and beyond

The NASA-TLX remains one of the most widely used subjective instruments for workload assessment in industrial experiments.

Its enduring adoption reflects its ability to evaluate workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Hart and Staveland (1988) [6] established its rigorous empirical grounding, and it continues to serve as a standardised benchmark. Within Industry 5.0-orientated reviews, the NASA-TLX is frequently used as the fundamental subjective baseline, ideally complemented by objective physiological data [12], [13]. Indeed, the broader cognitive ergonomics literature increasingly promotes multimodal workload assessment. Lucchese et al. (2025) [13] explicitly identify a methodological paradigm shift toward integrative approaches that fuse subjective questionnaires with continuous data from physiological sensors, such as EEG, eye-tracking, and HRV monitors.

In practice, within the scope of applied system evaluation and prototyping, a layered methodological approach is often the most pragmatic: deploying the NASA-TLX

(or similar validated scales) as a robust, highly deployable baseline while systematically integrating physiological and behavioural proxies only when technically feasible and environmentally justified [37].

2.6.3 Acceptance, trust, and organisational fit

User acceptance and operator trust play a central role in whether an assistance system is adopted and used consistently in practice.

The TAM, introduced by Davis (1989) [40], explains technology adoption mainly through two factors: perceived usefulness and perceived ease of use. Although the model was originally developed for conventional information systems, these concepts are also relevant to industrial assembly assistance. Operators are less likely to rely on a system when its benefits are unclear or when the interface adds unnecessary effort to the task [40].

Trust is equally important in automation and robotics because it affects when and how users rely on system outputs. This is closely connected to the failure modes discussed in Section 2.5. Repeated false alarms, inconsistent GenAI outputs, or unreliable prompts can gradually reduce confidence in the system, while opaque behaviour may make long-term acceptance more difficult [27], [38], [40].

2.6.4 Inclusive evaluation across user diversity and long-term effects

A distinctive methodological challenge in inclusive manufacturing is that evaluation should include representative user diversity, which is often difficult due to recruitment, privacy, and ethical constraints [3], [4], [31].

Studies like Vanneste et al. (2020) are notable precisely because they include participants with disabilities, allowing a more direct assessment of whether a modality supports the intended users [4]. Peltokorpi et al. (2023) similarly examine different instruction forms in a sheltered workplace, showing how evaluation methods can be adapted to inclusion-oriented settings [3]. What remains much less explored is what happens after these systems are used over longer periods.

Laboratory studies often report short-term improvements in performance, but they provide limited evidence about how workers are affected by continued use. Over time, assistance may encourage learning and skill development, but it may also create dependence on the system or contribute to the gradual loss of abilities that are no longer exercised.

For this reason, recent Industry 5.0 reviews increasingly call for evaluation approaches that go beyond short laboratory studies. More attention is needed on longitudinal research that examines real deployment conditions and assesses how human outcomes evolve over time [1].

2.7 Research gaps and thesis positioning

The literature reviewed in this chapter shows that AI-assisted assembly systems are moving towards more integrated and adaptive configurations. Real-time perception, decision logic, multimodal guidance, workload estimation, and collaborative interaction are increasingly combined within the same assistance architecture rather than developed as separate technical functions.

This development reflects both the practical demands of variable assembly environments and the broader direction of Industry 5.0. On the shop floor, systems are expected to support error prevention, recovery, and adaptation to changing conditions. At the same time, Industry 5.0 places greater emphasis on worker well-being, inclusion, and sustainable forms of human-machine collaboration. The technical literature has advanced rapidly in response to these needs, but governance approaches have not always developed at the same pace. Once questions of data protection and AI governance are included in the analysis, several unresolved issues emerge.

One of the clearest concerns is the transition from laboratory validation to industrial deployment [29], [39]. Many assistance systems are validated in controlled settings that do not fully reproduce the variability of industrial work. Noise, organisational constraints, differences in operator behaviour, and performance drift over time may be limited or absent during testing, making later behaviour on the shop floor harder to anticipate.

The problem extends beyond experimental validity [12], [13]. Once a system becomes part of everyday work, its scale, frequency of use, and dependence on personal data may alter the balance between necessity, proportionality, and risk. A configuration that appears minimally intrusive in a laboratory may become harder to justify when it operates continuously, processes more variable data, or becomes embedded in organisational routines.

A related challenge concerns workload estimation and adaptive personalisation [31], [33], [37]. Research proposes increasingly sophisticated methods for modelling cognitive workload, yet there remains a substantial distance between sensor-rich prototypes and systems that can be deployed at scale without becoming overly intrusive. The difficulty is partly technical, but it also concerns how much data should be collected to achieve a meaningful assistive benefit. Greater reliance on physiological, behavioural, or inferred data brings questions of minimisation, proportionality, retention, and lawful purpose closer to the centre of system design. Choosing between richer sensing and a less intrusive architecture is therefore not only an engineering decision [17], [22], [41], [42].

The use of generative AI creates a further area of uncertainty [34], [35], [43]. GenAI can make industrial assistance more flexible by generating instructions, explanations, or context-sensitive responses. However, this flexibility also raises questions

about reliability, traceability, output validation, transparency, and accountability. Manufacturing research has only recently begun to develop more stable approaches for constraining generated outputs and keeping them consistent with procedural rules and safety requirements. In workplace applications, usefulness alone is not enough. Outputs also need to be reviewable, verifiable, and limited to a clearly defined assistive purpose.

Long-term evaluation remains another weak point in the literature [3], [4]. Most studies examine short-term outcomes such as task speed, accuracy, workload, or perceived usability. Much less is known about what happens after repeated use. Trust may increase or decline, skills may be strengthened or gradually lost, dependence on the technology may emerge, and frequent prompts may lead to alert fatigue. Longer deployments may also normalise increasingly data-intensive forms of assistance. These effects are important because a system cannot be considered genuinely human-centred only on the basis of short laboratory trials. From a governance perspective, short-term validation may also fail to reveal how the balance between support, dependence, and intrusiveness changes over time.

Perhaps the most transversal gap is the limited integration of technical evaluation with legal-by-design governance [17], [23]. Compliance is often addressed only after the main technical architecture has already been established. The GDPR and the broader AI governance framework suggest a different approach, in which questions of data protection, transparency, risk, and human impact shape design decisions from the outset.

The main gap is therefore not the absence of legal discussion around industrial AI. What is still limited are practical methods for bringing technical evaluation, data governance, and ex-ante assessment into the same design process.

These unresolved issues provide the basis for the position adopted in this thesis. AI-assisted assembly systems are examined here as socio-technical architectures rather than as collections of separate technical modules. Their performance, workload effects, usability, robustness, trust, and adaptability are considered together with questions of data minimisation, proportionality, transparency, retention, and ex-ante impact assessment. Legal compliance is therefore treated as part of system design rather than as a downstream exercise.

The thesis is situated at this intersection between engineering and legal-regulatory governance. Its underlying premise is that the long-term viability of AI-assisted work systems depends not only on what they can technically achieve, but also on whether their architecture remains understandable, proportionate, and manageable over time [18], [19], [41].

Chapter 3

Data and data governance in AI-assisted production systems

This chapter examines the role of data in AI-assisted production systems and explains why data governance is central to their legitimacy. It distinguishes between different categories of data, including personal data, inferred data, behavioural data, physiological data, and special-category data, and connects these distinctions to GDPR principles such as purpose limitation, data minimisation, storage limitation, fairness, and transparency. The chapter also introduces the governance logic that will later be applied to the three selected systems.

3.1 The role of data in AI-assisted systems

In AI-assisted systems, data are not simply an input among many others. Data enable the system to observe what is happening, relate those observations to the task, adapt the support provided, and, in some cases, infer aspects of the operator's state.

Recent manufacturing research reflects this wide range of uses, from inclusive and multimodal assistance to digital work instructions, logic-based support, and cognitive-workload assessment in Industry 5.0 settings. The relevant point is that data rarely remain in their original form. They are processed into indicators, alerts, predictions, or other outputs that influence how the system responds.

Greco and Mantelero (2018) [41], in their analysis of Industry 4.0, robotics, and privacy by design, describe the contemporary smart factory as a dense informational ecosystem in which the data generated by the production environment also entail the processing of information relating to the human subjects operating within it. Their argument is particularly important because it shifts legal attention away from isolated data points and toward the broader chain through which information

is transformed into decisions, recommendations, and adaptive responses within human-machine interaction. The same concern emerges in more recent GDPR scholarship on AI. Voigt and von dem Bussche (2017) observe that AI systems generate data-protection tensions not only at the point of operational use, but also during development and training, since they are often trained on large quantities of personal data and may generate inferences that themselves qualify as personal data or, in some cases even special categories of data [23].

Accordingly, the key legal question is not simply whether an assistive system uses data, but what data are involved, why they are collected, at which stage of the AI lifecycle they are processed, whether individuals can be identified, and what safeguards are in place.

From a regulatory perspective, this issue must be considered in light of both the GDPR and the AI Act. The European Commission presents the AI Act as the EU's comprehensive legal framework for AI, based on a risk-orientated approach that aims to promote trustworthy AI while addressing potential impacts on health, safety, and fundamental rights. The AI Act entered into force on 1 August 2024 and will be fully applicable from 2 August 2026, subject to a staggered timetable for certain provisions [18], [19]. At the same time, the AI Act does not displace the GDPR but supplements it in important respects. For workplace AI, this means data issues cannot be treated as a secondary compliance matter. They are central both to lawful personal-data processing and to the broader governance of AI systems.

3.2 Types of data involved in industrial assistive and manufacturing systems

AI-assisted industrial systems can process a heterogeneous range of data, and any serious legal assessment must begin by distinguishing among them. At a basic level, these systems may process identification and authentication data, operational and process data, contextual and environmental data, interaction data, voice data, physiological data, and data derived from sensors embedded in wearables, workstations, interfaces, or collaborative devices. Recent studies increasingly examine concrete human-machine interaction scenarios involving biometric data, robotic exoskeletons, adaptive workstations, geo-referenced information, smart suits, voice data, voice-identification systems, and voice commands directed to robots or AGVs [22], [41].

This distinction is important because different types of data raise different legal and technical concerns. Operational data, such as cycle times, task completion records, workstation states, or command histories, may appear to be purely industrial, but they can still be linked to a specific worker. Voice data may simply be used as an input for commands, while in other cases they may qualify as biometric

data when processed through specific techniques for the unique identification of a natural person. Physiological data may instead be collected to support ergonomic adaptation, estimate fatigue, or improve safety. They constitute personal data when they relate to an identified or identifiable individual and may also fall within the category of health data when they reveal information about that person's physical or mental condition. The EDPS expressly states that health data include not only conventional medical records but also personal information more broadly relating to a person's health status, and such data are subject to particularly strict protection requirements [44].

For AI-assisted assembly, another important distinction is between raw data and transformed data. A camera feed, a vocal trace, or a heart-rate stream is rarely the endpoint of the system. In practice, these signals are often transformed into templates, indicators, scores, alerts, or adaptive responses. This is particularly clear in voice-based systems, where raw vocal data may be converted into templates that are later used for comparison or recognition [41].

A similar process occurs in adaptive workstations and cognitive-support systems. Here, raw parameters become operationally relevant because they can be used to adjust system behaviour or to estimate factors such as strain, workload, and usability. It is also useful to distinguish data according to the phase of the system lifecycle in which they are processed. Controllers must differentiate between data used as part of an AI system in operation and data used during development, training, or further refinement of the model because the risks, purposes, and potentially the legal basis may differ across those phases.

This distinction matters in industrial settings because data collected while an assistance system is operating usually serve an immediate function, such as supporting task execution or adapting the response in real time. Their later use for model training, optimisation, or system development cannot therefore be assumed automatically.

Before such reuse takes place, the organisation needs to examine why the data are needed again and whether the new purpose remains necessary and compatible with the context in which they were originally collected. Data governance, for this reason, begins with a functional mapping of which data are processed, at which stage of the system lifecycle, and for which specific use.

3.3 Personal data, special categories of data, inferred data, and behavioural data

The GDPR adopts a deliberately broad notion of personal data. Recital 26 clarifies that the Regulation applies to information relating to an identified or identifiable natural person, not only to data that directly reveal a person's name or identity.

This is particularly relevant in industrial settings, where a worker may become identifiable through the combination of information such as workstation assignments, credentials, timestamps, access logs, or task histories, even when no single explicit identifier is present. Indeed, production environments generate not only process information but also data relating to those who work within them [18].

Physiological and biometric data require particular caution in legal and governance terms, since their intimate connection with the bodily and, in some cases, health-related sphere of the individual may significantly increase the risks associated with identification, inference, misuse, or disproportionate processing. Within this broad field, a first distinction must be drawn between ordinary personal data and special categories of data. The latter includes, under Articles 4 and 9 GDPR, data concerning health and biometric data when processed for the purpose of uniquely identifying a natural person [45].

Greco and Mantelero's discussion of voice-identification systems is instructive in this regard: they explain that voice traces, once sampled and converted into templates used for identification, fall squarely within the logic of biometric processing [41]. In the same way, auto-adaptive workstations and exoskeleton-related systems may involve physiological or biometric information where the system calibrates itself to the anthropometric or physiological characteristics of the operator.

What the system derives from the data can be as important as the data collected at the start. AI may be used to draw conclusions about individuals or groups, and those conclusions can acquire legal relevance of their own. Where an inference identifies a person, makes that person identifiable, or reveals sensitive characteristics, it may itself qualify as personal data and, in some cases, fall within a special category of data.

For this reason, the legal assessment of an AI-assisted system cannot stop at the raw inputs. It also needs to consider the information produced through inference. Examples include estimates of cognitive load, anomaly alerts, assessments of fatigue, or predictions that support should be adapted or intensified. The data-protection analysis must therefore extend to what the system concludes, not only to what it initially records [46]. A further type of data, closely linked to inferred data, is behavioural data.

The Article 29 Working Party (2012) warns that new workplace technologies enable systematic and potentially invasive forms of data processing, and those secondary traces, including metadata and activity data, may reveal detailed behavioural patterns even where the content of communications is not directly examined [42]. Behavioural data, in the present context, should therefore be understood as data concerning movements, habits, interaction patterns, timing, responses to prompts, navigation through digital instructions, or command sequences. Such data are not necessarily special categories of data, but they may become highly sensitive because of the patterns and conclusions that can be extracted from them [47].

For present purposes, it is important to keep a methodological distinction in view. This thesis does not primarily examine profiling or worker surveillance from a labour-law perspective. However, inferred and behavioural data remain highly relevant because they show how assistive systems can affect workers through interpretation and adaptation. Even without explicit profiling, an AI-assisted system may still build a legally relevant representation of a worker's condition, capabilities, or support needs. That is why the classification of data should not stop at raw categories and must include the system's interpretive layer as well.

3.4 Aggregated, anonymised, identifiable, and pseudonymised data

A second indispensable distinction concerns the degree of identifiability of the data. Identifiable data, pseudonymised data, anonymised data, and aggregated data are often presented as forming a simple linear progression. In reality, their relationship is more complex, and complexity matters greatly in AI-assisted work systems. Data remain identifiable when a person can still be recognised directly or indirectly, which keeps them within the scope of data-protection law.

Pseudonymisation does not remove that link completely. Instead, it separates the identifying information and protects it through additional safeguards. Its practical value is greatest when the controller cannot readily reconnect the data to the individual, or does not routinely use the information needed to do so. However, Big Data and AI contexts make privacy by design, privacy by default, and impact assessment especially important precisely because complex analyses often remain based on pseudonymised rather than anonymised data.

In relation to auto-adaptive workstations in industrial settings, it is proposed to separate biometric and physiological data from workstation-related information into two distinct databases and to allow only exceptional and controlled procedures to reconnect them, for example, in the event of an accident, and with logging of the operator carrying out the re-linking procedure.

Analyses of voice-based systems follow the same logic: identifying data may be separated from biometric templates, stored in non-communicating databases, and re-linked only through exceptional procedures. This is highly relevant for AI-assisted assembly because it shows that pseudonymisation is not a merely abstract legal label and it can be translated into concrete architectural choices about database design, access control, and logging.

By contrast, anonymisation excludes data from the scope of data-protection law only where re-identification is no longer reasonably likely. Recital 26 GDPR makes clear that the standard is not formal but contextual: the relevant question is whether identification remains reasonably possible by means likely to be used [18]. In the

case of voice-identification systems, for example, they argue the biometric nature of the template and its functional role make true anonymisation impracticable without destroying the utility of the template itself. In the case of adaptive workstations, they further note that even where the operator is not directly identified, the specific workstation context may still make it relatively easy to reconnect physiological data to the person occupying that position at a given time.

Aggregated data, instead, occupy a different position. Aggregation may reduce risk, but it does not automatically achieve anonymisation. The Article 29 Working Party (2012) opinion on workplace processing expressly warns, in the context of wearable devices and health-related information, that even aggregated health information may remain problematic in employment settings because singling out can still occur where the employer has access to contextual information allowing re-identification [42]. Aggregation can be useful for historical analysis and statistical reporting because it may reduce the need to retain individual-level records. In a smart factory, for example, system-wide dashboards or workload trends may be less intrusive than data tied to specific workers.

That reduction in detail does not automatically remove legal concerns. The key question is whether individuals can still be singled out or re-identified once the aggregated information is combined with the surrounding operational context. This point also affects how the system should be designed. The key issue is not whether data are formally labelled as anonymous or aggregated, but whether the architecture genuinely reduces the possibility of linking them back to a specific worker [41]. For this reason, measures such as pseudonymisation, separation of datasets, and selective aggregation should be understood as practical governance techniques rather than as mere labels.

3.5 Core data protection principles

Once the nature and identifiability of the data have been clarified, the analysis turns to the principles governing their processing. Under Article 5 of the GDPR, purpose limitation, data minimisation, storage limitation, fairness, and transparency are not peripheral obligations but the normative core of lawful processing.

The Regulation requires that personal data be collected for specified, explicit, and legitimate purposes, be adequate, relevant, and limited to what is necessary in relation to those purposes and be kept in identifiable form no longer than necessary. Purpose limitation is particularly important in AI-assisted work systems because the same data may appear useful for several purposes that are only partly related [22].

Data collected for one operational purpose may later appear useful for very different activities. Information gathered to support ergonomics, adapt instructions,

or authenticate voice commands, for example, could later be reused for analytics, model retraining, optimisation, or even worker evaluation.

These uses should not be treated as part of the same processing context. Data needed to operate the system serve a different function from data retained for model development or training, and the legal assessment may change accordingly. For the present thesis, this means that data collected for assistance cannot simply be presumed reusable for model development or secondary analytics.

Data minimisation and proportionality require that only data which are adequate, relevant, and necessary for the specific purpose be processed. The Article 29 Working Party (2012) states this particularly clearly in the employment context: employers must undertake a proportionality assessment before deploying new technologies, must ask whether all data are genuinely necessary, and must minimise both the information collected and the information shown to the employer [42]. Although this thesis is not centred on workplace monitoring as such, the same logic is directly applicable to assistive systems.

The relevant question is whether the same functional objective can be achieved with fewer data, lower granularity, shorter retention, local processing, or less intrusive technical alternatives.

The collection of biometric data might in some cases be avoided through adjustable or less personalised configurations based on approximate bodily measurements rather than fully individual biometric modelling. In relation to robots responding to voice commands, for example, controllers should assess whether, considering minimisation and proportionality, different technical solutions, such as pseudonymisation or even non-voice command systems, could better balance operational needs with the necessity requirement.

Storage limitations must be treated with equal seriousness. Personal data should be retained only for the minimum period necessary in relation to the specific purpose of the processing, which means that controllers should establish retention and deletion policies identifying the categories of data stored, the purposes justifying their retention, and the corresponding deletion deadlines [23]. In parallel, where processing serves only the correct execution of the work activity, automatic deletion at the end of the shift or working day may be appropriate, while longer retention should be limited to anomalies, incidents, or specific safety-related reasons [41].

A clear distinction could be based on the area of interest: in ordinary areas, voice-command data, as an example, may be deleted immediately after identification or retained only for a short period, such as no more than twenty-four hours, where minimal security needs justify temporary storage. In the case of adaptive workstations, instead, physiological parameters may be deleted after the shift where no anomaly is detected, while longer retention may be justified where health-related verification is genuinely needed.

Fairness and transparency complete this framework. Workers must receive effective

information about the existence of processing, its purposes, and the surrounding circumstances. In this regard, the Article 29 Working Party warns that new technologies may operate in ways which are far less visible to employees than traditional monitoring tools [42]. In AI-assisted systems, this point is particularly relevant because the most consequential part of the processing may be inferential rather than visible: the user may see only a helpful interface, while the system internally derives workload indicators, behavioural patterns, or adaptation profiles.

Fairness therefore does not concern only the formal lawfulness of collection, but it also concerns whether the system's informational asymmetry remains compatible with the worker's reasonable expectations and with the declared assistive purpose.

3.6 Data protection by design and by default

Article 25 GDPR brings these principles directly into system design by requiring controllers to adopt appropriate technical and organisational measures by design and by default.

In practice, privacy by design may involve measures such as data minimisation and timely pseudonymisation. Privacy by default, instead, concerns the ordinary configuration of the system, including how much personal data are collected, how extensively they are processed, how long they are retained, and who can access them. Access should therefore be limited according to role and operational need, so that personal data are available only to those who genuinely require them for their tasks.

In the workplace context, the Article 29 Working Party (2012) [42] had already anticipated this logic by stating that where employers deploy new technologies, privacy-friendly solutions should be preferred, and minimisation must be taken into account at the design stage.

In AI-assisted assembly systems, this means that privacy and data protection cannot be added after deployment as a purely documentary compliance layer, but they must be built into the system architecture from the outset. The relevant design choices concern not only notices and policies, but also sensor selection, database separation, local versus centralised processing, retention logic, default settings, access permissions, and auditability. Figure 3.1 summarises how the main data protection principles can be translated into a sequential design logic for AI-assisted assembly systems, from purpose definition to retention management and lifecycle review.

Greco and Mantelero's contribution is particularly valuable because it shows how data protection by design can be translated into specific industrial configurations [41]. Their analysis proposes, depending on the scenario, anonymous ergonomic adjustment where feasible, pseudonymised database structures where continuous

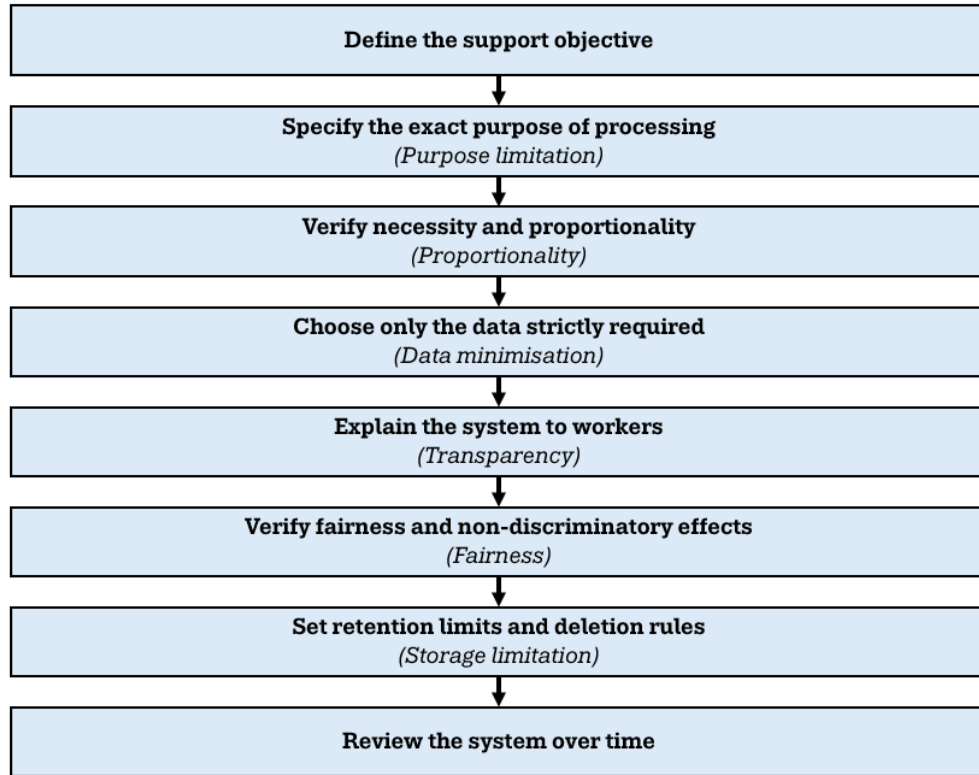


Figure 3.1: Operationalisation of core data protection principles in AI-assisted assembly systems.

adaptation is needed, decentralised or separated storage, exceptional and logged re-linking procedures, and automatic deletion mechanisms. In other words, privacy by design is not simply about limiting data use. It also provides a practical way to compare technical options in terms of necessity, proportionality, and potential impact.

This is especially relevant for assistive systems, which often depend on some degree of adaptation to function effectively. The aim is therefore not to remove personalisation altogether, but to identify the least intrusive data architecture that still allows the system to serve its intended purpose. In some cases, this may mean preferring local or on-device processing over remote centralisation, while in others, it may mean retaining only the adaptive output while deleting the raw signal, or it may mean making the more intrusive feature optional rather than default. The by-default dimension is therefore decisive: the ordinary mode of operation should not expose more data, for longer, or to more actors than the purpose genuinely requires.

3.7 DPIA as an ex-ante design and governance tool

The DPIA gives procedural form to this preventive logic. Article 35 GDPR requires a DPIA where a type of processing, especially one involving new technologies, is likely to result in a high risk to the rights and freedoms of natural persons. In the Article 29 Working Party (2012) on data processing at work, it is reiterated that a proportionality assessment should be conducted before the commencement of workplace processing and notes that this may form part of a DPIA [42].

For AI-assisted work systems, however, the DPIA should not be understood as a formal annex drafted after technical choices have already been fixed. Rather, it should operate as an ex-ante design and governance instrument. Before an AI system is deployed, it is necessary to assess whether the intended processing could be achieved through less risky alternatives and to consider, ex-ante, the possible adverse consequences of its use, together with the safeguards capable of mitigating them [23].

This makes the DPIA especially relevant where the system uses physiological, biometric, voice, or behavioural inputs, where it generates meaningful inferences, or where its functioning depends on complex and potentially opaque data flows. It is essential to apply GDPR principles to specific technologies and, at the same time, to develop impact-based methodologies in which risk assessment and mitigation measures are considered from the early stages of product and service design [41]. This preventive logic is at the core of a DPIA. The assessment is not limited to identifying a lawful basis for processing; it also examines why the processing is needed, whether it is proportionate, what risks may arise, and which safeguards should be in place before deployment.

The analysis should cover the processing operation as a whole, including its intended design and possible effects on data subjects. This requires a clear description of the proposed processing, an examination of necessity and proportionality, and an assessment of the risks to individual rights and freedoms.

For an AI-assisted assembly system, these general requirements can be translated into at least five practical areas:

- It should distinguish raw data from inferred data.
- It should separate operational use from training, validation, or system improvement.
- It should test whether less intrusive alternatives could achieve the same assistive purpose.
- It should define retention logic and re-identification controls.

- It should specify the organisational measures that make the chosen safeguards credible in practice, including access controls, logging, role allocation, and review procedures.

The AI Act further supports this approach. Rather than replacing the GDPR, it complements it. Article 26(9) links the governance of high-risk AI systems to the DPIA process by requiring deployers, where a DPIA is necessary under the GDPR, to make use of the information provided for those systems [19]. The European Commission’s overview of the AI Act also reflects its phased application and broader risk-based approach to governance. Even when an assistive system does not ultimately fall within the high-risk category, the broader point remains relevant: DPIA and AI governance are becoming increasingly connected through their shared focus on identifying and addressing risks before they materialise. This connection also opens the way to the FRIA introduced by Article 27 of the AI Act, since while the DPIA remains centred on risks arising from personal-data processing, the FRIA broadens the assessment toward the possible impact of high-risk AI systems on fundamental rights in the specific deployment context. This makes the FRIA a methodological bridge between legal obligations, risk analysis, and the concrete design of safeguards for human-centric and trustworthy AI [22].

3.8 Data governance issues for workplace AI

While data protection principles establish the conditions for lawful processing, data governance concerns how those conditions are maintained throughout the lifecycle of an AI-assisted system.

In workplace AI, governance goes beyond a general commitment to privacy or security. Organisations need clear rules for deciding what data are processed, why they are needed, who may access them, how long they are retained, and how these decisions are reviewed as the system moves from design to deployment and later evolves [17], [22], [41]. This is particularly important because new workplace technologies can enable increasingly systematic forms of processing, creating tension between organisational interests and workers’ reasonable expectations of privacy [15], [17], [48].

One basic distinction is between data needed for the immediate operation of the system and data used for development or later improvement [17], [23]. In an AI-assisted assembly setting, information may be processed in real time to support task execution, adapt instructions, or enable another immediate function. Similar data may also be valuable for model training, validation, debugging, or optimisation, but these are different uses with potentially different purposes, risks, and legal implications. Data collected during a shift to provide assistance should therefore not be treated as automatically available for later model development.

Clear governance is needed to prevent an unnoticed expansion from immediate support to secondary uses that were not properly defined when the data were collected [17], [22], [42].

Documentation is another essential part of this process [19], [23]. Its role is not limited to demonstrating formal compliance. It should preserve a clear account of the data architecture as the system changes over time. This includes the categories of data collected, the transformations applied to them, any inferences generated, the functions that depend on those data, and the boundaries between operational and secondary uses. This approach is consistent with Voigt and von dem Bussche (2017), who connect GDPR compliance with internal procedures, risk analysis, documented policies, and practical implementation measures rather than isolated legal statements [23].

In practical terms, documentation should make it possible to follow the main stages of the data lifecycle, from collection and transformation to storage, deletion, reuse, and possible retraining. Without such an overview, it becomes more difficult to explain why particular data remain necessary, assess whether a later use is compatible with the original purpose, or recognise when the system has moved beyond its initial scope [17], [23], [41].

Access control raises a related governance issue [41], [42], [44]. Lawful collection does not mean that all data should be available to all actors involved in the system. The same dataset may contain information needed for technical maintenance, immediate operation, or highly restricted analysis of physiological, behavioural, or inferred worker data. Access should therefore reflect actual functional need and the sensitivity of the information involved.

The EDPS, in relation to health-related data at work, emphasises that access should be limited to the information required for a specific function, that administrative staff should receive only what is necessary for their tasks, and that organisations should adopt appropriate access-control measures [44]. Although assistive systems do not always process formal health records, the same reasoning is relevant when physiological, behavioural, or inferred information is involved. In some cases, this may justify separating data across different databases and allowing re-linking only in exceptional circumstances through controlled and authorised procedures. Access control therefore becomes one of the practical ways in which proportionality is maintained inside the organisation [41], [42], [44].

Retention also needs to reflect the purpose and sensitivity of different data categories [18], [23], [41]. Storage limitation cannot be implemented through a single retention rule applied to the whole system. Real-time inputs, derived indicators, logs, and datasets used for training may serve different functions and should be treated accordingly. Workplace data should not be kept longer than necessary, and controllers need to reassess whether continued retention remains justified in relation to the original purpose [42]. The EDPS likewise stresses the need for

clear retention periods linked to the reason for processing [44]. For an AI-assisted assembly system, this may mean that some inputs exist only briefly while a task is being performed, others are retained for a limited period to support verification or incident analysis, and longer storage is permitted only where a separate justification exists. A differentiated approach of this kind is more consistent with storage limitation than retaining all available data under a single policy.

Purpose separation is equally important [17], [22]. Once a technical infrastructure is in place, the same data may appear useful for assistance, quality assurance, optimisation, analytics, retraining, safety review, or system evaluation. The availability of data, however, does not make these purposes interchangeable. Information collected to support a task should not automatically move into model development, behavioural analysis, or historical performance repositories simply because this is technically possible. Practical boundaries are therefore needed, both organisationally and technically, so that operational support remains distinct from later analytical or developmental uses.

Governance must also continue after deployment [19], [23], [41]. AI-assisted systems can change through software updates, new interfaces, additional data sources, different operating conditions, or subsequent retraining. Decisions that were reasonable at the start of the project may no longer remain appropriate after these changes. Changes in system function, scale, or data use can alter the justification for earlier design choices. Decisions about necessity, access, retention, and purpose should therefore be reviewed again when the system evolves. Governance cannot stop once the initial architecture has been approved.

Cross-border processing introduces another layer of complexity. Personal data may pass through cloud infrastructure, external providers, or storage environments located outside the European Union, raising additional legal and organisational questions. These issues are relevant but are not examined in depth here, since the thesis focuses mainly on data minimisation, purpose limitation, retention, and lifecycle governance.

The resulting approach is lifecycle-oriented. Operational data should remain distinct from data retained for training or development; transformations and data flows should be documented; access should follow actual functional need; retention periods should reflect purpose; and later uses should not quietly expand beyond the original assistive function. These arrangements also require periodic review as the system evolves.

Data governance is therefore not an administrative layer added after the technical architecture has been defined. It is part of the practical structure through which AI-assisted work systems can remain proportionate, understandable, and manageable over time [16], [17], [23].

3.9 Chapter synthesis

The analysis developed in this chapter shows that the legal relevance of data in AI-assisted work systems depends on more than the simple presence of AI or personal-data processing. What matters is how data move through the system: what is collected, how it is transformed, whether new inferences are generated, whether individuals remain identifiable, who can access the information, how long it is retained, and whether later uses remain consistent with the original purpose. From this perspective, GDPR principles are not external constraints on industrial design. They influence how the system is designed and how information moves through it.

One important example is the distinction between raw and derived data. In AI-assisted assembly, the legal analysis cannot stop at the initial inputs because behavioural indicators, inferred states, predictions, and other generated outputs may also relate to identifiable workers. In some cases, these outputs may themselves qualify as personal data or fall within special categories of data.

The degree of identifiability matters for the same reason. Identifiable, pseudonymised, anonymised, and aggregated data do not create the same level of risk and do not call for the same safeguards. Architectural choices such as separating databases, restricting re-linking, shortening retention, or deleting data automatically can reduce the chance of reconnecting information to a specific worker, even where full anonymisation is not feasible.

The core GDPR principles discussed in this chapter should therefore be considered together when designing AI-assisted work systems. Purpose limitation, data minimisation, proportionality, storage limitation, fairness, and transparency shape different parts of the same data architecture. This is particularly important in the workplace, where organisational interests and workers' expectations of privacy may not always align. Without clear limits on collection, retention, and reuse, the same data flows introduced for assistance may gradually support more intrusive forms of monitoring.

Legal principles alone, however, are not enough. Their practical effect depends on organisational arrangements that remain in place throughout the system lifecycle. Operational data should be distinguished from data used for training or further development; purposes should remain separated; retention periods should reflect actual need; access should be limited according to role; and these choices should be reviewed as the system evolves. The discussion in Section 3.8 shows that governance is credible only when such distinctions are maintained in practice.

Documentation supports this continuity. Data flows, access conditions, retention rules, and the boundaries between operational and secondary uses need to remain understandable as the system changes through maintenance, updates, adaptation, or retraining. An impact-based approach should therefore begin during design and

continue throughout operation rather than being treated as a one-time compliance exercise.

Data governance is thus part of the socio-technical architecture of an AI-assisted work system, not an administrative layer added after technical decisions have been made. A system may function effectively while still being difficult to justify if its data practices are unclear, excessive, or poorly controlled. Long-term viability depends on maintaining a proportionate relationship between the assistive purpose of the system, the data it uses, the access granted to different actors, and the duration of storage.

The broader regulatory framework supports this view. The GDPR continues to provide the main principles governing lawful personal-data processing, while the AI Act introduces an additional risk-based layer of AI governance within the European Union.

Chapter 4

Research design and analytical framework

This chapter presents the methodological structure of the thesis and explains how the comparative analysis is organised. The research design is based on the selection of three AI-assisted assembly systems that differ in terms of data type and AI paradigm, allowing the analysis to isolate relevant technical and governance trade-offs. The chapter also defines the engineering and legal dimensions used to reconstruct and compare the selected case studies.

4.1 Thesis rationale and comparative logic of case selection

This thesis adopts a structured comparative research design aimed at analysing how different configurations of AI-assisted assembly systems influence both technical performance and governance implications. The selection of the case studies is grounded in a deliberate analytical choice: to isolate and combine two key variables that are increasingly central in Industry 5.0 systems, namely, the type of data processed and the type of AI employed.

The first variable concerns the distinction between physiological data and non-physiological data. Physiological data include signals related to the operator's cognitive and physical state, such as heart rate, temperature, or behavioural indicators, while non-physiological data typically refer to environmental, visual, or task-related information. This distinction is particularly relevant in human-centric manufacturing, where cognitive load estimation and adaptive assistance are increasingly based on multimodal sensing strategies, particularly in cognitive workload estimation contexts [4].

Physiological data also create a more demanding design context because they are harder to integrate technically and may raise more sensitive legal questions under

data-protection law.

The other dimension used in the case selection is the type of AI. Conventional systems typically classify, predict, or apply predefined rules, whereas GenAI, particularly LLMs, can produce instructions, explanations, and other context-dependent outputs. Recent research suggests growing interest in these models for manufacturing applications, especially where human-machine interaction and adaptive support are involved [35].

The framework brings these two dimensions together to compare how data characteristics and AI paradigms affect system behaviour. The analysis is deliberately limited to three systems, each representing a different configuration, so that the main technical and governance trade-offs can be examined in greater depth. This makes it possible to preserve analytical depth while still comparing the main technical differences and structural trade-offs across the systems.

This selection approach is particularly useful for examining both the technical and governance dimensions that lie at the centre of the thesis.

4.2 The comparative matrix: data type and AI paradigm

The comparative design is operationalised through a conceptual matrix defined by the two variables introduced above. The matrix in Table 4.1 provides the structural backbone of the thesis, positioning each case study within a clearly defined analytical space.

The first system, System A, represents configurations in which assistance is based on environmental and task-related data, while leveraging GenAI to produce adaptive outputs. In the selected case, generative models are used to generate context-aware assembly instructions, supporting operators through flexible and personalised interaction.

The configuration of System B focuses on systems that estimate the operator's cognitive state through physiological and behavioural data, using non-generative AI techniques. More specifically, the system combines fuzzy inference-based data fusion with ML classification to estimate cognitive load and adapt the assistance provided.

The final configuration, System C, represents the most advanced and complex scenario, combining multimodal physiological data with generative or large-scale AI models. In the selected case, different input modalities are used to infer human intentions and support real-time HRC, bringing together perception, reasoning, and adaptive interaction.

The remaining configuration, namely non-physiological data combined with non-GenAI, is not examined here because it reflects a more conventional design space

and is less central to the specific tensions addressed in this thesis. The matrix thus functions as a comparative lens, enabling the systematic identification of differences and commonalities across systems.

Table 4.1: The comparative matrix.

System	Paper positioning	Technologies and architecture	Data collected, processing, storage
System A - A framework for digital assembly instructions as a step towards manufacturing inclusiveness [10]	Non-physiological data and GenAI	• Intelligent vision and object recognition with YOLOv7 • Error characterisation with matrix • NLP for contextual and semantic interpretation of instructions and error narratives • LLM (ChatGPT 4.0 through a custom Python interface) • T2S audio rendering • Visual support and multi-channel interface	The system processes images of components, tools, and operations, assembly instructions, error logs, context descriptions, operator name, date, time, location, workstation, and generated corrective instructions. Error logs are filled automatically and then fed into OpenAI ChatGPT 4.0 via a custom interface, so an external API.

System	Paper position- ing	Technologies and architecture	Data collected, pro- cessing, storage
System B - Inclu- sive manufactur- ing: A contribution to assembly pro- cesses with human- machine RL [8]	Physiological data and non-GenAI	• Wearable, nearable sensing units for physiology • Behavioural monitoring via non-intrusive camera tracking • Fuzzy inference system for cognitive load fusion • ML classifier for operator-state categorisation • Collaborative robot (OMRON TM5-900) • YOLOv7 for object detection • RL loops: HITL and MITL	The system collects physiological data (heart rate, respiratory rate, body temperature or blood pressure), behavioural video data, workplace parameters (noise, lighting, temperature, layout), and process/task variables (task difficulty, number of subtasks or items, execution time, workstation design). These data are normalised, fuzzified, fused into six cognitive load descriptors, and then classified into classes. It also uses camera images for object recognition and relabelling.

System	Paper position- ing	Technologies and architecture	Data collected, pro- cessing, storage
System C - A Multimodal Large Model to Enhance Robot Understand- ing of Human Inten- tions for Accurate Human Robot Col- laborative Manufac- turing [49]	Physiological data and GenAI	• Multimodal sens- ing: depth camera for hand gestures, laptop microphone for speech, Emotiv EPOX EEG head- set with Emotiv Pro software • YOLOv5 for vision- based gesture and object-related recog- nition • OpenAI Whisper for speech recognition, and OpenAI GPT API support for intent interpretation • Service layer for operator safety mon- itoring, natural lan- guage interaction, and energy monitor- ing • ROS-based execu- tion pipeline linking inferred commands to robot action	The system collects syn- chronised multimodal data from the same spa- tiotemporal context, mainly vision, audio, and EEG signals gener- ated while the operator issues commands. Vision data capture hand ges- tures are associated with five predefined command generations. A dataset of synchronised samples is used, with separate splits processed sepa- rately, and the resulting intermediate predictions are then fused through an attention mecha- nism. The fused output is transmitted to the robot control pipeline for execution. In the virtual layer, fused data support digital twin rep- resentation, simulation, and system optimisa- tion, and a service layer linked to monitoring and cloud-enabled interac- tion.

The overall methodological logic adopted in this thesis is summarised in Figure 4.1. The framework shows how the analysis moves from the initial research framing and case selection to the extraction of evidence from the selected scientific papers, the reconstruction of the three systems through a common analytical template, and the final comparative synthesis. In this way, the figure provides a visual overview of the analytical pathway that connects the case-selection logic introduced in this section with the structured case-study methodology developed in the following sections.

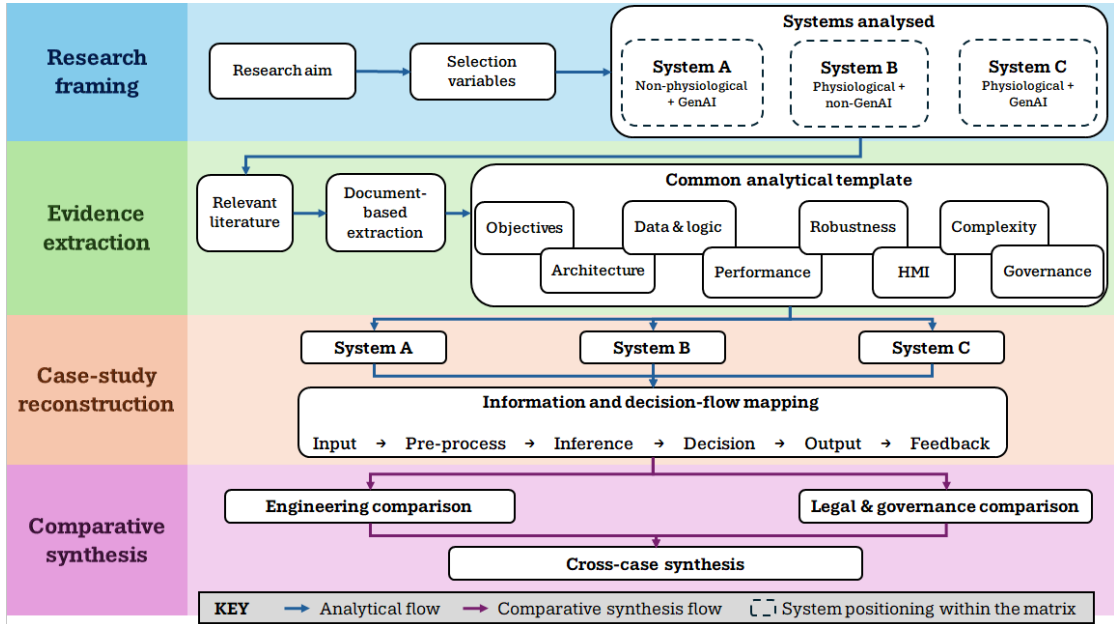


Figure 4.1: Methodological framework adopted for the comparative analysis of the three AI-assisted assembly systems.

4.3 Structured case-study methodology

As illustrated in Figure 4.1, the comparative analysis is based on a document-based extraction process applied uniformly to the three selected systems. Given the exploratory and comparative nature of the research, this thesis adopts a document-based comparative case-study methodology. Each system is analysed through a structured template, ensuring consistency and comparability across cases.

The methodology is based on the systematic examination of published scientific articles describing the selected systems. The analysis relies on reconstruction and interpretation of documented evidence. This approach is consistent with established research practices in human-centric manufacturing, where case studies are frequently used to analyse emerging technologies and design paradigms [3], [29]. Each case study is analysed according to a common structure, including:

- System objectives and target users.
- Architectural configuration.
- Data processing and decision logic
- Performance evaluation.

- Robustness and limitations.
- Human-machine interaction design.
- Implementation complexity.
- Data and governance implications.

This structured approach enables a controlled comparison, reducing interpretative variability and ensuring that differences observed across cases can be attributed to the variables under investigation.

4.4 Engineering analysis dimensions

To support a systematic and meaningful comparison, the thesis defines a set of engineering analysis dimensions, which serve as analytical lenses applied uniformly to all cases. These dimensions are not merely descriptive, rather, they are specifically selected to capture the main engineering trade-offs that emerge when combining different data types and AI paradigms. In this sense, they reflect key aspects of human-machine systems and are grounded in the literature on automation, human-centred design, and Industry 5.0.

More specifically, the analysis considers the overall system architecture, understood as the structural organisation of sensing, processing, and interaction components, together with the information and decision logic through which data are transformed into actionable outputs. Performance, measured by efficiency, error rate or reduced mental workload, and robustness, i.e., the system's capacity to function effectively in the face of unknowns and changes, are also key.

The HMI is examined through the communication channels it uses and its ergonomic suitability for different users. Technical complexity and industrial implementability are then assessed by considering integration effort, adaptability, and deployment under real operating conditions. These criteria provide a common basis for describing each system and for the cross-case comparison developed in Chapter 6.

4.5 Information-flow and decision-flow mapping

A central element of the analytical framework is the reconstruction of each system's information flow and decision flow, which allows a deeper understanding of how data are processed and translated into actions.

This step corresponds to the case-reconstruction layer of the methodological framework shown in Figure 4.1. Each system is decomposed into a sequence of functional stages:

1. Input acquisition, including environmental, behavioural, or physiological data.
2. Pre-processing and transformation, such as normalisation, filtering, or data fusion.
3. Inference mechanisms, where AI models interpret the processed data.
4. Decision points, where the system determines appropriate actions.
5. Output generation, including instructions, feedback, or robotic actions.
6. Feedback loops, enabling system adaptation and learning.

Figure 4.2 presents the generic information and decision flow used to reconstruct the three case studies. In System B, heterogeneous inputs are normalised and combined through a multi-layer fuzzy inference process, producing cognitive load indicators that are then classified with ML models [8]. The resulting decision logic combines predefined rules with learned patterns.

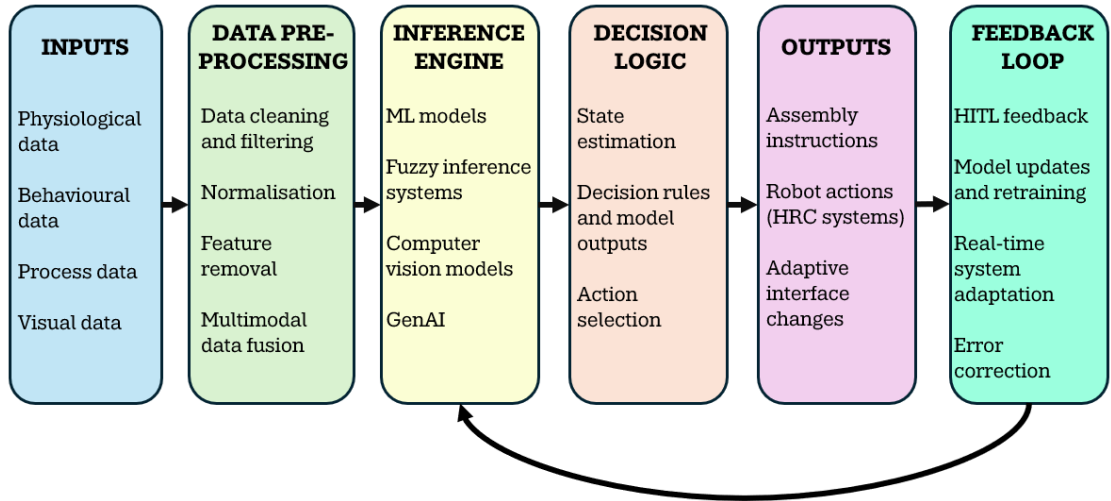


Figure 4.2: Generic information and decision flow in AI-assisted assembly systems.

By contrast, systems integrating GenAI (Systems A and C) introduce more dynamic and less deterministic decision processes, where outputs are generated rather than selected from predefined options. This mapping enables the identification of critical processing stages, potential failure points, and differences in system adaptability.

4.6 Comparative analysis criteria

Building on the analytical dimensions outlined above, the thesis further defines a set of comparative criteria that guide the cross-case analysis developed in Chapters 6

and 7. These criteria are intended to capture the most relevant aspects along which the selected systems can be meaningfully compared, both from an engineering and a governance perspective.

The analysis, in particular, looks at data usage intensity, i.e., how much and what variety of data each system needs, along with overall system complexity, i.e., the number of components involved and how tightly they must be integrated. It also focuses on robustness and reliability, especially in terms of performance under real-world variations, as well as scalability, i.e., the ability to be implemented in industrial settings. Besides that, the criteria of adaptability and personalisation, which are very important in human-centred systems that are meant to react to individual user traits and changed conditions, are included. The comparison also accounts for governance-related differences between the systems. Governance sensitivity captures how the type of data processed, especially physiological data, changes the need for safeguards and organisational control. Legal exposure addresses the compliance issues that may arise under the GDPR and the AI Act.

These criteria are included because legal and ethical concerns can shape technical design choices directly, particularly in workplace AI. In this respect, governance sensitivity and legal exposure will be further examined in Chapter 7, where the selected systems are analysed in relation to applicable data protection and AI regulation requirements.

4.7 Scope, assumptions, and limitations

The research design presented in this chapter has several limitations, which should be clearly stated.

Firstly, the analysis relies entirely on published academic sources and does not have direct access to the systems, data, or code. Therefore, the study is inevitably reliant on the correctness, completeness, and granularity of the published papers. Secondly, the thesis does not aim to reproduce the results of the experiments reported in the selected papers. Its goal is rather to reconstruct and compare the systems' architectures, design decisions, and consequences. Hence, the performance results validity reported in the articles is taken for granted and not questioned independently [8], [10], [11], [49]. Thirdly, while the choice of three case studies is justified, it unavoidably restricts the possibilities of generalising the results. However, the selected cases are distinct and relevant from the theoretical point of view, so it is possible to discern broader patterns and issues.

Despite these limitations, the adopted methodology still allows for a rigorous and systematic comparative analysis, which can produce interesting results regarding the interplay of data, AI, and governance in AI-augmented assembly lines.

Chapter 5

System-level case-study analysis

This chapter reconstructs the three selected AI-assisted assembly systems at system level. Each case study is analysed through a common template covering system purpose, architecture, information and decision flow, reported performance, robustness, human-machine interaction, implementation complexity, and data-governance profile. This structured reconstruction provides the empirical and analytical basis for the comparative chapters that follow.

5.1 System A: non-physiological data and GenAI

5.1.1 System purpose and target users

The system is best understood as an inclusive digital work-instruction framework for manual assembly, designed to reduce the interpretive burden imposed by conventional text-based instructions in industrial tasks. Its primary engineering purpose is not generic task automation, but the real-time detection of assembly deviations and the generation of corrective guidance in a format that is easier to process for operators who may struggle with written instructions.

In this sense, the system addresses two related challenges. The first is operational: identifying errors involving components, sequences, tools, or specific actions while the task is being performed. The second concerns accessibility: presenting procedural information through a communication channel that better accommodates different cognitive profiles, particularly users with reading difficulties.

From the perspective of the intended users, the paper places the solution within the broader context of inclusive manufacturing and identifies neurodiverse operators with reading difficulties as its main beneficiaries. However, the broader framework is presented as adaptable to several cognitive profiles, including memory, attention, language, writing, and logic difficulties through different channels and

interface choices. This broader claim should be considered cautiously, since the conceptual framework is wider, but the evaluated case study presented is much narrower and focuses on an operator with simulated dyslexia-related reading problems. Accordingly, the safest reconstruction is that the system is primarily a reading-accessibility-orientated assembly support system, with only prospective extension to other neurodiversity-related needs.

The context of use is a manual industrial assembly workstation, specifically a case study on the assembly of a horizontal bare-shaft centrifugal pump. The application domain is therefore not abstract human-computer interaction, but discrete-part mechanical assembly under sequence constraints, tool requirements, and object-specific operations. This matters analytically because the system’s usefulness depends on the existence of a sufficiently formalised process description, a stable bill of materials, and a visually observable execution environment.

It is evident that the approach is especially suitable for complex mechanical assembly of customised products, but the evidence remains tied to a bounded proof-of-concept rather than a full production deployment.

5.1.2 System architecture

Architecturally, the framework can be reconstructed as a modular perception-to-guidance pipeline organised into three main layers: diagnosis, instruction development, and instruction implementation.

The diagnosis layer is initiated when the operator starts a task. An intelligent vision subsystem monitors the work area and identifies components, tools, and operator actions to verify adherence to the predefined assembly sequence. When a deviation is detected, the system classifies the error across four dimensions, which are ‘component’, ‘sequence’, ‘tool’, and ‘operation’, and represents the result in a confusion-matrix-based diagnostic output. This diagnostic block is therefore the trigger condition for the rest of the assistance pipeline.

The instruction development layer combines structured process knowledge with language-processing modules. It is specified that assembly instructions are supplemented by a manually prepared context description containing the relevant tools, components, and operations. In parallel, the system uses customised error-log templates, including date, time, operator name, location, error description, error details, and corrective actions.

The intended logic is that diagnostic output from the vision module is transformed into a structured error log, which is then semantically enriched through NLP analysis before being passed, together with a prompt, to an LLM. The LLM then generates corrective instructions designed to be easier for users with dyslexia to follow and suitable for voice-based delivery. In the complete conceptual architecture, this stage includes syntactic parsing, semantic role labelling, and contextual

integration. However, the case study later clarifies that semantic analysis was not included in the implemented evaluation. This distinction is important, as the paper describes a broader reference architecture than the one actually tested.

At the output stage, the generated instruction is converted into speech through T2S and delivered to the operator via earphones, while additional visual support is shown on a display. The operator is therefore placed in a multimodal but guidance-centred loop rather than in a bidirectional conversational loop. The broader conceptual description also mentions speech-recognition-based interaction for operators with writing difficulties, but this functionality is not part of the validated proof-of-concept reported in the study. In implementation terms, it is explicitly reported that error logs were fed into OpenAI ChatGPT 4.0 through a custom interface developed in Python 3.10.7, while the object-recognition module was trained using YOLOv7 on Google Colab with an L4 GPU. This suggests a hybrid architecture combining local or edge-side sensing with at least partially external model services for generative processing, although the paper does not specify deployment topology in full detail.

A further architectural actor, although secondary, is the supervisor. The system includes an escalation rule under which three consecutive identical errors trigger an alarm for supervisor intervention. This means that the architecture already includes a fallback mechanism beyond full automation. From an engineering perspective, this is important because the system is not intended to act as a fully autonomous corrective agent, but rather as a monitored assistive layer with a defined recovery strategy. The possibility of rephrasing instructions after two repeated errors is mentioned only as a future enhancement and was not part of the tested system.

Figure 5.1 reconstructs the architecture of System A by distinguishing between the broader conceptual framework proposed by the authors and the modules actually implemented in the proof-of-concept case study. While the overall framework includes semantic analysis and visual support, shown as dashed components in the figure, these elements were not included in the reported evaluation. The implemented pipeline instead focuses on vision-based error detection, generation of an error log, LLM-based instruction generation, and voice delivery. The figure also includes supervisor escalation as a fallback mechanism, as explained before.

5.1.3 Informational and decision flow

The information flow begins with two input streams: a static process-description stream and a dynamic execution-monitoring stream. The static stream consists of the assembly instructions, which define the expected sequence, required components, and required tools for each step. The dynamic stream, instead, consists of the visual observation of the workstation by the intelligent vision system. The system compares observed execution against expected execution. Once a discrepancy

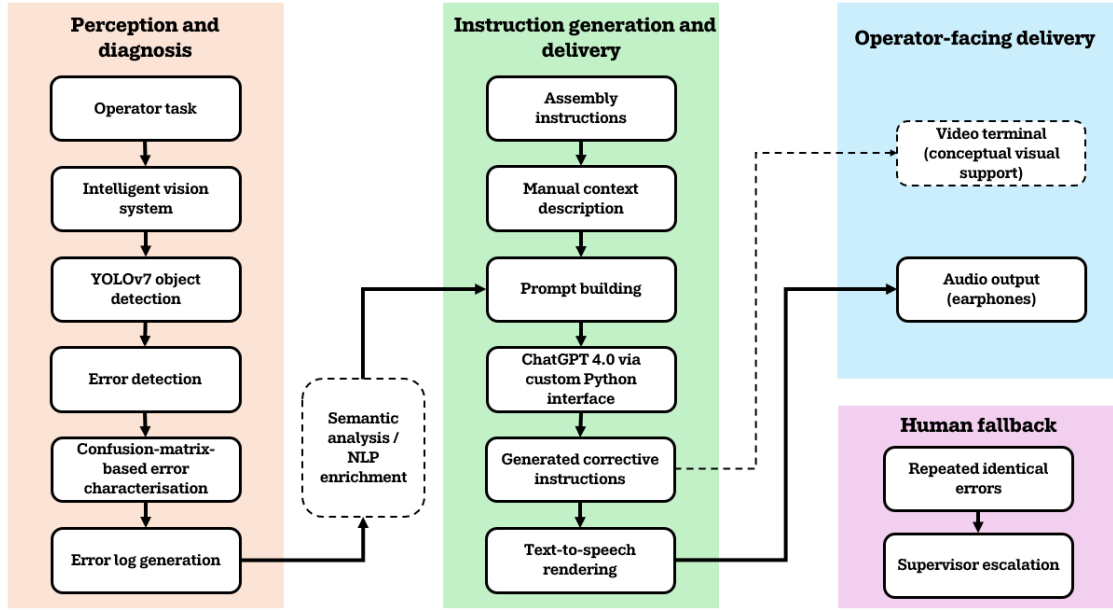


Figure 5.1: Reconstructed architecture of System A.

is detected, the error is categorised into one or more diagnostic classes, namely ‘component’, ‘sequence’, ‘tool’, or ‘operation’. In the conceptual framework, this classification is formalised through a confusion matrix that becomes the structured diagnostic representation of the event.

The next stage is pre-processing and structuring. Here, data from the assembly instructions and the diagnostic output are inserted into a standardised error-log template. It is specified that such logs include operational metadata such as date, time, operator name, and location, plus the description of the deviation and the associated corrective action. In the full conceptual model, this textual material is then processed by an NLP system that performs syntactic analysis and semantic role labelling in order to interpret the error narrative within the product-process context. This enriched representation is then combined with manually prepared contextual information and provided to the LLM through a prompt designed to generate simple, supportive, and dyslexia-friendly corrective instructions. These instructions are structured into numbered steps and include a final verification stage. The decision process therefore involves several successive transformations rather than a single classifier output.

First, the vision system determines whether the ongoing task deviates from the expected procedure. Secondly, the diagnostic module decides how that deviation should be characterised. Thirdly, the language-generation stage decides how corrective action should be verbally formulated for the operator. This is analytically

important because the system’s effective decision logic is distributed across heterogeneous modules rather than concentrated in one model. The final recommendation delivered to the operator is thus the product of a cascading sequence of perception, classification, templating, prompt construction, and generative reformulation. This architecture is powerful in terms of flexibility, but it also means that upstream recognition errors can propagate downstream into incorrectly framed corrective instructions.

In the implemented case study, the actual flow is more constrained than the full framework suggests. The proof-of-concept uses object recognition to detect two induced errors, then generates error logs, sends them to ChatGPT 4.0 via a custom interface, and produces voice-rendered instructions. Semantic analysis and image-based support were excluded ‘to simplify the scope’ [10], which means that the validated pipeline is effectively the one shown in Figure 5.2.

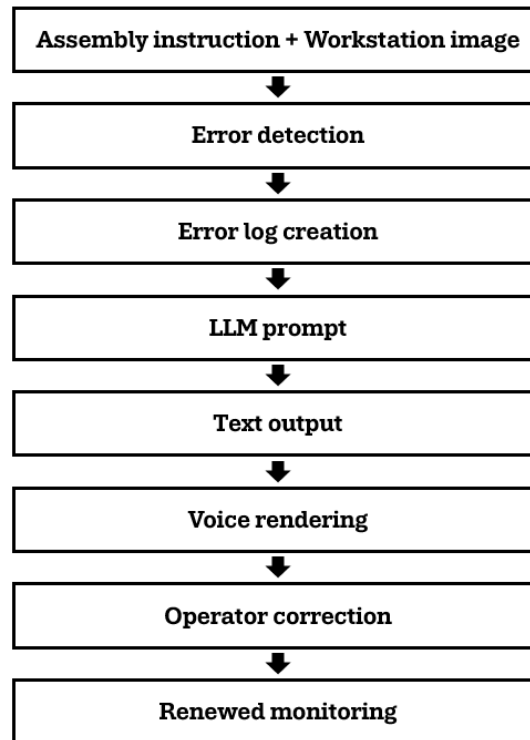


Figure 5.2: Validated pipeline of the case study of System A.

5.1.4 Reported performance and evaluation evidence

The evaluation evidence reported is limited, but still analytically useful if described with strict caution. The object-recognition subsystem was trained on only a limited

part of the overall assembly domain: 5 components, 2 tools, and 3 operations, compared with a full bill of materials containing 26 components and 7 tools. Training was carried out with YOLOv7 on Google Colab using an L4 GPU, a learning rate of 0.01, a batch size of 16, and more than 600 epochs, for a total training time of 2.3 hours.

The case study describes the training results as satisfactory, reporting correct classification for most categories and only minor confusion between the ‘Nut’ and ‘Washer’ classes because of their similar geometry. However, it does not provide standard evaluation metrics such as precision, recall, or a complete confusion matrix. Therefore, detector performance can only be discussed qualitatively, not benchmarked quantitatively.

The experimental evaluation is not a broad human-subject study, but a proof-of-concept timing comparison conducted with a single operator over 10 repetitions. The baseline condition simulates dyslexia-related reading difficulty through reading and understanding written instructions, while the alternative condition uses audio instructions generated by the digital system. Two errors were deliberately introduced during the experiment: a component-selection error at step S6 and a sequence error at step S9. Given that the study relies on a single participant, simulated reading difficulties, deliberately induced errors, and only two selected task steps, the results should be interpreted as exploratory and specific to the tested scenario. They cannot be readily generalised to neurodiverse populations or to industrial performance at scale.

Moreover, the reported numerical improvements relate specifically to how task time was distributed, rather than to broader measures of productivity or quality. For error E1 at step S6, instruction input decreased from 27.8 s to 14.8 s, error detection from 45 s to 2 s, and error correction from 33 s to 3 s, producing an overall saving of 94 s and a 61% reduction in step duration, as reported in Figure 5.3. For error E2 at step S9, instruction input decreased from 14.8 s to 9.8 s, error detection from 25 s to 2 s, and total step duration decreased by 36 s, corresponding to a 41% reduction, as reported in Figure 5.4. These figures should be reported exactly as time-accounting results within the experimental setup, not as universally comparable measures of assembly efficiency. The paper itself notes that human-operation values are averages, while digital-instruction times are fixed, which further limits strict comparability.

A further methodological constraint is that part of the conceptual system was not evaluated: semantic analysis and image-based support were excluded, and no direct measurement of cognitive load was performed. The system is likely to reduce the cognitive burden associated with task errors, but it remains a plausible hypothesis rather than a directly validated result and is left for future work.

Hence, this case study reports preliminary evidence of shorter recovery time under two error scenarios, but it does not yet provide validated evidence on workload,

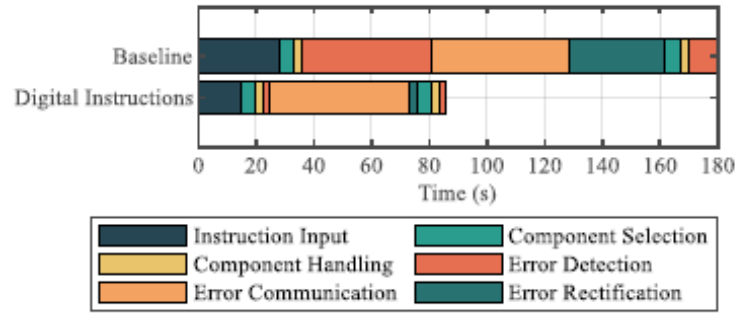


Figure 5.3: Time comparison for error scenario E1. Adapted from [10].

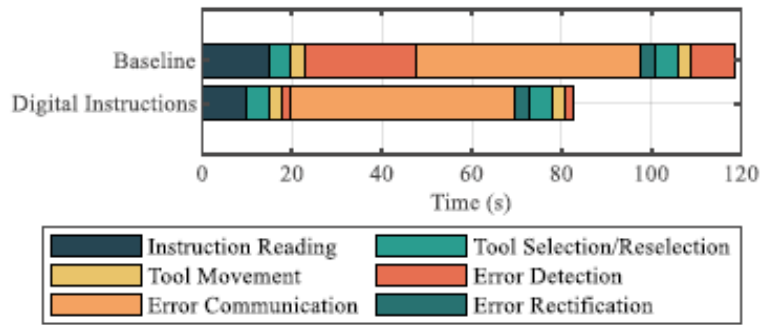


Figure 5.4: Time comparison for error scenario E2. Adapted from [10].

long-term learning, or error-rate reduction across a real operator population. A more detailed breakdown of the reported evaluation evidence, including the scope of the experimental setup and the main methodological limitations, is provided in Appendix A, Table A.1.

5.1.5 Robustness and technical limitations

An initial robustness issue concerns perceptual dependency. The vision subsystem is evaluated under favourable lighting and camera-placement conditions, while image augmentation is used to expand the training set instead of testing the recogniser across a wider range of shop floor environments. Its reported performance should therefore be read as evidence from a relatively controlled visual setting.

Deployment in a real assembly cell would introduce additional sources of variability. Changes in lighting, occlusion, reflections, workspace clutter, hand position, or component wear could all reduce recognition reliability compared with the proof-of-concept results. The limited training scope, covering only 5 components, 2 tools,

and 3 operations, also means that robustness has been demonstrated for only a small part of the full assembly domain. A second limitation concerns how errors can propagate through the architecture. The system links recognition, error characterisation, templating, prompt construction, and generative reformulation in sequence, so an early misclassification can affect every stage that follows. An incorrectly detected part may lead to an inaccurate error log, which may in turn prompt the LLM to generate an unsuitable corrective message. Because the resulting message may still sound plausible and fluent, the operator may also find it more difficult to recognise and question the underlying error. Systematic testing of these failure modes is not reported, and the behaviour of the generative layer under ambiguous or conflicting inputs is not discussed. The architecture is vulnerable to errors that propagate across modules. Because perception, diagnosis, language processing, and instruction generation are connected in sequence, an otherwise strong pipeline can still fail when one stage produces an incorrect or incomplete output. Average performance at module level is therefore less informative than the behaviour of the full chain under error propagation.

Transferability is constrained by the amount of contextual knowledge built into the system. The NLP layer relies on manually prepared descriptions and customised templates, which can support accuracy in a bounded setting but make adaptation to new products or processes less straightforward. A change in product family, workstation configuration, or process variant may require updates to process descriptions, prompt logic, expected error patterns, and the mapping between detector outputs and corrective instructions.

For this reason, the system is better understood as a configured assistance architecture than as a ready-to-deploy solution. Scaling it across industrial settings would require continued maintenance of these contextual resources, and that effort may eventually outweigh the cost of initial model training.

Finally, the robustness of the interaction channel itself remains underexplored. Voice output through earphones is assumed as an accessibility advantage, which is plausible for reading-related difficulties, but it does not test how the channel behaves under realistic industrial noise, time pressure, or competing attention demands. This omission is relevant because studies of speech interaction in industrial settings show that noisy acoustic conditions can reduce both system performance and user acceptance [50]. A 2024 study of speech interaction with HoloLens 2 reported a marked decline in performance under industrial noise and found ASR insufficiently reliable for that setting [36]. The study concerns speech input rather than audio guidance, so the comparison should be treated cautiously. Even so, it shows that voice-based interaction cannot be assumed to transfer reliably to industrial environments without dedicated acoustic testing.

5.1.6 Human-machine interface and interaction design

From an HMI perspective, the strongest feature of the system is that it converts assembly correction into a structured, low-literacy interaction format. The prompt explicitly instructs the LLM to greet the operator, explain the error in simple language, provide numbered correction steps, include a verification step, and maintain a supportive tone.

The resulting examples for E1 and E2 are operationally clear, sequential, and task-focused. The design is well aligned with the needs of the target users because it replaces dense written instructions with spoken, step-by-step guidance while still preserving the necessary procedural detail. Rather than simply informing the operator that an error has occurred, the system provides clear instructions on how to recover from it.

The interface is also designed around a multimodal concept. In the complete framework, audio guidance is combined with visual support on a terminal, while the broader design also considers elements such as icons, colour coding, process maps, simplified multilingual content, and combined visual and auditory cues to address different cognitive needs. This is a notable strength at the design level because it avoids treating ‘neurodiverse operators’ as a uniform category.

At the same time, it is practical to carefully evaluate the implemented HMI: the validated case study did not test the whole multimodal interface stack, and image-based support was explicitly excluded. What has been depicted is not the performance of a fully multimodal neurodiversity-adaptive interface, but the feasibility of one specific assistance mode, LLM-generated voice correction after detected errors.

In explainability terms, the system offers what may be called operational explainability rather than model explainability. It explains what the operator should do next and why the immediate error matters, but it does not explain how the vision model classified the state or how the LLM derived the wording. This distinction is relevant, because, for assembly support, operational explainability may be more valuable than technical transparency at the moment of action.

However, when misclassification occurs, the absence of deeper inspectability may complicate debugging, auditing, and trust calibration. The interface is therefore well aligned with operational task execution, but less aligned with stronger forms of AI transparency, interpretability, and traceability discussed in the explainability literature [30], [28], [38].

5.1.7 Technical complexity and industrial implementability

The system has moderate-to-high software complexity despite relatively ordinary sensing hardware. The core architecture combines computer vision for object and action recognition, structured process modelling, error logging, NLP-based text

processing, prompt engineering, LLM inference, T2S conversion, and multimodal delivery of the resulting guidance.

Even in the simplified form tested in the case study, the prototype already relies on YOLOv7 training, a custom Python interface, template-based logging, and integration with ChatGPT 4.0. By current AI standards, this is not an unusually complex architecture, but it is still considerably more demanding than conventional digital work-instruction systems based on static displays, pick-to-light solutions, or rule-based alerts.

Its practical implementation in industry therefore depends less on the novelty of the individual models than on the effort required to integrate and maintain the overall system. Deployment would require calibrated cameras, sufficiently stable workstation layouts, annotated visual data for the relevant components, tools, and operations, as well as process-specific context descriptions, error templates, prompt management, audio-delivery devices, and monitoring mechanisms. The claim that the system could support complex assembly of customised products is therefore plausible only when such customisation remains within a sufficiently controlled product and process structure. Without continued maintenance, operating costs could increase quickly.

The generative stage also introduces dependence on an external service. The paper reports that error logs were sent to OpenAI ChatGPT 4.0 through a custom Python interface, which suggests reliance on a remotely managed model unless a different enterprise arrangement was in place. The paper does not clarify this point.

Such a setup is convenient for prototyping, but production deployment would require closer attention to response time, service availability, operating cost, and data-governance implications. More broadly, recent manufacturing literature sees substantial promise in LLMs for instruction following, knowledge extraction, and manufacturing support, but also makes clear that practical value depends on domain adaptation and reliable integration into operational workflows [35]. In other words, LLM capability alone does not remove the engineering work of making the full stack dependable on the shop floor.

Overall, the architecture appears suitable for a targeted industrial pilot, but it is not yet a solution that could be scaled with little effort. The reported software stack appears technically manageable, and the use case is clearly bounded. Moving beyond the prototype would still require sustained effort, including data curation, product-specific reconfiguration, detector updates, interface validation, and monitoring of generated outputs.

The greater difficulty is likely to emerge after initial development, when the system has to be maintained, adapted to new conditions, and kept reliable over time.

5.1.8 Data and governance profile

The data profile of this system is more governance-sensitive than a simple ‘instruction assistant’ label might suggest. At minimum, the architecture processes visual data from the workstation, including components, tools, and operator actions, but it also processes document data from the assembly instructions.

Structured operational metadata such as date, time, location, and operator name in the error logs are collected too, followed by textual descriptions of deviations and corrective actions, with textual and audio outputs generated by LLM. A detailed reconstruction of the system’s data and governance profile is provided in Appendix A, Table A.2. The table identifies the main data items, their role within the system, and the points at which governance concerns may arise.

The first concern relates to identifiability. The examples of error logs explicitly include information such as the operator’s name, timestamp, assembly line, and workstation number. In addition, visual monitoring of the task area may capture information related to the operator, including posture, hand movements, or other contextual features, even when the primary purpose of the camera system is to recognise tools and components. Official EDPB guidance confirms that processing through video devices falls within GDPR analysis when personal data are involved, which is relevant for any workplace-facing vision system [20]. Another sensitive point is external processing. During the process, error logs were fed into OpenAI ChatGPT 4.0 via a custom interface. That means at least some task-relevant and operator-contextual information is transmitted beyond the local sensor layer, unless an undisclosed isolated deployment was used. The research does not specify retention periods, processor allocation, storage location, transfer regime, or whether the generated logs are archived. It also does not explain whether outputs are reviewable before delivery or whether prompt histories are stored. From a governance-mapping perspective, these omissions are not minor, since they mark exactly the areas that later legal analysis under GDPR and the AI Act would need to examine in more depth.

A final point concerns inference and its possible organisational effects. The system does not simply record what happens during the task; it also turns those events into error descriptions and corrective instructions linked to a specific operator session. In a workplace setting, repeated error logs could therefore become indirectly related to performance assessment. This possibility is partly reflected in the escalation mechanism, which involves a supervisor after the same error occurs repeatedly. Although the system does not rely on physiological data, it can still raise significant governance concerns when computer vision, operator-linked records, and GenAI-based assistance are combined.

5.1.9 Preliminary assessment

This case represents a strong example of a non-physiological-data and GenAI assembly assistance architecture. Its main technical strength lies in the closed-loop design that links task monitoring, error detection, structured logging, generative reformulation, and multimodal corrective delivery. Compared with conventional digital work-instruction systems, System A provides a more responsive form of support because it reacts when execution departs from the expected procedure instead of only presenting nominal steps. Its use of GenAI is also relatively constrained. The LLM operates on assembly instructions, error templates, and contextual information, so the generated guidance remains tied to a defined task setting rather than being produced from an open-ended prompt.

The evidence, however, is narrower than the overall framework might suggest. The architecture is presented as broadly inclusive, but the implemented evaluation covers only a limited proof-of-concept scenario. At this stage, the system is best interpreted as a promising prototype whose industrial validity still requires broader testing.

Its main trade-off concerns the relationship between adaptability and architectural dependence. Tailored support is made possible by combining several modules, but every additional stage introduces another point at which technical error, maintenance burden, or governance risk can emerge. This case is especially useful because it isolates a configuration in which the system does not rely on physiological sensing but does rely on GenAI. That makes it analytically valuable as a baseline for comparing how governance sensitivity changes when one moves from conventional operational data to physiological data, and from bounded instruction logic to more open generative layers.

5.2 System B: physiological data and non-GenAI

5.2.1 System purpose and target users

The system can be reconstructed as a HMC framework for inclusive assembly, whose primary purpose is not simply to automate task execution, but to detect cognitively stressful work states and adapt support strategies accordingly. More specifically, the framework aims to infer operator cognitive load from multimodal data, classify the resulting state into decision-relevant categories, and activate a suitable mitigation strategy, with RL as the support modality explored in the case study.

From an engineering perspective, the system addresses two closely related challenges. The first is to make cognitive strain observable in a way that is both non-intrusive and useful in practice. The second is to translate that assessment into adaptive

support during assembly tasks. The system therefore lies at the intersection of cognitive-state monitoring, decision support, and collaborative assistance in manual assembly.

From the perspective of the intended users, the framework is explicitly designed to support the inclusion of medically identified neurodiverse workers in industrial environments. The study adopts a broad neurodiversity perspective and refers to conditions such as autism spectrum disorder and dyslexia. Its fuzzy cognitive-load model also spans six domains: attention, memory, language, maths, logic, and reading. The empirical validation is more limited. Instead of involving a substantial group of neurodiverse operators, the case study reproduces memory-related difficulty by assigning warning values to the memory component and introducing delays intended to generate errors.

System B should therefore be read as a general framework for neurodiversity-sensitive assistance whose current evidence remains centred on a memory-focused proof of concept. The available results do not yet demonstrate equivalent performance across different neurodiverse profiles.

The framework is tested in a manual assembly setting shaped by strict task sequences, repetitive operations, time pressure, and error recovery. This places the evaluation within an industrial context where cognitive burden interacts with HRC and operational stress, rather than within a generic human-machine interaction experiment. The target operator is an assembly worker whose state and support needs are assessed using multimodal behavioural, physiological, workplace, and process data. In the case study, the operator interacts with a collaborative robot, object-recognition models, and a rule-based decision architecture. The system should therefore be understood not as a static assistive interface, but as a monitored and adaptive socio-technical system.

5.2.2 System architecture

Architecturally, the framework is organised as a layered monitoring-and-mitigation pipeline. The overall framework shown in the study begins with medical assessment, neurodiversity characterisation, and cognitive load thresholds, then combines online operator, workplace, and process data acquisition with a fuzzy-based cognitive load assessment block, an ML-based decision-making block, and a support-strategy selection layer that can lead to workspace design interventions, digital enhancement, or RL. The same diagram also distinguishes between considered methodological steps, the narrower case-study focus, and out-of-boundary steps, which is analytically important because it makes explicit that not every architectural element was validated experimentally.

The first major layer is multimodal sensing and data acquisition. The framework collects four categories of input: physiology, behaviour, workplace settings, and

process characteristics. According to the paper, physiology may include heart rate, respiratory rate, and body temperature. Behaviour is captured through non-intrusive camera tracking, while workplace data include noise, lighting, and temperature. Process data, instead, include assembly complexity, tool usage, workstation design, number of components, and information load. Physiological data collected during the case study were recorded in the field using commercially available smartwatches, while a camera streamed real-time activity to the HRC system, and a fixed adjustable light was used to stabilise image capture of industrial components.

The second layer consists of the fuzzy-based cognitive load assessment block. The input variables are pre-processed, normalised to a common range, and expressed through low, medium, and high fuzzy membership levels. A multi-layer inference process then combines them into six cognitive-load descriptors covering attention, memory, language, maths, logic, and reading. Expert knowledge from occupational health, neurodiversity, and process engineering informs this stage.

The output is not a single workload score. Instead, the model produces task-linked dynamic charts that show how the estimated cognitive demands change over time. Cognitive load is therefore represented through a set of structured descriptors derived from heterogeneous proxy variables rather than measured directly from one physiological signal.

These descriptors are subsequently used in the decision layer. Predefined thresholds, informed by medical assessment and expert knowledge, support the labelling and classification of the operator state into three categories: ‘Suitability’, ‘Medical Aid’, or ‘Technological Support’. The classifier is described as being pretrained on labelled cognitive-load data, with SVMs, neural networks, and decision trees mentioned as possible techniques. The paper does not clarify which model was actually implemented in the proof of concept, leaving the exact classifier underspecified.

When the outcome indicates a need for technological support, the framework can direct the response towards workplace redesign, digital enhancement, or RL-based assistance. The study concentrates on the latter within a HRC assembly scenario. The resulting RL architecture links object detection with the collaborative assembly process. Image acquisition, model training and testing, relabelling, and retraining form part of the perception pipeline, while task execution generates feedback for both the operator and the model. Detected errors may trigger corrective support, allow progression to the next task, or feed additional relabelling and model refinement. Two coupled loops organise this interaction: ‘human in the loop’ improves machine perception, whereas ‘machine in the loop’ supports the operator during task execution. The physical hardware stack includes an OMRON TM5-900 collaborative robot, camera, control box, joystick, TMflow software, and operator workstation.

Figure 5.5 reconstructs the overall architecture of System B by distinguishing

between the broader framework proposed in the paper and the narrower subset of components effectively mobilised in the proof-of-concept study, the ones that are highlighted. While the framework includes six cognitive load dimensions and multiple support strategies, the reported experiment focuses mainly on memory-related cognitive load and on RL as the selected technological-support measure.

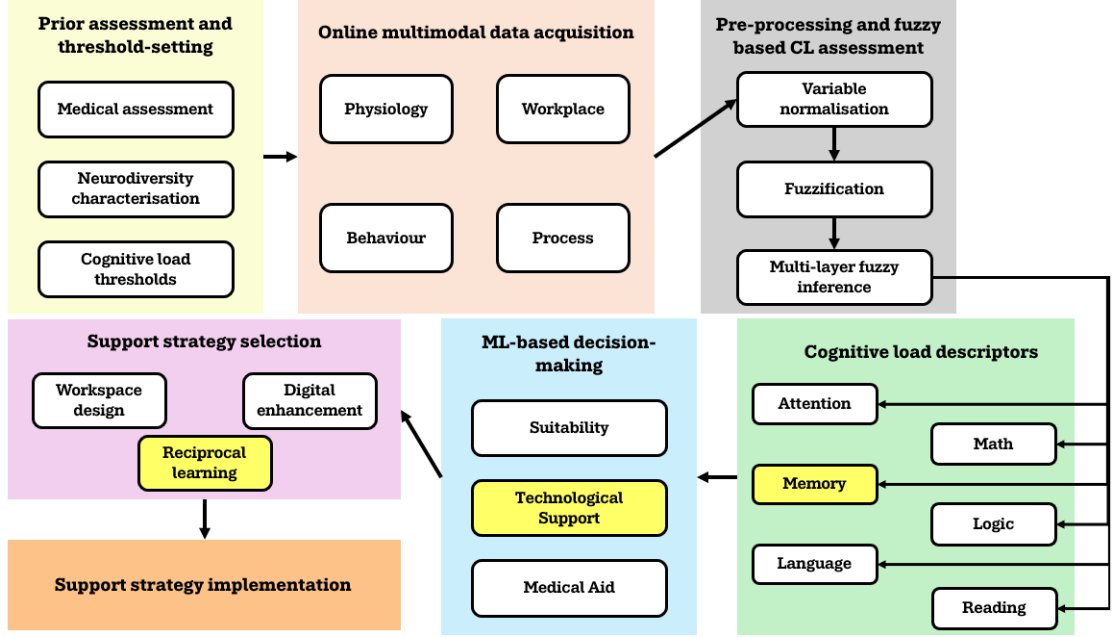


Figure 5.5: Reconstructed overall architecture of System B.

5.2.3 Informational and decision flow

The information flow begins with multimodal acquisition rather than with a single dominant sensor. Physiological signals, behavioural observations, workplace conditions, and process characteristics are collected in parallel and normalised into a common representation.

The first substantial transformation occurs when heterogeneous inputs are fused through the fuzzy-logic block. Membership functions and inference rules convert the available observations into six domain-specific cognitive-load descriptors. By this point, the system has already moved beyond raw-data handling, since physiological, behavioural, workplace, and process variables are being interpreted as cognitively meaningful states.

These descriptors feed a decision process that combines expert-defined thresholds with ML classification. After defuzzification, the outputs are evaluated against cognitive-load thresholds informed by medical assessment and expert knowledge,

while a pretrained classifier assigns the operator to one of three categories. The resulting logic mixes fuzzy inference, expert judgement, and learned classification. This gives the system more flexibility than a purely rule-based approach, but each layer also introduces assumptions that may influence the final outcome.

A ‘Technological Support’ outcome activates the assistance branch. Although the broader framework includes several possible responses, the case study focuses on RL in a human-robot collaborative assembly scenario. YOLOv7-based perception monitors task execution and detects object- or sequence-related deviations, providing input for corrective support. Operator feedback serves a different function: relabelling is used to update and retrain the object-recognition model.

The resulting interaction forms a feedback cycle. System outputs influence task execution, operator corrections modify the perception model, and the updated model shapes later assistance.

The pipeline evaluated in the case study is, however, more limited than the complete framework. Behavioural and workplace data were forced to non-warning values, and the analysis focused only on the memory-related component of cognitive load, as it is represented in Figure 5.6. The two RL loops embedded in the case study are made explicit: a MITL supporting task execution through object detection and correction, and a HITL improving model performance through relabelling and retraining. This narrower reconstruction is important because otherwise the broader conceptual model might be mistaken for a fully validated end-to-end system.

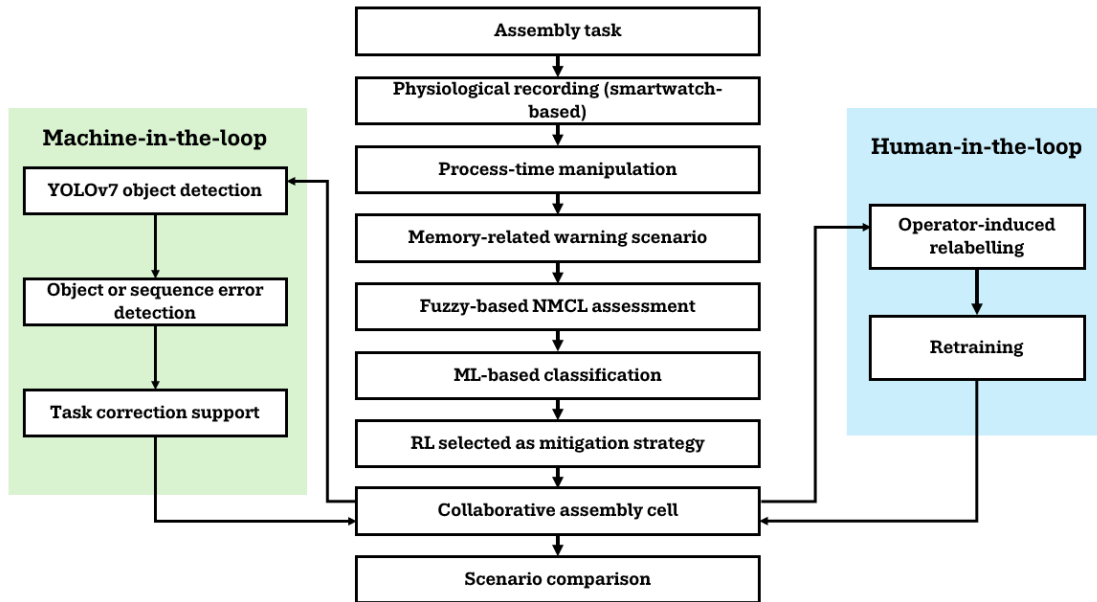


Figure 5.6: Validated pipeline of the case study of System B, highlighting the MITL and HITL RL loops.

5.2.4 Reported performance and evaluation evidence

The reported evidence is more substantial than in System A in one respect, namely that the study reports both cognitive load dynamics and object-detection results. However, it remains a tightly bounded proof-of-concept.

The experimental case study evaluates three scenarios: a baseline without errors, an error scenario with manual correction, and an error scenario with RL-based correction. The assembly is divided into ten steps with reported nominal times, and the scenario definitions are made explicit: in the manual-correction case, the operator detects the error after 5 seconds into the next step, calls a supervisor, and undergoes a 10-second supervisor inspection. In the RL case, object detection takes 1 second, and robot-assisted tasks include a 3-second robot handling time. This gives a relevant transparent temporal decomposition, but still within a small, highly staged experiment.

The paper reports cognitive load outcomes through the NMCL chart, which compares baseline, manual check, and RL scenarios. Scenario (b), the one with manual check, shows a significant increase in NMCL after error encounter and during correction, whereas scenario (c), the one with RL, exhibits a less pronounced increase and a quicker return to baseline, which is interpreted as evidence that RL reduces the cognitive disruption associated with error handling. Even though this is a meaningful result, the study does not provide inferential statistics, confidence intervals, or a broader multi-participant evaluation. The results point to a reduction in memory-related cognitive strain under the RL condition, but the evidence is too limited to support a broader claim about cognitive-load reduction across neurodiverse workers.

The timing results require similar caution. Scenario (c) is completed more quickly than the baseline, and the study links this improvement to the combined use of RL and collaborative robotics. Robot-assisted object selection could reasonably shorten parts of the assembly process, but the compared scenarios are not fully equivalent. In Phase 2, the RL condition includes robot handling and applies different definitions of step duration.

The reported times are therefore informative for the prototype as configured, but they should not be generalised into evidence of productivity gains over conventional assembly. What the experiment does show is that, under the stated assumptions, the RL architecture performs better than the manually corrected error scenario and can also achieve a shorter completion time than the nominal baseline.

The most concrete numerical evidence concerns YOLOv7 object recognition. For Phase 1, most components achieved 100% recognition, but gear unit #40 showed a 9% error rate due to confusion with gear unit #45, with an overall confidence level of 0.837. Phase 2 training achieved 100% recognition and 0.985 confidence. It is reported a HITL improvement loop too: after operator-induced relabelling

and 6 epochs of retraining, taking about 2.5 minutes, the classification error of gear unit #40 decreased from 9% to 3%. This is one of the strongest engineering contributions because, as it is reported in Figure 5.7, it shows a concrete RL effect on the perception subsystem rather than merely on operator experience. Even in this case, however, the evidence remains limited to a very specific setting: two visually similar gears, a small-scale experimental setup, and a controlled retraining process.

A final limitation is that the case study does not validate the full scope of the proposed framework. Although the overall architecture covers six cognitive load domains and several possible support strategies, the empirical analysis focuses only on memory-related difficulties. It also relies on simulated warning conditions, keeps some data channels effectively constant, and does not evaluate the system with a large sample of medically identified neurodiverse participants.

Accordingly, the system provides promising but preliminary evidence for the feasibility of RL as a cognitively supportive assembly strategy, rather than generalisable proof of robust neurodiversity-adaptive deployment. A more detailed breakdown of the reported evaluation evidence, including the scope of the case study and its main methodological limitations, is provided in Appendix A, Table A.3.

5.2.5 Robustness and technical limitations

The first technical limitation concerns the validity of cognitive load inference itself. The framework does not directly observe cognitive load, instead, it transforms a heterogeneous set of physiological, behavioural, environmental, and process variables into six cognitive load descriptors through normalisation, fuzzification, hierarchical inference, and threshold-based interpretation. This creates a chain of dependence between the different stages of inference, meaning that noise or uncertainty in an upstream signal can affect the final representation of cognitive load.

In practice, a high memory-load estimate or a recommendation for support may be based on signals that only approximate the operator's underlying cognitive state. This use of proxy variables is common in industrial cognitive-load assessment [33], but it places an additional burden on validation. Robustness depends on more than reliable sensing: the selected proxies must also correspond closely enough to the cognitive states the system is intended to represent.

A second limitation is the dependence on expert-defined levels and rules. The fuzzy logic process uses membership classes and inference rules grounded in expert knowledge, while the subsequent categorisation into suitability, technological support, or medical aid relies on predetermined thresholds established through medical assessments and expert guidelines. The system relies on thresholds and decision rules that embed assumptions about how observed signals relate to cognitive states. Values that are appropriate for one setting or operator profile may not generalise to

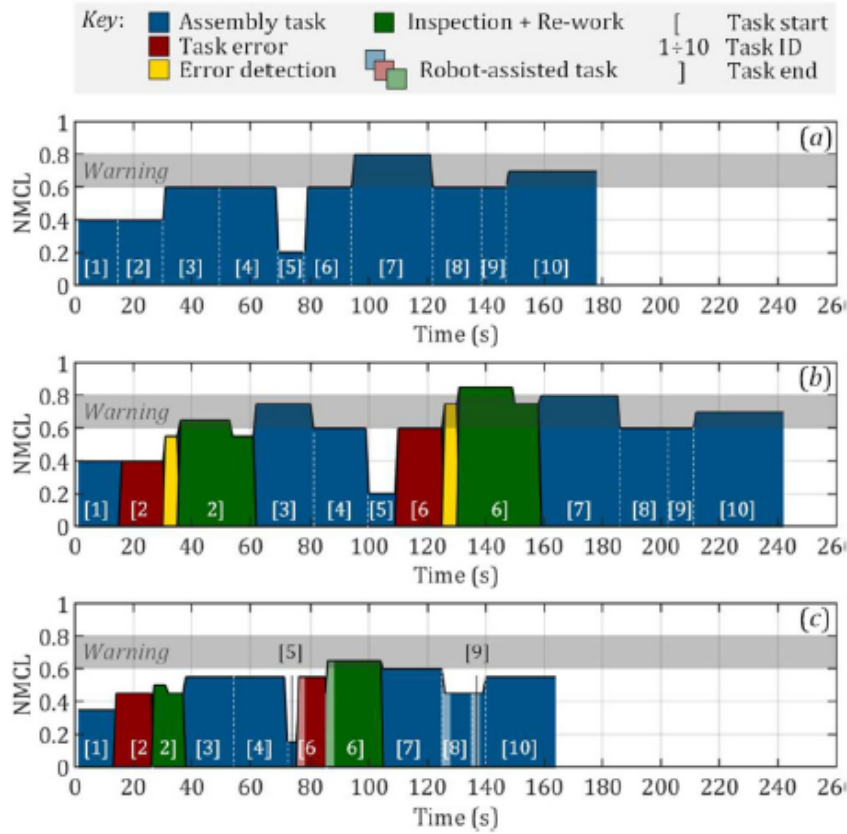


Figure 5.7: NMCL and the three scenarios' evidence. Adapted from [8].

another. Measurement error is therefore only part of the concern; the interpretation of otherwise valid inputs may also lead to inappropriate classification. An operator, for example, could be assigned to 'Technological Support' or 'Medical Aid' because the underlying rules do not adequately capture individual variation or the specific conditions of use.

A third limitation concerns the robustness of the object-detection subsystem under realistic industrial conditions. The case study uses YOLOv7 in a controlled environment with adjustable lighting, a fixed camera setup, and a limited variety of parts. Even under these conditions, the model initially confuses gear unit #40 with gear unit #45, resulting in a 9% error rate for one component class. Although HITL relabelling reduces this error to 3%, the need for human correction still points to an underlying weakness in the recognition process. Performance would likely be harder to maintain in less controlled shop-floor conditions, where occlusion, dirt, reflections, motion blur, changes in object position, and a broader variety of parts can affect recognition. The experiment shows that the system can respond adaptively to a

specific perception failure, but it does not establish robust generalisation across uncontrolled industrial environments.

The scope of the case study is also environmentally and contextually restricted. Behavioural and workplace variables are kept at non-warning levels, while the simulated difficulty is limited to the memory domain. This choice helps isolate the mechanism being tested, although it leaves much of the framework’s intended multimodal complexity outside the experiment. The proposed architecture is designed to combine several potential sources of cognitive strain, whereas the validation exercises only a narrower subset of that capability. The study therefore demonstrates feasibility in a controlled scenario rather than robustness under the full variability of a real multimodal industrial environment.

A fifth limitation concerns the risk of excessive intervention and reduced operator autonomy. Because the system continuously interprets the operator’s state and selects possible support strategies, assistance could gradually become overly directive. The HITL branch partly reduces this risk within the RL process, but the broader framework still assumes that the system can interpret the operator’s state reliably enough to decide when intervention is appropriate. From an engineering perspective, an overly sensitive system could issue too many corrections or warnings and limit operator discretion, particularly when temporary fluctuations are mistaken for persistent cognitive strain.

This limitation is not directly measured in the study, but it follows from the architecture’s decision logic and should be acknowledged.

5.2.6 Human-machine interface and interaction design

In HMI terms, this system differs substantially from System A because its main interface is not a single operator-facing instruction channel, but a broader ecology of sensing, classification, and collaborative assistance.

The HMI is spread across several elements of the system, including wearable or nearable sensing, visual monitoring, collaborative-robot support, and potentially audio or visual communication. The paper notes that human-machine communication can range from standard interfaces to more tailored solutions, such as speech synthesis, audio instructions, and visual aids, depending on the specific neurodiversity-related challenge being addressed. This gives the framework considerable flexibility, since inclusive support is not tied to a single interaction modality.

However, the implemented case study reflects only part of this broader multimodal vision. The proof of concept focuses on RL within a human-robot collaborative assembly cell, where the robot supports object handling and the perception layer contributes to error detection and correction. In practice, the operator mainly experiences assistance through the task itself: robot support, sequence monitoring, and error-handling mechanisms are more prominent than any fully developed

dialogue-based interface.

A notable HMI feature is the RL structure. As shown in the study, the operator is not simply a passive recipient of machine decisions. Through relabelling and feedback, the human helps improve object recognition, while the system supports task sequencing and execution. This reflects a more human-centred approach to automation and avoids a purely top-down model of assistance.

In practice, however, the balance between support and additional burden remains uncertain. HITL correction can improve system performance, but it also requires the operator to contribute to system maintenance during or around the work process. The study presents this as synergistic learning, which is plausible, but does not separately assess the cognitive or organisational effort involved.

From an explainability perspective, the system makes its operational behaviour relatively easy to understand, while offering limited transparency into the internal reasoning of the underlying models. The operator can in principle see the consequences of intervention (i.e., error correction, robot assistance, and altered task flow), but there is not any reported interface mechanism that would explain why a certain cognitive load class was inferred or why a specific support strategy was selected. This is especially relevant because the architecture uses fuzzy rules, expert thresholds, and ML classification, yet the traceability of those transitions for end users is not addressed. As in the broader explainability literature, this suggests a gap between action-guiding support and deeper inspectability of the decision logic [37], [33], [31]. The latter matters for trust calibration, contestability, and debugging, especially when misclassification may lead to sensitive classes such as ‘Medical Aid’.

Overall, the interaction design is conceptually rich but empirically under-specified. It is strongest where it embeds collaboration into the support logic itself, especially through the paired HITL and MITL loops. It is weakest where it invokes multimodal neurodiversity-sensitive communication without directly validating those channels in the reported case study. The strongest claim, instead, is that the HMI design is coherent with human-centric and RL principles, but only partially evidenced at the interface level.

5.2.7 Technical complexity and industrial implementability

This system is technically more complex than System A because it combines two different forms of adaptive intelligence: cognitive-state inference and collaborative task support. The software stack combines multimodal sensing, data normalisation, fuzzy inference, threshold-based reasoning, ML classification, YOLOv7-based object detection, HITL retraining, and collaborative-robot coordination.

On the hardware side, the system relies on wearables or nearables for physiological data collection, a camera, controlled lighting, a collaborative robot, the necessary

control infrastructure, and programming tools such as TMflow. Even before considering the additional work required for industrial integration, this is already a relatively complex architecture. The architecture is better viewed as an integrated cyber-physical assistance and decision system than as an isolated AI component. Deployment would require several elements to remain aligned. Multimodal inputs need to be acquired with sufficient consistency, including at least the physiological and visual signals used by the framework. Process steps must also be formalised in enough detail to connect task conditions with cognitive-load descriptors and corresponding support responses. Beyond this, the organisation would need to maintain the object-detection models together with the fuzzy rules, calibration logic, and mappings between detected conditions and assistance actions.

Such an effort may remain manageable within a stable and well-defined family of assembly tasks. The burden is likely to grow in high-mix or frequently reconfigured environments, where both the process representation and the support logic would require repeated adaptation. The current framework therefore appears most realistic as a targeted implementation in a structured collaborative cell, while broader deployment would still involve substantial integration and maintenance work.

The case study nevertheless provides some encouraging indications regarding deployability. By current standards, YOLOv7 training requirements are relatively modest: 300 epochs, a batch size of 16, an image resolution of 640 x 640, approximately 2 hours of total training across the two datasets, and retraining after relabelling in about 2.5 minutes. Likewise, the cobot-assisted object-handling logic is technically plausible for industrial pilot deployments. Retraining may be relatively straightforward in a small gear-and-screw assembly, but the same result cannot be assumed for richer tasks involving more object classes, similar geometries, or less stable workspaces.

The larger deployment burden may come from validation and governance rather than computational requirements. Because the framework assigns operators to distinct outcome categories, industrial use would call for stronger evidence that those classifications remain reliable across users and operating conditions. Continued calibration, auditability, and human-factors validation would therefore become part of normal system maintenance. These requirements are especially relevant when physiological data contribute to decisions that may affect how an operator's status is interpreted within the organisation.

Scaling the architecture is likely to depend less on raw technical feasibility than on whether its classifications can remain reliable, explainable, and appropriate as both the system and its working environment change.

5.2.8 Data and governance profile

This system has a substantially higher governance sensitivity than System A because physiological data are integral to the architecture.

The framework processes heart rate, respiratory rate, body temperature, and blood pressure, alongside behavioural monitoring, workplace data, and process variables. It then transforms these data into cognitive load descriptors and further into operator-state classes such as ‘Suitability’, ‘Technological Support’, and ‘Medical Aid’. From a data-governance perspective, this is not merely an instruction-assistance system. It is a system for inferring sensitive states from multimodal monitoring and using those inferences to structure intervention. A detailed reconstruction of the system’s data and governance profile is provided in Appendix A, Table A.4, which identifies the main data categories, their technical role, and the first governance-sensitive points generated by physiological monitoring and operator-state classification.

Physiological monitoring creates one of the most sensitive aspects of the system. Although the intended purpose is inclusive assistance rather than medical diagnosis, the framework processes bodily signals and applies expert-informed thresholds that may be associated with fatigue, well-being, or a perceived need for support. The case therefore illustrates how physiological sensing can introduce legal and ethical questions directly into the technical architecture, even when the system is not presented as a medical tool.

Visual and behavioural monitoring raises a related set of issues. The broader framework includes camera-based observation for behaviour tracking and crisis prediction, while the case study streams real-time activity to the HRC system. Such monitoring does not need to rely on face-centred identification to generate detailed information about workers. Once linked with task steps and intervention logic, patterns of movement, speed, posture, hesitation, and error recovery may acquire a performance-related meaning. In this sense, EDPB guidance [20] on video devices is therefore directly relevant to workplace surveillance because the system exemplifies a shift from simple task monitoring to interpretive surveillance of operator state and behaviour [15].

Another sensitive point is inferential classification. The architecture does not merely store raw signals, it generates normative categories that can shape organisational treatment of the worker. This is technically efficient for intervention logic, but governance-sensitive because classification errors may have exclusionary or stigmatising effects. It is worth stressing that the main sensitivity lies not only in the input data, but also in the inferences and status assignments produced from them. A final point concerns data minimisation and the gap between the conceptual framework and its practical deployment. At a conceptual level, the framework covers a wide range of data, including behavioural, physiological, workplace, and

process-related information. However, the case study validates only a narrower configuration focused on memory-related difficulties, with some data channels kept under controlled conditions. Here the system’s legal and ethical exposure arises not only from what it does but also from how much it proposes to know about the operator in order to do it.

5.2.9 Preliminary assessment

This system is a strong example of a physiological data, non-GenAI-inclusive manufacturing architecture. Its main technical strength lies in the integration of multimodal cognitive-state inference with a RL assistance logic. Unlike simpler digital-support systems, it does not merely react to observable task errors. The system aims to identify and respond to operator support needs by combining physiological, behavioural, environmental, and process-related signals. Unlike systems focused mainly on monitoring, it also includes a collaborative adaptation mechanism in which both human and machine contribute to improvement through HITL and MITL loops. This makes it a particularly ambitious example of human-centred manufacturing support.

Its main limitation is the gap between the scope of the proposed architecture and the evidence provided by the case study. The conceptual scope of System B is broader than the evidence produced by the case study. The framework covers six cognitive-load domains, several mitigation options, and a possible route for supporting workers with medically identified neurodiverse conditions. The implemented experiment, by contrast, centres on a memory-related simulation using controlled data channels, a small assembly cell, and a limited set of object classes, without substantial testing involving neurodiverse participants. The most convincing results therefore apply to the specific RL configuration that was evaluated, rather than to the full neurodiversity-sensitive framework.

System B also illustrates a clear tension between adaptation and governance. Richer data allow the assistance logic to respond more closely to the inferred condition of the operator, but they also make the architecture more dependent on modelling assumptions, expand the number of possible failure points, and increase the sensitivity of data management. This distinguishes it from System A: GenAI is not used, while physiological monitoring and inference about the operator’s state play a much larger role. More broadly, the case shows that substantial technical and governance complexity can emerge even without generative models when cognitive-state classification becomes a central part of system operation.

5.3 System C: physiological data and GenAI

5.3.1 System purpose and target users

The system can be reconstructed as a multimodal intention-recognition framework for human-robot collaborative manufacturing, designed to improve robot understanding of human intentions in complex manufacturing environments. Its core engineering purpose is not simply to support task execution, but to infer operator commands more accurately by combining multiple synchronised modalities captured in the same spatiotemporal context. In that sense, the system addresses a specific coordination problem in HRC, where the robot must correctly interpret what the human intends to do or request, even when single channels such as gesture, speech, or EEG are individually noisy, ambiguous, or incomplete. The architecture therefore targets the translation of human multimodal behaviour into reliable robot-executable commands.

From a target-user perspective, the system is not framed explicitly as a neurodiversity+orientated or disability-orientated assistance platform. The system is presented as a HRC solution for manufacturing operators who share their workspace with collaborative robots. The target user is therefore an operator involved in collaborative tasks who communicates commands to the robot through multiple input modalities. This is important for the analysis because the paper focuses on how accurately the system can interpret human intentions and translate them into coordinated robot actions. The operator is not primarily treated as a recipient of assistive support, but as an active collaborator whose multimodal signals need to be combined into a reliable representation of the intended action.

The system is evaluated in a laboratory simulation of a manufacturing disassembly task involving a drone, a UR5e collaborative robot, hand gestures, speech, and EEG signals. The application is therefore more specific than a generic conversational human-robot interface. It concerns a shared manufacturing workspace in which operator commands must be interpreted and acted upon under real-time collaborative conditions. The main interest of the system lies in how it combines perception, reasoning, and action across several human-generated signals within the same manufacturing workflow, rather than relying on a single interaction modality.

5.3.2 System architecture

Architecturally, the framework is organised into four major layers: physical layer, multimodal large model layer, virtual layer, and service layer. This layered organisation clearly separates sensing, fusion, optimisation, and application services of the system. At the same time, it is important to note that the validation evidence concentrates far more heavily on the physical layer and large-model fusion

layer than on the virtual and service layers. In other words, the paper presents a relatively broad ecosystem architecture but experimentally validates only part of it in depth.

The physical layer contains the human operator, the collaborative robot, and the manufacturing resources and activities. It is the source of the raw multimodal information streams. The operator continuously provides multimodal information such as speech, gestures, EEG, and, in principle, other ‘bio-signals’ like EMG, while the cobot operates in a shared manufacturing environment equipped with its own sensing and actuation capacities. In the validated experiment, the implemented modalities are more specific, since there is a depth camera that captures gestures, a laptop microphone which records spoken commands, and an Emotiv EPOX EEG headset with Emotiv Pro software that captures EEG data. These signals are collected simultaneously in the same spatiotemporal context and passed to the next layer.

The multimodal large model layer is the inferential core of the architecture. It performs two main operations: single-modal processing and attention-based multimodal fusion. Vision is processed using YOLOv5 for gesture and object recognition, while audio is processed through OpenAI Whisper, then supplemented by keyword matching and OpenAI GPT API support for intent interpretation. EEG is classified using an SVM-based pipeline after pre-processing and feature extraction. The outputs of the three modalities are not treated as final decisions, but as intermediate predictions. These are then fused through an attention mechanism in which each modality is represented by a query, key, and value vectors, weighted dynamically through scaled dot-product attention, and optimised by backpropagation. This means that the system’s ‘large model’ character is hybrid rather than monolithic: it combines an LLM-enabled audio-intent interpretation stage with a dedicated multimodal attention-fusion mechanism rather than relying on one end-to-end general model.

The virtual layer uses fused output to support simulation, digital twin design, and system optimisation. Fused data can be used to create a real-time representation of the physical system, support prediction of future operations, identify potential issues, and optimise resource allocation and robot strategies. Other components of this layer are digital human, digital HRC, and digital robot models. However, these virtual-layer functions are described more conceptually than experimentally validated, indeed the case study does not provide the same depth of empirical evidence for the digital twin uses as it does for intention-classification accuracy.

The service layer lies at the highest functional level and includes operator safety monitoring, human-robot natural language interaction, and resource and energy monitoring and optimisation, with cloud-service integration. This is architecturally significant because it suggests that the fused intention output is not the end of the pipeline, but a building block for broader manufacturing services. Yet, from an

analytical perspective, this is also where the gap between architectural ambition and experimental validation becomes most visible. In fact, the service layer is present in the framework, but the reported experiment remains focused on the intention-recognition and command-execution chain.

Figure 5.8 reconstructs the overall architecture of System C by distinguishing between the broad four-layer framework proposed in the paper and the narrower subset of components more directly supported by the reported experiment. In particular, the empirical core of the study lies in synchronised multimodal sensing, single-modal prediction, attention-based fusion, and command execution, while the virtual and service layers, which are represented with dashed boxes, remain more broadly conceptualised.

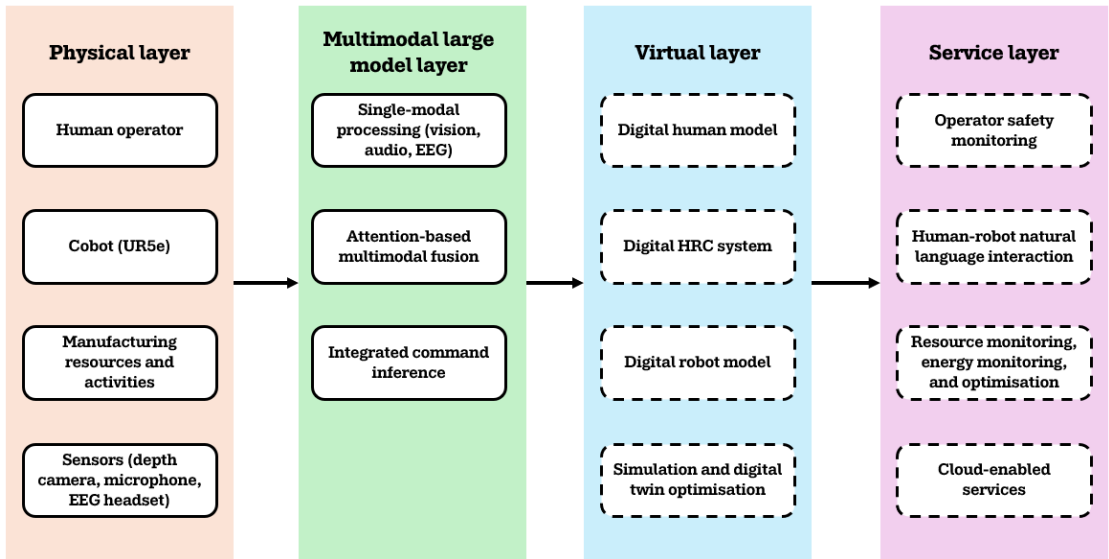


Figure 5.8: Reconstructed overall architecture of System C.

5.3.3 Informational and decision flow

The information flow begins with synchronised multimodal acquisition in the physical layer. Gesture, speech, and EEG are recorded simultaneously from the same operator in the same experimental context while commands are issued during the disassembly task. This synchronisation is not a marginal implementation detail: it is explicitly presented that this is a core methodological contribution, because many multimodal HRC studies suffer from weak spatial and temporal alignment between data streams. Here, all three modalities are intended to describe the same operator intention at the same moment in the shared manufacturing context.

The next stage is single-modal pre-processing and prediction. Each modality is

handled separately according to its own signal logic. Vision data are split into training and testing sets and processed through YOLOv5. Audio samples are processed with Whisper, then supplemented with keyword matching and GPT-based intent interpretation to improve recognition of predefined commands. EEG signals are pre-processed and classified through an SVM-based approach after the usual stages of signal acquisition, pre-processing, characteristic estimation, classification, and computer interaction. Crucially, the output of this stage is not yet the final command, but a set of intermediate modality-specific predictions.

The central inferential stage is the attention-based fusion mechanism. The modality-specific outputs are transformed into query, key and value vectors, while similarities are computed through scaled dot-product attention. Attention weights are normalised via softmax, a mathematical function used by AI to convert a vector into real numbers, and a fused context vector is built as a weighted sum of the value vectors. The model is trained by comparing the fused output against ground truth through a task-specific loss function and adjusting weights through backpropagation. Conceptually, this means that the system does not simply average modalities. It dynamically assigns different relevance weights depending on how informative each modality appears for the task at hand. This stage is therefore the decisive point at which raw multimodal heterogeneity is converted into a single inferred operator command.

Once the fused intention is generated, the system moves to the execution stage. The inferred command is transmitted to the ROS-based control pipeline, which exchanges data with the robot software and triggers cobot action. In the demonstration example reported in the paper, the fused command ‘remove the object’ is executed by the UR5e cobot, which identifies the target object through the external vision system, grasps it, and relocates it outside the working zone. The virtual and service layers may additionally consume the fused output for optimisation and monitoring purposes, but the experimentally validated action path is primarily this perception-to-command-to-actuation chain.

Figure 5.9 isolates the validated intention-to-action pipeline of System C. In contrast with the broader architecture shown in Figure 5.8, the reported experiment focuses on synchronised multimodal input, single-modal prediction, attention-based fusion, integrated command inference, ROS-based transmission, and robot execution within a bounded five-command drone-disassembly scenario.

5.3.4 Reported performance and evaluation evidence

The evaluation is more structured than in System A and more explicitly multimodal than in System B, although it still remains a laboratory-scale proof of concept. The study uses a dataset of 1,000 synchronised samples, each combining vision, audio, and EEG data for one of five predefined commands. Different subsets are used

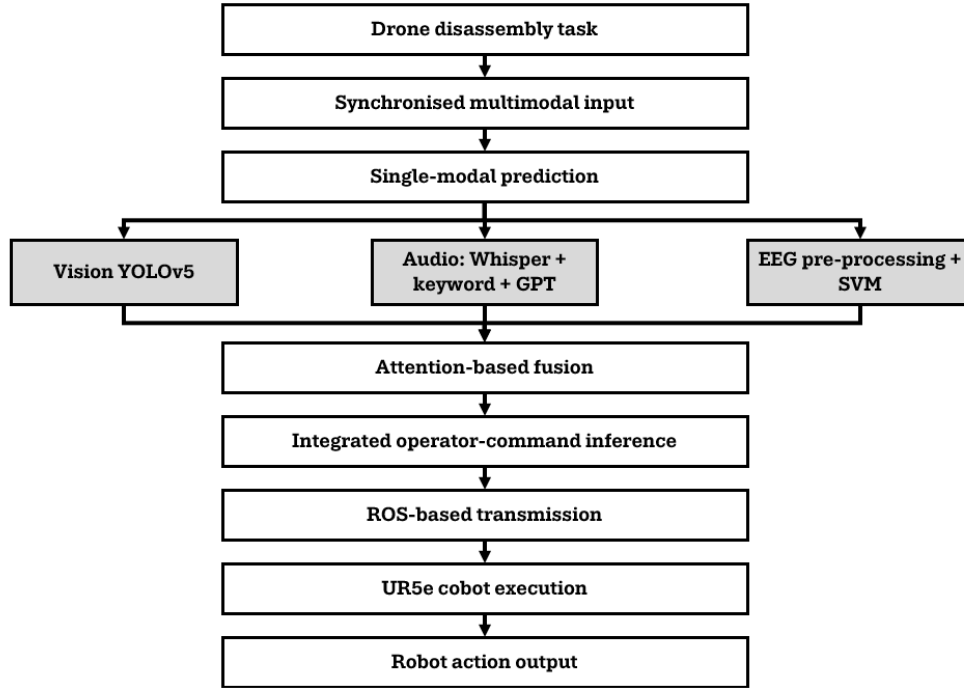


Figure 5.9: Validated intention-to-action pipeline of the case study of System C, from synchronised multimodal input to robot execution.

for the single-modality and fusion experiments: samples 1–500 for single-modality training, 501–700 for testing, 501–850 for fusion-model training, and 851–1000 for fusion-model testing. This provides a clearly defined experimental setup, but it also highlights the limited variety of commands and tasks included in the evaluation. The main quantitative result concerns the comparison between individual modalities and multimodal fusion. The fused model reaches an accuracy of 96.7%, compared with 88.8% for vision, 66.6% for audio, and 30.4% for EEG. These results suggest that combining multiple modalities can compensate for weaknesses in individual inputs and improve overall recognition performance. However, this conclusion should still be treated with caution. The evaluation covers only five predefined commands in a controlled disassembly scenario, while accuracy alone does not capture other important aspects such as latency, the consequences of incorrect predictions, or robustness under more variable industrial conditions.

The audio results are especially revealing. OpenAI Whisper initially performs below 40% accuracy because of mispronunciations by a non-native operator. After manual error correction, accuracy improves but remains insufficient. A keyword-matching system is then added to trigger predefined commands, and GPT support is introduced to merge transcribed speech with predefined command options and

infer intent more reliably. This is an important methodological distinction. The reported ‘audio modality’ is not based on a single speech-recognition model, but on a layered recovery process that combines Whisper, keyword matching, and GPT-assisted interpretation. The final audio accuracy should therefore not be interpreted as the performance of one standalone audio model.

The EEG results show a different pattern. Although EEG is presented as a potentially information-rich modality, its standalone accuracy is only 30.4%. The paper links this weak performance to the difficulty of collecting sufficiently large and stable datasets in the manufacturing setting, as well as to the sensitivity of EEG signals to fluctuations. This is analytically relevant because EEG is one of the main modalities in the proposed architecture, yet the reported evidence suggests that it is more useful as an additional input to multimodal fusion than as a reliable independent channel.

The demonstration of robot execution is also useful, but its evidentiary scope remains limited. The fused command can be transmitted through ROS and executed by the cobot in the drone-disassembly scenario, but the study does not report broader human-factors measures, user acceptance, workload outcomes, or longitudinal performance. The paper should therefore be viewed as a promising proof of concept for multimodal intention fusion in HRC, rather than as a comprehensive validation of a deployable system under real industrial conditions. A more detailed summary of the evaluation evidence, including the scope of the case study and its main methodological limitations, is provided in Appendix A, Table A.5.

5.3.5 Robustness and technical limitations

A first limitation concerns multimodal synchronisation dependency. The paper explicitly relies on vision, audio, and EEG being collected in the same spatiotemporal context. This is presented as a strength, but it also creates a fragility: if synchronisation is imperfect, the system may fuse signals that do not actually refer to the same intention instance. In such a case, the attention mechanism may assign high relevance to temporally misaligned or semantically inconsistent inputs, degrading command inference. The architecture therefore depends not only on model performance, but also on accurate and consistent acquisition of multimodal data.

A second limitation concerns the uneven reliability of the individual modalities, which may be partly hidden by the strong performance of the fused model. Vision is relatively robust, while audio remains sensitive to pronunciation and noise, and EEG performs poorly as a standalone predictor. The high overall fusion accuracy may therefore mask weaknesses in the underlying input channels. In deployment, if one modality degrades more strongly than in the laboratory, for example due to speech noise, occlusion, or poor EEG signal quality, the fused system may not

preserve the same performance. Put differently, the fusion result is promising, but it is not evidence that each modality is independently industrially reliable.

Another limitation concerns the audio pipeline under realistic industrial conditions. The paper itself shows that speech recognition alone is not sufficiently reliable and needs to be supported by manual correction, keyword matching, and GPT-assisted interpretation. This suggests that the audio channel still depends on a relatively limited command vocabulary, compatibility with different pronunciation patterns, and additional post-processing. This is an important limitation because, although the system may appear capable of natural-language interaction, its actual robustness still relies on a fairly constrained set of meanings and commands. This concern is also highlighted in other research, where industrial noise degrades speech-interaction performance and user acceptance [36], [51].

The EEG channel represents a limitation for the system too. The standalone EEG accuracy of 30.4% is extremely weak relative to the other modalities, and this is recognised to be a challenge with regard to collecting sufficiently domain-specific EEG data. This raises two engineering questions: first, whether EEG currently adds enough practical value to justify its burden. Second, the system's apparent multimodal sophistication partly depends on a modality that remains difficult to stabilise, scale, and use reliably in routine manufacturing settings. Multimodal fusion can reduce the impact of this weakness, but it does not eliminate the underlying fragility of the channel.

A final limitation concerns the narrow and controlled setting used for validation. The case study is limited to a simulated drone-disassembly scenario with five pre-defined commands, fixed sensing positions, and a relatively structured environment. This creates a generalisation risk, because the framework may work well for narrow command vocabularies and carefully bounded HRC tasks, but the evidence does not yet show how it behaves in richer, more variable, or less disciplined manufacturing environments. From a robustness perspective, the system is therefore validated under controlled feasibility conditions rather than real-world industrial variability.

5.3.6 Human-machine interface and interaction design

From a HMI perspective, the strongest feature of the system is that it does not depend on a single interaction channel, as represented in Figure 5.10.

Instead, it distributes human-robot communication across gesture, speech, and EEG, then uses a fusion mechanism to infer the operator's intended command. This is a meaningful design choice in principle because it recognises that human intention in collaborative manufacturing may not be reliably expressed through one modality alone.

At the same time, the implemented HMI is narrower than the broad architecture may initially suggest. The experiment constrains operator communication to five

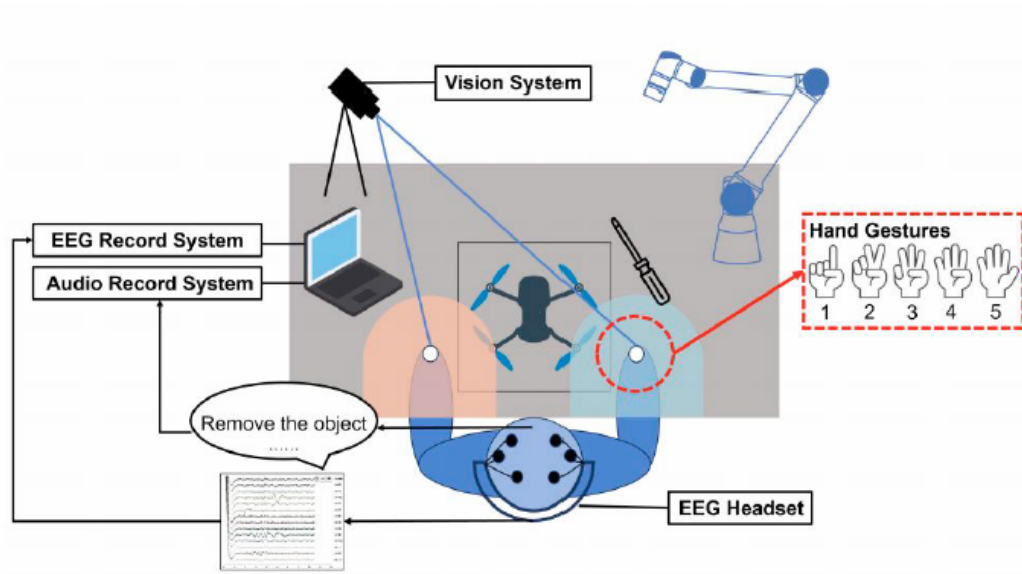


Figure 5.10: Diagram of the described setup. Adapted from [49].

predefined commands associated with corresponding hand gestures, and the robot executes those commands after multimodal interpretation. This means that the interface is multimodal in signal input, but still relatively narrow in expressive range. The apparent naturalness of the system should therefore not be overstated since it is not a fully open-ended human-robot interaction environment, but a bounded command-recognition system enhanced by multiple channels and GPT-assisted intent mapping.

GPT plays a practical role in the audio pathway by converting imperfect speech transcriptions into commands that can be matched to the robot’s predefined action space. Interaction quality still depends on whether noisy or incomplete transcripts are interpreted correctly. The main benefit is therefore increased tolerance to ambiguity in spoken input, rather than open-ended conversational capability.

At interface level, the robot’s executed actions are easy to observe and can be compared with the operator’s intended instruction. The reasoning behind those actions is much less visible. The system does not show why the attention mechanism prioritised one modality, how GPT resolved a particular command interpretation, or why one input signal overrode another. This is precisely the kind of gap discussed in the explainability literature and is one of the main issues that future developments of the system would need to address [28], [30], [38].

Overall, the HMI design is conceptually advanced because it treats human intention as multimodal and context dependent. Yet the experimentally validated interface remains a structured command interface rather than a fully general collaborative

language and intent interface. Therefore, HMI is multimodal and promising, but still bounded, partially opaque, and only preliminarily validated as a manufacturing interaction paradigm.

5.3.7 Technical complexity and industrial implementability

This is the most technically complex of the three systems, since it combines multimodal sensing, single-modal specialised models, an attention-based fusion mechanism, GPT-assisted audio intent interpretation, digital-twin integration, service-layer functionalities, and ROS-based robot execution. The architecture combines multimodal perception, state inference, robot control, and higher-level optimisation within the same system.

Deploying it in an industrial setting would require tight coordination across a large number of components. Depth cameras would need stable positioning, audio capture would have to remain usable under factory noise, and wearable EEG equipment would need to operate together with its supporting software. Each modality also relies on a dedicated processing pipeline before the data are combined through the attention-based fusion model. Beyond sensing and inference, the implementation includes ROS integration, cobot execution, digital-twin infrastructure, and cloud-based services associated with the virtual and service layers.

This level of integration may be achievable in an advanced pilot cell, but it creates a substantial burden for calibration, validation, and maintenance. Deployment would also remain sensitive to operating conditions such as camera stability, the restricted command vocabulary, operator acceptance of EEG wearables, and the level of environmental noise. For this reason, the system appears more plausible as a research demonstrator or advanced pilot than as an immediately scalable industrial product. This assessment is also consistent with Keshvarparast et al. (2024) study, which emphasises that collaborative-robot assembly systems often face significant deployment and integration burdens beyond their proof-of-concept promise [32].

A further complexity point is that the paper’s ‘large model’ architecture is actually a composition of heterogeneous AI techniques, not a single unified multimodal foundation model. That is not a criticism, but it matters analytically because lifecycle maintenance must cover multiple subsystems: YOLOv5, Whisper, GPT integration, EEG-SVM processing, attention-fusion training, ROS communication, and service-layer functions. The main implementation burden lies not in compute alone, but in the orchestration and governance of interacting subsystems.

5.3.8 Data and governance profile

This system has the highest governance sensitivity of the three cases, because it combines multimodal physiological and behavioural monitoring with GenAI-enabled interpretation and robot execution.

The system collects synchronised vision, audio, and EEG data from the same operator and processes each modality through a dedicated model. The resulting predictions are then combined through an attention mechanism to generate an integrated command that guides robot behaviour and also supports functions in the virtual and service layers. From a governance perspective, the architecture is therefore more than a sensing system or a GenAI application. It combines multimodal data to interpret both the operator’s state and intended actions.

A detailed reconstruction of the system’s data and governance profile is provided in Appendix A, Table A.6. The table identifies the main data categories, their technical role, and the points at which governance concerns may arise, particularly in relation to physiological monitoring and operator-state classification.

The first concern relates to the use of EEG data, which are highly sensitive and difficult to regard as low risk. Even when EEG is used for intention recognition rather than diagnosis, the recorded signals may contain neurophysiological information unrelated to the operator’s intended command. This makes EEG demanding from both an engineering and a governance perspective, particularly where the processing could support inferences about internal operator states.

The combination of modalities creates a separate concern. Hand gestures, speech, and EEG are captured within the same spatiotemporal context, producing a detailed representation of the operator during task execution. Although the stated objective is more effective collaboration, such an architecture can considerably expand the scope of workplace observation and narrow the distinction between assistance and intensive monitoring.

Part of the processing chain also depends on external services. The audio pathway uses OpenAI Whisper and GPT API support, while the service layer connects to cloud-based infrastructure. Depending on the implementation, operator-generated data or intermediate outputs may therefore leave the local processing environment. However, the paper does not clearly specify the related data flows, storage locations, retention periods, or allocation of processing responsibilities. In governance terms, this is a major omission rather than a minor implementation detail, because it directly affects transfer risk, storage limitation, access control, and purpose separation.

Finally, one last sensitive point is the actionability of inferred outputs. The system does not stop at classification or logging, instead, it converts fused multimodal inference into robot action. This greatly raises the stakes of classification error: in System B, a wrong classification could mislabel operator state, while here, a wrong

inference can directly trigger the wrong robot behaviour in a shared workspace. Governance sensitivity for System C arises not only from the intimacy of the data, but also from the immediacy with which inferred output becomes physical action.

5.3.9 Preliminary assessment

This case is a strong example of a physiological data and GenAI HRC architecture. Its main technical strength lies in the attempt to overcome the limits of single-modality HRC by fusing synchronised vision, audio, and EEG into a shared intention-recognition pipeline and linking the result to robot execution. Compared with System A, this architecture is more directly connected to robot action. Compared with System B, it relies more explicitly on large-scale AI models and a richer multimodal fusion process.

Its main limitation is that several sources of fragility are concentrated within the same system. The main sources of vulnerability include the need for precise synchronisation, the limited standalone performance of EEG, sensitivity in the audio pathway, the restricted command vocabulary, uncertainty around cloud and API processing, and the possibility that incorrect outputs directly influence robot behaviour.

Multimodal integration can improve performance by combining complementary signals, but it also enlarges the number of components that must remain reliable and understandable. Errors may emerge during sensing, fusion, interpretation, or action execution, while the use of richer operator data increases both governance sensitivity and the difficulty of tracing how a final decision was reached.

System C therefore gains adaptability and direct action capability at the cost of a more complex architecture, less transparent decision pathways, and greater sensitivity in the data it processes.

For the comparative analysis that follows, System C is particularly informative because it clearly shows how greater technical sophistication and human-robot flexibility can also bring higher data intensity, more demanding validation requirements, and increased governance sensitivity.

Chapter 6

Comparative engineering analysis

This chapter compares the three systems from an engineering perspective. The analysis focuses on architecture, sensing strategies, data intensity, decision logic, performance evidence, robustness, HMI, technical complexity, and industrial implementability. The aim is to identify the main engineering trade-offs that emerge when AI-assisted assembly systems become more adaptive, multimodal, and personalised.

6.1 Comparative architecture patterns

The three systems display three different architectural patterns. System A can be reconstructed as a ‘perception-to-guidance’ pipeline, System B as a ‘monitoring-to-mitigation’ architecture, and System C as a ‘multimodal intention-to-action’ architecture. This distinction is central because the systems do not merely differ in the technologies they use, but in the way they organise sensing, inference, feedback, and action [8], [10], [49].

System A is the most linear of the three. Its core architecture moves from vision-based monitoring to error diagnosis, from diagnosis to structured error logging, and from error logging to LLM-generated corrective instructions.

In the paper by Fan et al. (2025) [10], the diagnosis module detects deviations in components, tools, sequences, or operations, characterises the error through a confusion-matrix logic, and then feeds structured information into an instruction-generation process based on NLP, LLM support, and T2S rendering. The architecture is clearly modular, with separate components for vision, diagnostic reasoning, error-log generation, language generation, and audio delivery. However, these modules mainly operate in sequence. The system does not continuously adapt through operator feedback, as in System B, and it does not combine several intention signals at the same time, as in System C. Its feedback mechanism is more limited and corrective: once an error is identified, the system generates guidance intended to

bring the operator back to the expected procedure.

System A is relatively straightforward to interpret from an engineering perspective because each module performs a defined function. The generated output is also constrained by the assembly instructions, the recorded error state, and the prompt supplied to the LLM.

This modular clarity does not remove dependency across the pipeline. An incorrect workstation classification can distort the error log, which then provides the LLM with a faulty representation of the situation. The resulting instruction may still sound plausible despite being based on an upstream mistake. When delivered through voice, such guidance may be based on an upstream mistake. When delivered through voice, such guidance may also carry a sense of authority that is not matched by the reliability of the underlying information. In this sense, System A's architecture is modular but also fragile, because its reliability depends on the correct functioning of each stage in sequence.

System B has a more layered and adaptive architecture. In the paper by Simeone et al. (2024) [8], the system collects operator, workplace, and process data, normalises and fuzzifies them, produces cognitive load descriptors, classifies the operator state, and selects a mitigation strategy. Compared with System A, the architecture is less a single pipeline and more a closed-loop adaptive system. It contains two loops: one in which the machine supports the human by detecting and correcting task errors and another in which the human improves the machine by relabelling and retraining the object-recognition model.

This architectural pattern is technically richer than System A because it does not only respond to local task deviations, but it also attempts to infer the operator's cognitive condition and select an appropriate support strategy. The system therefore moves from task assistance to state-based adaptation. However, this richer architecture increases dependency on modelling assumptions. System B requires predefined cognitive load domains, expert-defined fuzzy rules, thresholds, and a classifier capable of assigning the operator to classes. The system is therefore more adaptive than System A, but also more dependent on the validity of its cognitive load model.

System C represents the most sophisticated architecture. In the framework proposed by Li et al. (2025) [49], the architecture is divided into four main layers: physical, multimodal large-model, virtual, and service. The physical layer collects vision, audio, and EEG signals from the operator. These inputs are first processed separately and then combined through an attention-based fusion mechanism. The resulting command is sent through ROS to the UR5e cobot, while the virtual and service layers support functions such as simulation, monitoring, natural-language interaction, and optimisation.

Compared with the other two systems, System C follows a less linear logic. It does not wait for an error to occur, as System A does, and its main purpose is

not workload estimation, as in System B. Instead, it attempts to infer operator intention directly from several signals collected at the same time and convert that interpretation into robot action. As a result, it covers a broader range of functions, from sensing and multimodal inference to command interpretation, robot control, and digital-twin-related optimisation. At the same time, it also has the tightest coupling requirements, because the three modalities must be collected in the same spatiotemporal context, processed through separate pipelines, fused correctly, and transmitted to the robot control system without ambiguity.

The comparison therefore reveals a first major engineering trade-off. System A is the most modular and easiest to decompose, but its intelligence is mostly corrective and pipeline-based. System B is more adaptive and reciprocal but depends heavily on cognitive-state modelling and fuzzy decision logic. System C is the most integrated and functionally ambitious, but its architecture is also the most sensitive to synchronisation, calibration, and subsystem coordination.

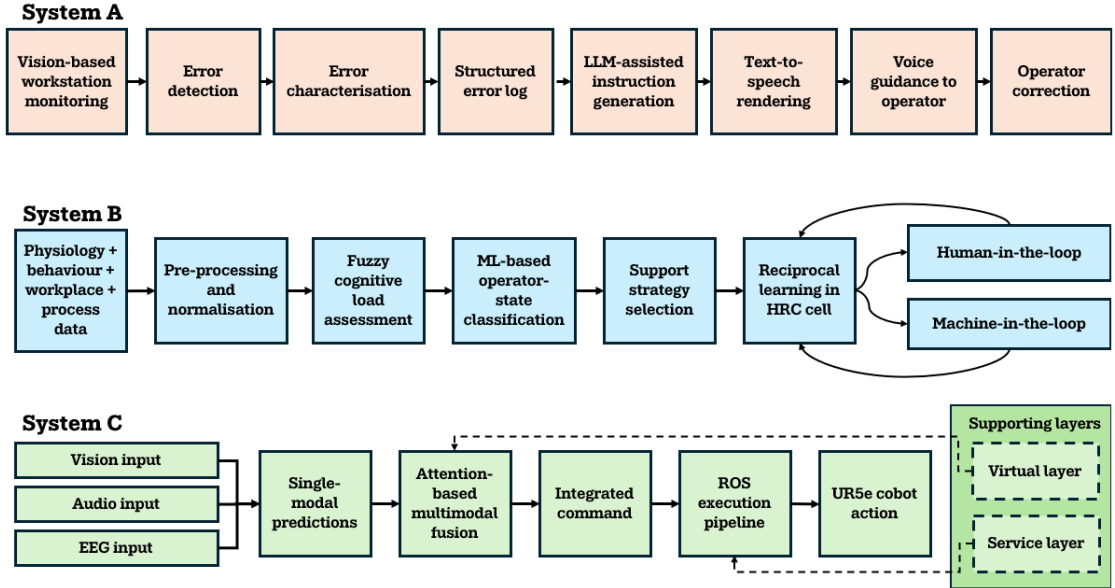


Figure 6.1: Comparative architecture patterns of the three systems.

6.2 Comparative sensing configurations and data intensity

The systems differ even more clearly when analysed through their sensing strategies. System A relies mainly on environmental and task-related data, System B introduces physiological, behavioural, workplace and process data, while System C combines

vision, audio and EEG in a synchronised multimodal command-recognition setting. This creates a progressive increase in data intensity from A to B to C.

System A has the lowest data intensity, although it should not be considered data-light in an absolute sense. Its main input is visual observation of the workstation, used to identify components, tools, and operator actions. It also uses assembly instructions, manually prepared context descriptions, structured error logs, operator name, time, location, and generated corrective instructions.

The system is therefore non-physiological, but is still based on workplace monitoring and operator-linked event data [10]. From a technical perspective, this makes its sensing strategy relatively manageable, because it requires a camera, stable lighting, identifiable components, and a formalised process sequence. It does not rely on physiological calibration or continuous estimates of the operator's cognitive state. System A is therefore less invasive than Systems B and C, although this does not make it free from monitoring concerns. A camera observing assembly activities may still capture information about hand movements, mistakes, hesitation, or error events linked to a specific operator. These data are less sensitive than physiological signals, but they can still become relevant in the context of workplace monitoring. The engineering advantage is that the system can provide assistance without directly sensing the operator's body or cognitive state. However, its limitation is that it can only infer errors from observable task deviations, and it cannot assess whether an operator is overloaded, fatigued, uncertain, or in need of a different type of support unless such states manifest through visible procedural errors.

System B moves to a more intensive sensing model, since the framework explicitly uses real-time acquisition of operator behaviour, physiology, workplace settings, and process characteristics [8]. Physiological variables include heart rate, respiratory rate, and body temperature, while workplace variables include noise, lighting, and temperature. Process variables, instead, include assembly complexity, tool usage, workstation design, number of components, and information load, performing behavioural monitoring through camera tracking. This sensing strategy is technically richer because it allows the system to infer cognitive load descriptors rather than merely detect procedural deviations, supporting a more personalised and state-sensitive form of assistance.

This richer data architecture also comes with practical costs. Physiological signals require sensors to be worn by the operator or positioned nearby, which can create additional demands in terms of integration, maintenance, calibration, and user acceptance, even when the technology is described as 'non-intrusive'. The signals themselves are also difficult to interpret in isolation. Heart rate, body temperature, or respiratory rate may vary because of task difficulty, physical effort, stress, heat, fatigue, or individual differences. Therefore, the system does not use these signals directly, it normalises, fuzzifies, and fuses them into cognitive-load descriptors during the process. The data intensity is therefore not only higher in quantity but also higher in interpretive complexity.

System C has the highest sensing intensity. The experiment integrates three external modalities: a depth camera for hand gestures, a laptop microphone for audio commands, and an Emotiv EPOX EEG headset with Emotiv Pro software for

brain-signal acquisition. These signals are collected at the same time and within the same workspace, then used to infer one of five possible operator commands in a simulated drone-disassembly task. The system therefore brings together behavioural, vocal, and neurophysiological information. Compared with System B, it relies on a narrower physiological data channel, but one that is more demanding to acquire and manage, since EEG monitoring is generally more sensitive to recording conditions and more cumbersome in practice than smartwatch-based sensing.

The sensing strategy of System C is attractive because it addresses a real limitation of single-channel HRC interaction: gestures may be occluded, speech may be noisy, and EEG may be unstable. Multimodal fusion can compensate for these weaknesses, but this advantage depends on successful synchronisation. If the operator's gesture, spoken command, and EEG signal are not aligned, the fusion model may combine signals that do not correspond to the same intention. Moreover, EEG introduces a major industrial implementability issue: a headset may be acceptable in a laboratory demonstrator, but it is much less obviously compatible with ordinary shop floor work, PPE, comfort requirements, and long-duration use.

The comparative hierarchy is therefore clear, and Table 6.1 is helpful to compare the technical sensing configurations of the three systems. Specifically, System A is the least sensing-intensive configuration and the easiest to justify from an engineering proportionality perspective, but it provides limited insight into operator state. System B is more personalised because it estimates cognitive load, but it requires continuous physiological and contextual data. System C is the most multimodal and potentially robust at intention recognition, but it is also the most invasive and operationally demanding because it combines audio, vision, and EEG.

Table 6.1: Comparative sensing architectures and engineering data intensity.

System	Sensing configuration	Main input stream	Processing function	Data intensity	Main technical burden
System A	Vision-based workstation monitoring combined with process-document inputs	Images of components, tools and operations; assembly instructions; context descriptions; error logs	Detection of component, tool, sequence, operation deviations and generation of corrective guidance	Medium	Stable visual setup, annotated components, visible mapping between detected errors and instruction templates

System	Sensing configuration	Main input stream	Processing function	Data intensity	Main technical burden
System B	Multimodal operator-state and process monitoring	Physiological signals; behavioural data; workplace variables; process variables; object-recognition images	Fuzzy cognitive-load assessment, ML-based operator-state classification, RL support	High	Sensor calibration, physiological variability, fuzzy-rule tuning, threshold definition, integration with cobot support
System C	Synchronized multimodal intention sensing	Depth-camera gestures; audio commands; EEG signals; intermediate predictions; fused command outputs	Human-intention recognition and translation into robot-executable commands	Very high	Synchronisation of modalities, EEG stability, speech robustness, attention-based fusion, safe robot-control integration

6.3 Comparative information and decision logic

The systems differ substantially in how processed data are converted into operational responses. System A uses visual and procedural information to produce corrective guidance when task execution deviates from the expected sequence. In System B, operator and process data are interpreted to estimate cognitive-load conditions and select a corresponding form of support. System C goes further towards direct action, combining synchronised multimodal inputs to infer commands for robot execution.

The distinction is therefore not only in the data being collected, but in what the decision layer is expected to do: generate guidance, interpret operator state, or translate inferred intention into physical action.

In System A, the decision flow begins with a comparison between observed execution and expected assembly sequence. The vision system identifies whether the operator has selected the wrong component, used the wrong tool, followed the wrong sequence, or performed an incorrect operation. The resulting diagnostic output is inserted into a structured error log and then passed to the LLM, which generates a corrective instruction. The paper makes clear that the LLM is not used as an unconstrained reasoning agent, since it receives assembly instructions, an error log, and a prompt requiring simple, dyslexic-friendly, numbered correction steps.

System A uses a hybrid decision process. Error detection begins with classification,

after which the identified problem is organised through a structured log template. The LLM then reformulates this information into guidance for the operator, combining the error description with the required corrective action. Compared with a purely rule-based instruction interface, this gives the system greater flexibility in how the final message is expressed and adapted to the user. At the same time, the actual decision about what is wrong remains dependent on the upstream diagnostic layer, given that the LLM does not independently verify the physical state, but it translates a structured diagnosis into accessible instruction.

System B has a different logic. It does not primarily generate language, but it estimates a state. Raw variables are first normalised and converted into fuzzy membership classes, then fuzzy inference produces cognitive load descriptors across attention, memory, language, math, logic, and reading. The resulting descriptors are evaluated against predefined thresholds and mapped to decision categories, giving the system a way to convert heterogeneous observations into an estimate of the operator's support needs.

Compared with a less constrained neural approach, this structure offers some degree of inspectability because the fuzzy rules and thresholds can, in principle, be reviewed. The complete decision path is still difficult to follow. Sensor acquisition, normalisation, fuzzification, rule-based inference, defuzzification, threshold evaluation, and ML classification all contribute to the final outcome. The logic is therefore structured, although the operator may have little visibility into how a specific classification was reached.

Reliability also depends heavily on the assumptions built into the cognitive-load model. Poorly calibrated rules, thresholds, or classifier behaviour could alter the inferred support need and lead the system towards an unsuitable intervention.

System C follows a different route by translating multimodal inputs into robot actions. Visual information is processed with YOLOv5, the audio pathway combines Whisper, keyword matching, and GPT-assisted intent interpretation, while EEG signals are handled through an SVM-based classification pipeline. The intermediate predictions are then fused using an attention mechanism, which assigns weights to the different modalities and produces a final command, which is transmitted to the robot through a ROS-based pipeline.

The decision logic of System C is therefore closer to intention inference than to error diagnosis or workload classification. The system does not ask, 'what error occurred?' as in System A, nor, 'what support does the operator need?' as in System B, but it asks 'what does the operator intend the robot to do?'. This is technically significant because the output is not only informational. It becomes physical action by the robot. A wrong inference in System A may produce a wrong instruction, while a wrong inference in System C may produce a wrong robot action.

The risk profile of the decision logic is therefore more immediate. GenAI plays

a different role in Systems A and C. In System A, the LLM is responsible for producing the final corrective message shown or delivered to the operator. In System C, GPT is used more narrowly within the audio pipeline, helping to interpret imperfect speech transcriptions and match them to a predefined command set. Its generative role is therefore more limited: it improves the handling of ambiguous or noisy language, but the overall system still operates within a constrained command-recognition framework rather than as an open-ended conversational robot. This distinction is important because both systems use LLM-related technology, but the location of the LLM in the decision chain is different.

Overall, System A is strongest in human-facing reformulation, System B in state-based support selection, and System C in multimodal command fusion. The systems therefore occupy different positions along the continuum from information support to adaptive decision support to robot action.

6.4 Comparative performance evidence

The performance evidence available for the three systems is heterogeneous, and it cannot be compared by simply placing the numerical results side by side because the systems measure different outcomes under different experimental conditions. The three systems are evaluated through different types of evidence, because System A reports reductions in error-correction time, while System B examines cognitive-load patterns, task efficiency, and improvements in object recognition through RL. System C focuses, instead, on command-classification accuracy, comparing individual modalities with multimodal fusion. These results are not directly comparable in absolute terms. The analysis should therefore consider the quality, scope, and industrial relevance of the evidence provided by each study [8], [10], [49]. Among the three, System A provides the narrowest evidence base, although its results are relatively easy to interpret from a practical perspective. The system is evaluated on two deliberately induced errors in a centrifugal-pump assembly task, using a single operator over ten repetitions. The reported results show substantial time savings for both errors' configurations, 94 s saved for E1, corresponding to a 61% reduction, and 36 s saved for E2, corresponding to a 41% reduction.

These results are relevant because they address a concrete production problem, which is error detection and recovery time. The evidence is still limited. The study does not include a broad sample of neurodiverse operators, does not directly measure cognitive load, and examines only two induced errors within a restricted part of the assembly process. In addition, semantic analysis and image-based support were not included in the implemented case study. The results for System A provide encouraging evidence that corrective guidance can support faster recovery under controlled conditions. They do not yet establish comparable effectiveness in industrial settings or across different operator groups.

System B is evaluated through a wider set of outcomes, including cognitive-load-related indicators and model performance. The case study examines three conditions: task execution without error, error followed by manual correction, and error followed by RL-based correction. The reported results indicate that RL reduces the NMCL linked to error handling and supports a faster return toward baseline levels. The study also evaluates object recognition and reports a marked reduction in classification errors. This evidence is stronger than System A in one respect: it demonstrates both task-level support and improvement of the machine perception subsystem. The RL loop is not only conceptual, but it also produces a measurable reduction in one classification error.

However, System B's evidence also remains bounded, since the neurodiversity-related condition is simulated, the cognitive-load analysis focuses on memory, and the broader framework covering six cognitive load dimensions is not fully validated. The case study therefore supports the feasibility of RL as a mitigation strategy, but it does not yet prove robust cognitive load adaptation in real industrial populations. System C provides the strongest algorithmic evidence but weaker human-factors evidence. In the experiment, the dataset includes 1,000 synchronised samples of vision, audio, and EEG associated with five predefined commands. The multimodal fused model reaches a higher accuracy of 96.7%, compared with the one for vision (88.8%), the one for audio (66.6%), and for EEG (30.4%). This is a clear quantitative result and directly supports the claim that multimodal fusion improves intention recognition compared with single modalities. The reported performance is obtained within a tightly bounded command space. The experiment uses five commands, a simulated drone-disassembly environment, controlled sensor positions, and a limited dataset, which constrains how far the results can be generalised. The evaluation also leaves several human- and deployment-related aspects unexplored, including workload, usability, operator trust, task completion time, error severity, latency, and long-term robustness.

Performance is uneven across the individual modalities. The audio pathway requires additional interpretation through keyword matching and GPT-assisted processing, while EEG shows limited effectiveness when used as a standalone input. The reported results therefore support the value of multimodal fusion for improving classification accuracy in the tested scenario, but they do not by themselves demonstrate industrial readiness.

Overall, the three systems provide different types of evidence. System A offers the clearest indication of time savings during error recovery, but its experimental basis is also the most limited. System B provides a more varied set of results by combining cognitive-load dynamics, task timing, and improvements in object detection, although the study remains controlled and partly simulated. System C reports the strongest classification accuracy, but gives much less attention to human-factors outcomes.

Table 6.2 summarises the comparative criteria for performance evidence, where systems are not ranked based on absolute performance, but are compared based on the type, scope, and industrial relevance of the evidence. The most industrially relevant evidence would require combining all three types of evaluation: time

and error recovery, workload and usability, and algorithmic reliability under real production variability.

Table 6.2: Comparative strength of performance evidence.

System	Main outcome	Evidence type	Strength of evidence	Main limitation	Industrial relevance
System A	Reduction of error-recovery and instruction-processing time in two induced-error scenarios	Proof-of-concept timing comparison	Moderate-low	One operator, ten repetitions, two induced errors, simulated reading difficulty, no direct cognitive-load measurement	Relevant for reducing recovery time in manual assembly, especially where reading-heavy instructions create barriers
System B	Reduction of memory-related cognitive load and improvement of object-recognition performance through RL	Cognitive-load and RL proof-of-concept	Moderate	Simulated memory-related condition, simplified case study, focus mainly on memory rather than all six cognitive-load domains	Relevant for adaptive support, cobot-assisted assembly, and human-machine co-learning
System C	Improvement of command-recognition accuracy through multimodal fusion	Algorithmic classification experiment	Moderate-high	Five predefined commands, controlled drone-disassembly setting, limited human-factor evidence, EEG and audio fragility	Relevant for HRC intention recognition, but not yet sufficient to prove broad shop floor deployability

6.5 Comparative robustness and failure modes

Robustness must be analysed at the level of the entire architecture, not only at the level of individual models. All three systems are vulnerable to sensor noise, environmental variability, and model uncertainty, but the dominant failure modes differ.

System A is mainly vulnerable to vision errors, context errors, and LLM output uncertainty, and the vision module must distinguish components, tools, and operations in a real workstation. In the proof of concept, the object-recognition model is trained on only a limited set of components and tools, and the paper reports some confusion between visually similar objects.

Performance may therefore become less reliable in a real industrial environment, where visual conditions and part variability are harder to control. A false positive could trigger an unnecessary corrective message, while a false negative could allow an assembly error to go unnoticed. These are not merely algorithmic errors, since they also affect human-machine interaction: repeated unnecessary alerts may reduce trust, while missed errors may allow defect propagation. Contextual dependence is also another issue regarding System A, because it can reduce the portability of the system. The risk is not that the LLM invents the entire procedure from nothing, but that it generates a fluent instruction from an incomplete or incorrect structured input. The risk is especially relevant when a fluent output gives a stronger impression of certainty than the underlying interpretation actually supports.

System B is vulnerable in different ways. Its main uncertainties concern physiological variability, calibration of the fuzzy rules, and the possibility of selecting an unsuitable form of support. The variables used to estimate cognitive load are proxies rather than direct measurements of the operator's mental state. Heart rate, body temperature, respiratory rate, task duration, and hesitation can all vary for reasons unrelated to cognitive difficulty. The fuzzy inference process combines these variables into cognitive load descriptors, but in doing so it also becomes dependent on the chosen rules, thresholds, and expert assumptions. Poor calibration to the individual operator or the specific context could therefore lead the system to overestimate or underestimate the need for support.

The robustness issue is different from that of System A. There, the system may misidentify an object or an assembly error; in System B, the incorrect interpretation may concern the operator's state. The consequences also differ. An object-recognition error mainly affects the guidance provided during the task, whereas an incorrect cognitive load classification may influence the level of assistance, escalation decisions, or the operator's degree of autonomy.

Categories such as 'Technological Support' and, in particular, 'Medical Aid' make this issue especially relevant. They are useful within the proposed decision framework, but they also produce outputs that may have significant consequences for the operator. In practice, this calls for cautious threshold setting, meaningful human oversight, and a way for system interpretations to be reviewed or corrected.

System C is exposed to yet another group of weaknesses, including synchronisation errors, degradation of individual channels, uncertainty in the audio pipeline, and instability in EEG acquisition. The paper reports considerably lower accuracy

for the individual modalities than for the fused model. Multimodal fusion clearly improves performance in the tested setting, but the results also reveal substantial differences between the channels. Vision may fail because of occlusion or an inadequate field of view; audio remains sensitive to noise, accent, pronunciation, and command ambiguity; and EEG signals are difficult to stabilise and may fluctuate considerably.

The attention mechanism can compensate for these weaknesses only when some of the available signals remain informative and are correctly synchronised. Similarly, the apparent natural-language capability should not be confused with robust free-form conversation. The audio pathway uses several recovery steps to map imperfect speech onto a limited command set. This does not reduce the value of the approach, but it does place clear boundaries around claims of robustness.

Looking across the three systems, a broader pattern becomes visible. System A performs reliably only while the visual scene and process conditions remain sufficiently stable. System B, instead, is robust only if physiological and behavioural proxies remain valid indicators of cognitive load. Finally, System C is robust only if multimodal signals remain synchronised and at least one or more reliable modalities dominate the fusion process.

Therefore, multimodality does not automatically guarantee robustness, but it improves robustness when modalities fail independently and when fusion can correctly weight reliable channels. If all modalities are affected by the same environmental disturbance, or if synchronisation fails, multimodality may increase complexity without increasing dependability.

6.6 Comparative human-robot interface models

The three systems embody three different HMI models: System A is an instructional guidance interface; System B is a reciprocal collaborative support interface; and System C is a multimodal command interface. These models differ in inclusiveness, transparency, and operator agency.

System A has the clearest inclusive rationale at the HMI level. It aims to reduce reliance on conventional written instructions by turning detected errors into simple, numbered corrective guidance delivered by voice. The paper specifically considers operators with reading difficulties and uses spoken instructions as a way to make assembly support easier to access. This is a sensible design choice from an inclusion perspective, since dense written procedures can be slower and more difficult to follow for some users. Audio guidance may reduce this barrier and help the operator keep attention on the physical task.

Nonetheless, System A's interface is mostly unidirectional. The system observes, diagnoses, generates, and speaks, while the operator receives and acts. The operator

is not strongly positioned as a co-adaptive partner, except through task execution and possible supervisor escalation. The interface is therefore inclusive in the sense of accessibility, but less reciprocal in the sense of shared learning. It supports the operator by simplifying instruction delivery, but it does not deeply model the operator's changing cognitive state or allow broad conversational negotiation.

System B offers a more collaborative but less explicit HMI. RL is the most distinct feature, since the machine supports the operator through error detection and robot assistance, whereas the operator can improve the machine through relabelling and feedback. The system is based on HITL and MITL mechanisms, and this is valuable because it avoids treating the operator as a passive recipient of automation. The human contributes to system improvement, while the system contributes to task execution. System B addresses a wider range of cognitive demands than System A. Its framework considers attention, memory, language, maths, logic, and reading, creating the possibility of linking different support responses to different forms of difficulty.

The case study tests only a limited portion of that broader design. The implemented scenario concentrates on memory-related cognitive load and RL, without demonstrating a fully developed interface that adapts across multiple neurodiverse profiles or support modalities. The HMI is therefore more developed at the conceptual level than in the empirical evaluation. There is also a possible downside to the feedback process: involving operators in relabelling and retraining may improve the model, but it can also introduce extra cognitive or organisational effort if the interaction is not designed carefully.

System C offers the richest input configuration, combining gesture, speech, and EEG. This is appealing in HRC because operator intention does not always have to be expressed through a single channel. Speech and gesture, for example, may help resolve ambiguity in one another, while EEG can provide an additional source of information. In practice, however, the tested interface remains quite constrained. The operator chooses among five predefined commands, each linked to corresponding hand gestures. GPT improves the interpretation of imperfect speech, but the system still works within a limited command structure rather than supporting open-ended conversation.

Its inclusive value is also less straightforward than the number of modalities might suggest. Offering several interaction channels may help users with different preferences or abilities, yet EEG equipment and tightly defined command structures may also create new barriers in terms of comfort, usability, and accessibility.

The comparison therefore shows that inclusive HMI design depends on how interaction is actually experienced by the user, not simply on how many modalities are available. System A uses fewer modalities but has a clearer accessibility target. System B offers deeper personalisation but requires sensitive cognitive-state inference. System C offers the broadest multimodal input but remains bounded and

technologically demanding. In inclusive manufacturing, the most suitable HMI is not necessarily the most sophisticated one but the one that reduces the relevant barrier with the lowest cognitive and operational burden.

6.7 Comparative technical complexity and industrial implementability

From an industrial implementability perspective, System A appears the most transferable in the short term, System B appears feasible as a structured pilot in advanced collaborative cells, and System C appears the most sophisticated but least immediately scalable.

System A has moderate technical complexity, because it has a more complex structure than static digital work instructions, but it does not require physiological sensors, EEG equipment, digital-twin infrastructure, or robot control.

System A can operate with relatively limited hardware, but deployment depends heavily on how well the assistance logic is configured for the target process. Each application would require a stable bill of materials, annotated data for relevant components and tools, a formalised task sequence, suitable error templates, and prompt structures that have been tested for the intended use case.

Among the three systems, this architecture appears the most transferable to a bounded assembly setting, especially where tasks are repetitive or semi-repetitive, components remain visually distinguishable, and the process is already well documented. The main implementation difficulties are likely to concern object-detection robustness, validation of generated guidance, and integration with existing shop-floor ?? infrastructure. A local or enterprise-grade LLM, combined with suitable validation and monitoring mechanisms, could make the architecture a realistic option for targeted industrial assistance.

System B involves a denser combination of sensing, inference, learning, and physical support. Its architecture connects cognitive-load estimation with ML classification, object detection, RL, and cobot assistance. Depending on the implementation, the hardware setup may include wearables or nearables, cameras, controlled lighting, a collaborative robot, and the associated control infrastructure.

The software stack is equally layered. It includes data normalisation, fuzzy inference, threshold logic, ML classification, YOLOv7-based perception, relabelling, retraining, and robot coordination. System B is therefore closer to a complete cyber-physical assistance architecture than to a standalone digital support tool. In a well-structured cell with repetitive tasks and a clear role for collaborative robotics, such an architecture could be feasible as an advanced pilot, although its deployment would require a relatively mature organisation.

The RL component is one of the more promising aspects of System B because operator feedback can be used to improve recognition performance. The cognitive-load layer is harder to scale. Fuzzy rules and thresholds need calibration, physiological sensing must be acceptable to operators, and detected support needs must be linked to interventions that are actually useful in practice. Implementation effort would continue well beyond initial installation. Sensors, models, thresholds, and the mappings between system outputs and support actions would all require maintenance as the process and operating conditions evolve.

System C is the most complex of the three architectures and appears the hardest to transfer beyond its experimental setting. It combines a depth camera, microphone, EEG headset, modality-specific models, attention-based fusion, GPT-assisted speech interpretation, ROS integration, and cobot execution. The wider architecture can also extend to digital-twin functions and cloud-based services.

The paper demonstrates this configuration in a simulated drone-disassembly task. Industrial deployment would add further requirements, including stable sensing, robust speech capture under factory noise, usable EEG equipment, accurate synchronisation, safety validation, robot integration, and continued model maintenance.

One of the most demanding features is the direct connection between multimodal inference and robot behaviour. An incorrect output is not simply a poor recommendation; it may affect the action executed by the robot. This places greater emphasis on conservative safety constraints, confirmation steps, fallback modes, and meaningful human oversight. The virtual and service layers could add value in advanced smart-factory settings, but they also increase dependence on additional infrastructure.

Taken together, the three systems show very different levels of industrial readiness. Among the three cases, System A appears the most straightforward to introduce into an existing assembly environment. It builds on familiar work-instruction practices and requires relatively limited changes to the surrounding cyber-physical infrastructure.

System B involves a heavier technical and organisational commitment because sensing, cognitive-state estimation, adaptive decision logic, and collaborative robotics must be coordinated within the same deployment. That effort may still be reasonable in applications where personalised support and HRC are central requirements. System C remains closer to an advanced research demonstrator. Its transition to routine industrial use is complicated by the multimodal hardware setup, the practical limitations of EEG, strict synchronisation needs, and the safety consequences of allowing inferred commands to influence robot action directly.

6.8 Engineering synthesis: main trade-offs across the three systems

The comparison identifies four main engineering trade-offs.

The first trade-off is simplicity versus functional richness. System A is the simplest architecture and therefore the most understandable and transferable. The system identifies deviations during assembly and responds with corrective guidance designed to be easy for the operator to follow.

Its relative simplicity is an advantage because it keeps integration requirements and data use comparatively limited. At the same time, this also constrains how far the system can adapt to the operator. System A does not estimate cognitive state, combine multiple signals to infer intention, or intervene directly in task execution through a robot. System B extends this functionality by adding cognitive-load estimation and RL-based support. It can adapt support based on inferred operator state and can improve machine perception through human feedback.

However, this richness depends on physiological sensing, fuzzy inference, thresholding, and ML classification, the system gains adaptiveness at the cost of greater calibration burden and interpretive uncertainty. System C, instead, maximises functional richness by combining multimodal intention recognition with robot execution. It offers the most advanced form of human-robot communication among the three systems, but also the highest technical burden. Its performance depends on synchronised multimodal sensing, attention-based fusion, speech-repair logic, EEG processing, and ROS-based actuation. The architecture is powerful, but less easily transferable.

The second trade-off is personalisation versus complexity. System A personalises mainly at the level of output format, since instructions are generated in a simple, supportive, voice-orientated style. This is useful for reading-related difficulties, but it does not require continuous modelling of the operator. System B, instead, personalises more deeply by estimating cognitive-load dimensions and selecting support strategies. System C personalises interaction by accepting multiple intention signals, but the actual command set remains narrow. The message is that personalisation can be achieved in different ways through the system: by adapting communication, by adapting support strategy, or by adapting input interpretation, and the deeper the personalisation, the greater the data, modelling, and validation burden.

A third trade-off concerns multimodality and practical implementation. Combining several input channels can clearly improve system performance, as System C shows through the higher accuracy achieved by multimodal fusion. At the same time, this approach introduces additional requirements for synchronisation, hardware integration, calibration, and maintenance. A similar issue appears in System B. Bringing

together physiological, behavioural, workplace, and process data may support a richer estimate of cognitive load, but it also adds more sources of uncertainty and raises further governance concerns. System A demonstrates, instead, the opposite strategy where fewer modalities make deployment easier, but reduce the system's ability to understand complex operator states.

Therefore, multimodality should not be treated as an intrinsic good and should be justified only when the additional modality contributes enough reliability or inclusiveness to offset its cost.

The fourth trade-off is automation of assistance versus human oversight. In System A, the system automates corrective instruction but leaves physical correction to the operator. System B automates parts of support selection and robot assistance, but still includes HITL feedback. System C goes further by translating inferred operator intention into direct robot action. As the system moves from providing advice to triggering physical action, stronger verification becomes necessary. An unsuitable instruction can often be questioned or ignored by the operator, whereas an incorrect cognitive-state classification may need human review, and a wrong robot command may require immediate safety intervention.

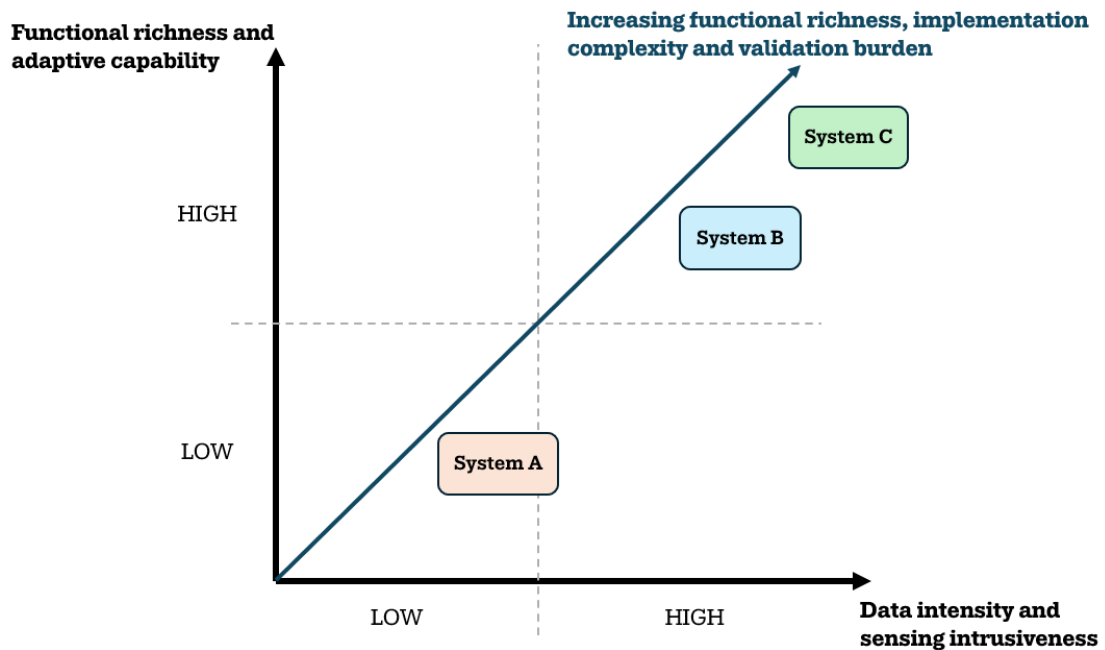


Figure 6.2: Qualitative engineering trade-off map of the three systems.

Overall, the comparison does not identify a single system as clearly superior in every respect. System A appears the easiest to deploy in practice and has the strongest accessibility focus, but the available evidence is still preliminary and its

reliability depends heavily on the stability of the full vision-to-guidance pipeline. System B is more closely aligned with the idea of adaptive and human-centred manufacturing because it connects cognitive-load estimation with support selection and RL-based assistance. This added adaptability, however, comes with greater dependence on sensitive data, calibration choices, and the reliability of the underlying inference process.

System C is the most technologically advanced because it integrates multimodal intention recognition and robot execution, but it is also the least mature for broad industrial transfer. In conclusion, inclusive AI-assisted assembly should not pursue maximum technological sophistication by default, and the appropriate design depends on the support objective.

System choice should follow the dominant support need. When reading-related barriers and error recovery are the main priorities, an architecture similar to System A may provide sufficient assistance without adding unnecessary technical complexity. Applications that require adaptation to changing cognitive strain point instead towards a System B-type configuration, although such use would demand stronger validation and careful calibration.

System C is most relevant where multimodal intention recognition and fluid HRC are central to the task. Its broader technical capability comes with a substantially greater burden in integration, validation, and safety assurance. The same features that improve technical capability also increase governance sensitivity, and engineering choices, therefore, cannot be separated from data protection, transparency, accountability, and human oversight.

The comparative result is not simply that the systems differ in performance, but that each architecture embodies a different balance between assistance, complexity, intrusiveness, and deployability.

Chapter 7

Legal and governance comparison

This chapter extends the comparison to the legal and governance implications of the three systems. It analyses how differences in data type, inference intensity, external services, transparency, retention, and system purpose affect the risk profile of each configuration under the GDPR and the AI Act. The chapter then derives design and governance requirements for deploying AI-assisted assembly systems in a proportionate, transparent, and legally robust manner.

7.1 Differentiating the legal profiles of the three configurations

The three configurations examined in this thesis are built around the same general industrial objective: supporting the operator during cognitively demanding assembly or disassembly tasks. However, their legal profiles differ because the technical means through which they pursue this objective are not equivalent.

The legal distinction does not map simply onto a binary opposition between 'AI' and 'non-AI', rather it emerges from the interaction between four elements: the nature of the input data, the inferential function performed by the system, the degree of identifiability of the outputs, and the organisational consequences attached to those outputs.

System A, as described by Fan et al. (2025) [10], processes images of the workstation, assembly instructions, error logs, context descriptions, and operator-linked metadata, and then uses an LLM to generate corrective instructions. Its legal profile is therefore shaped primarily by workplace video-based observation, operator-linked error logs, external GenAI processing, and the risk that error records may become performance records. The system does not process physiological signals, and this reduces the likelihood that Article 9 GDPR will be triggered by the raw data [18].

Nevertheless, the system still processes personal data when images, timestamps, workstation identifiers, or operator names make the worker identifiable. It also generates inferred data in the form of error characterisations and corrective narratives. Its lower sensitivity compared with Systems B and C should therefore not be confused with the absence of legal risk.

System B, as explained in the work of Simeone et al. (2024) [8], changes the legal analysis because it moves from task-state observation to operator-state estimation. The system combines physiological, behavioural, workplace, and process-related data, processes them through fuzzy inference, and assigns the operator to categories such as ‘Suitability’, ‘Technological Support’, or ‘Medical Aid’.

This changes the nature of the system from a legal perspective. It is no longer limited to checking whether the correct part or tool has been selected; it also builds an interpretation of the worker’s cognitive or functional state. Even where the purpose is inclusive support, the inference may become highly consequential because it can influence task assignment, escalation, assistance strategy, or organisational judgement about the worker’s capacity [28], [47]. For this reason, the GDPR analysis must include not only the raw physiological signals, but also the cognitive load descriptors and the final operator-state classification.

System C, as described by Li et al. (2025) [49], represents the most legally sensitive configuration. The architecture combines depth-camera gesture recognition, speech and audio processing, EEG acquisition, multimodal fusion, GPT-supported intent interpretation, cloud-enabled service layers, and ROS-based robot execution. Its legal profile is shaped by both the information being processed and the consequences of the resulting decisions. Neurophysiological and voice-related streams may reveal information beyond the immediate purpose of recognising command intent, making them more sensitive than ordinary process data. At the output stage, the system can move beyond guidance or support classification and translate an inferred command into robot action within a shared workspace. Legal and governance risk therefore arises from the data architecture as well as from the operational effects of an incorrect or poorly controlled decision.

The GDPR and the AI Act approach these issues through partly different regulatory lenses. Under the GDPR, the analysis centres on the processing of personal data and the conditions that make such processing lawful and proportionate. Relevant questions concern the nature of the data, the applicable lawful basis, necessity, retention, the involvement of external processors, and the extent to which workers are informed and able to exercise their rights [18]. The AI Act places additional attention on the AI system, its intended purpose, applicable risk classification, transparency, logging, human oversight, accuracy, robustness, and conditions of deployment [19].

The two perspectives do not always move together. A system may involve relatively limited data-protection concerns while still requiring strong AI governance because

its outputs directly influence work execution. Conversely, a non-generative system may raise substantial GDPR concerns when it processes physiological signals and uses them to infer or classify the worker’s cognitive state.

This comparative logic is represented in Figure 7.1, which positions the three systems along two analytical axes: data sensitivity and inferential intrusiveness on the one hand, and AI complexity and operational actionability on the other. The figure does not provide a numerical ranking of risk, but visualises the qualitative legal-risk gradient developed in this chapter. System A occupies a lower-risk position because it relies mainly on non-physiological task and error-log data, although its use of GenAI prevents it from being treated as legally neutral. System B moves upward because physiological and behavioural data are transformed into cognitive load and support-related inferences. System C is placed in the upper-right area because it combines multimodal and neurophysiological data, external GenAI and API components, and robot-action implications.

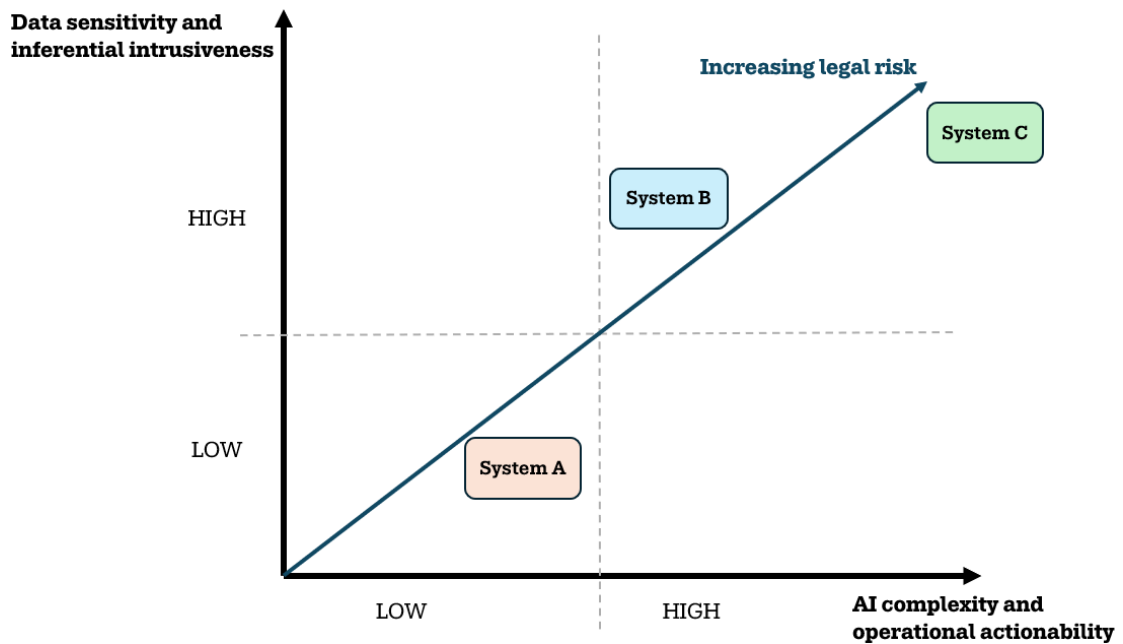


Figure 7.1: Conceptual positioning of the three systems according to data sensitivity, inferential intrusiveness, AI complexity, and operational actionability.

7.2 Risks arising from personal, inferred, and special category data

The first comparative question concerns data sensitivity. In all three systems, at least some data are likely to qualify as personal data when they relate to an identified or identifiable worker. The GDPR threshold is broad, because identifiability may arise not only from the operator’s name, but also from workstation assignment, timestamps, shift schedules, video context, login credentials, or repeated task histories. This is particularly relevant in production environments, where a workstation can often be associated with a specific worker even if the dataset does not include a formal identity field [18].

In System A, the raw data are comparatively less intimate, but they remain legally relevant. The paper reports error logs that include date and time, the person reporting the error, location, expected and actual components or tools, and corrective actions. The system also uses an intelligent vision module to identify components, tools, and operator actions, and then sends the error log to ChatGPT 4.0 through a custom Python interface. The immediate personal data risk therefore arises from the linkage between the operator, the workstation, the timestamp, and the recorded error. The legal sensitivity is amplified by the fact that the event log does not merely record a neutral process state, and it records that an operator made a specific component or tool selection error. In a workplace setting, these records may gradually take on a performance-related meaning, even when they were originally created for assistance.

System A also generates inferred information. It determines that a deviation has occurred, identifies it as a component, sequence, tool, or operation error, and then produces a corrective explanation. Such an inference does not in itself amount to special-category data, but it may still have consequences for the worker if repeated error logs are later used to identify recurring patterns, such as frequent tool-selection mistakes or persistent difficulty in following written instructions. The dyslexia-friendly voice-guidance function also requires care in the way user-related information is handled. Operational records should not be stored or exposed in a form that could reveal or imply a diagnosis, reading disability, or neurodiversity status unless that additional processing is separately justified and protected by appropriate safeguards.

For System A, the more immediate governance risk comes from the gradual transformation of routine operational records into behavioural information associated with individual workers, rather than from physiological monitoring.

System B has a substantially higher sensitivity, and it is described as a framework that acquires operator behaviour, physiology, workplace settings, and process characteristics, then uses fuzzy inference to generate cognitive load descriptors across attention, memory, language, math, logic, and reading.

Physiological metrics may include heart rate, respiratory rate, and body temperature, while behavioural data may be captured through camera tracking. From a legal perspective, these data become especially sensitive when they are used not

simply to describe the surrounding environment, but to draw conclusions about the worker's cognitive or functional state.

Another concern is the possibility of discrimination. When physiological, biometric, or similar signals are converted into indicators of fatigue, cognitive load, stress, attention, suitability, or need for support, they may indirectly reveal information linked to health, disability, neurodivergence, age-related conditions, or other personal vulnerabilities. In the workplace, such classifications may contribute to unequal treatment if they influence task allocation, access to specific activities, referral for medical review, performance assessment, or assumptions about a worker's competence. The concern therefore goes beyond privacy and data protection. It also involves the risk that workers are treated differently on the basis of bodily signals or characteristics inferred from them [17], [28], [47].

The risk increases when thresholds are calibrated around an assumed 'standard' worker, when training or validation data fail to represent different user profiles, or when support-related classifications become accessible to supervisors and are subsequently reused in managerial decisions.

Physiological data require particular care when they are processed to infer stress, fatigue, cognitive load, or a need for medical assistance. In such cases, the assessment should consider whether the resulting inference reveals information about the individual's physical or mental health and may therefore qualify as data concerning health under Article 9 GDPR.

The most delicate element of System B is the classification layer. The categories 'Suitability', 'Technological Support', and 'Medical Aid' are useful from an engineering perspective because they translate cognitive load assessment into operational support decisions. These categories are not legally neutral. 'Suitability', for example, may be read as an assessment of the worker's capacity, while 'Technological Support' may create a record of reliance on assistance. 'Medical Aid' moves even closer to the domain of occupational health.

For this reason, assistive decisions should remain clearly separated from managerial evaluation. If supervisors can access these classifications for productivity monitoring, performance assessment, or disciplinary purposes, a system designed to support inclusion could gradually become a tool for profiling workers.

System C raises the most significant data-protection concerns of the three cases because it combines vision, audio, and EEG signals collected from the same operator and within the same context. Visual input is used for gesture interpretation, the audio pathway combines speech recognition with GPT-assisted processing, and EEG data are collected through an Emotiv headset. The neurophysiological component is especially sensitive because the recorded signals may contain information unrelated to the specific command being inferred.

The intended function is to improve the robot's interpretation of human intention, but reuse or combination with other data sources could enable broader conclusions about the operator than the immediate interaction requires. Depending on the processing involved, this might include information related to attention, workload, fatigue, emotional arousal, or neurological patterns. The concern here is therefore not only collection, but also the possibility of drawing conclusions that extend

beyond the original purpose of the system.

Voice and audio data require a separate assessment. A voice command may constitute personal data when it relates to an identifiable worker, but it should not automatically be treated as biometric data. Under the GDPR, biometric qualification depends on the use of specific technical processing for the purpose of uniquely identifying a natural person.

In the system described, audio appears to be used for speech recognition and command interpretation rather than voice identification, so biometric status should not be assumed. Even so, recorded audio may contain other personal information, including background conversations, emotional cues, or details about the surrounding context. The minimisation question is, therefore, whether the system needs to retain audio files or whether it can process them rapidly and store only the final validated command.

Across the three systems, the comparative conclusion is that legal sensitivity increases when the system moves from observing task states to inferring worker states and increases again when sensitive worker-state signals are fused with GenAI and robot action, as summarised in Table 7.1.

The relevant object of protection is not only the raw data stream, but also the inferred representation of the worker that the system creates. In this respect, the three cases confirm the thesis developed in Chapter 3: AI-assisted work systems must be assessed at the level of the full data chain, from raw input to inference, classification, output, and possible organisational reuse.

Table 7.1: Comparative data and inference risk profile.

System	Main data profile	Main AI function	Primary legal sensitivity	Overall governance exposure
System A	Non-physiological, visual and operational data	GenAI-generated corrective guidance	Operator-linked error logs; external LLM processing; possible video monitoring	Moderate, if limited to real-time support and minimised logs
System B	Physiological, behavioural, workplace and process data	Fuzzy cognitive-load descriptors and ML state classification	Health-related inferences; suitability/-support/medical categories; worker monitoring	High, especially if outputs affect task allocation or evaluation
System C	Vision, audio, speech and EEG data	Multimodal large-model fusion and robot-executable intent inference	EEG sensitivity; voice and API processing; cloud service layer; direct robot action	Very high, because sensitive data and actionability are combined

7.3 Purpose limitation, data minimisation, and storage issues

Purpose limitation is especially important once assistive systems begin to accumulate data over time. In all three cases, the original objectives are legitimate and potentially beneficial, including improved accessibility, faster error recovery, lower cognitive strain, and safer HMC.

Problems may arise when information collected for those purposes is later reused in a different context. Operational records or inferred states could become attractive for quality analytics, workforce comparison, training-data generation, productivity assessment, disciplinary decisions, or medical screening. The legal problem is not that all secondary use is impossible but that secondary use cannot be silently derived from the technical availability of data. It must be specified, assessed, documented, and limited [17].

System A is particularly exposed to purpose expansion through logging. The error log is operationally useful because it structures the deviation and allows the LLM to generate corrective guidance. However, the log also creates a record of who made what error, at which workstation, and at what time. If these logs are accumulated without strict retention limits, they could become a sort of performance archive. The minimum implementable version of System A should therefore separate the real-time assistance function from longer-term analytics. A system designed for real-time correction can operate without retaining raw video or operator-linked records after the instruction has been delivered and the correction has been verified. A privacy-protecting setup used in industry might include a temporary session identifier, disposal of regular error logs at the end of the shift or after quality closure, and retention of only the aggregated, non-identifiable statistics for process improvement [41], [42].

In System A, data minimisation should also extend to prompts and API calls. To generate a corrective instruction, the LLM should receive only the information needed for that specific task, such as the step number, the expected and detected part or tool, the relevant assembly rule, and the required corrective action.

Information such as the operator's name, detailed location, shift history, supervisor notes, or raw video frames should remain outside the prompt unless there is a clear and separate need for them. Prompt templates should therefore be structured to exclude personal data by default. Where personalisation is needed for accessibility, it should be implemented through local interface preferences, such as preferred language, audio pace, or visual and audio modality, rather than through disability labels or long-term individual profiles [17].

System B raises a different minimisation problem. Its functionality depends on physiological and behavioural data because it aims to estimate cognitive load. Industrial deployment requires a clear distinction between information needed for immediate assistance and data collected mainly to enrich the model. For System B, this raises a practical question about whether all six cognitive-load dimensions need continuous physiological sensing. Some may be estimated through task- or

interaction-based proxies that are less intrusive. Raw physiological streams should also not be retained by default.

A more proportionate architecture would process data locally where feasible, use short rolling buffers, remove raw signals once the support decision has been made, and retain only coarse-grained support events where continued storage is justified. Records such as ‘additional guidance delivered’ may be sufficient for some operational purposes without making detailed physiological traces available to line management.

Assistive functions should also remain clearly separated from occupational-health processes. The ‘Medical Aid’ category requires a defined protocol and should not become a routine output for production management. Where the system detects a pattern that may justify health-related escalation, the DPIA should specify who receives the information, which conditions trigger further action, what human review is required, and how any resulting records are kept separate from ordinary performance-management data. A credible industrial design would place health-related outputs under occupational-health governance rather than supervisor dashboards, with strict access controls and retention [44].

System C has the most demanding minimisation profile because it combines three rich modalities. From an industrial standpoint, the system should not begin from the assumption that EEG is justified simply because it improves multimodal accuracy in a laboratory experiment. For System C, data minimisation starts with a practical question: is EEG genuinely needed to recognise the operator’s intention, or could the same goal be achieved through less intrusive combinations, such as vision with confirmed speech or gesture followed by manual confirmation?

The reported results show that multimodal fusion performs better than the individual channels in the tested scenario. This, however, does not mean that every modality is necessary in every deployment. The use of EEG should be justified by a demonstrable benefit that cannot be achieved adequately through less intrusive alternatives, particularly where safety or usability is concerned. Similar restraint is needed for storage. The mere availability of raw EEG, audio, or video data does not justify their routine retention.

Data collected for model improvement or retraining should follow a separate procedure with a clearly defined purpose. Where feasible, this work should be carried out in controlled validation settings rather than during normal production. Appropriate safeguards may include pseudonymisation, predetermined retention periods, and an explicit prohibition on reuse for worker evaluation.

Routine operation can follow a narrower logging model. Storage may be limited to the validated command, relevant confidence information, and records required for safety-related events. Retention periods should then reflect the type of event: ordinary commands may require little or no long-term storage, whereas incidents, near misses, or contested decisions may justify longer preservation.

An implementable industrial solution across the three systems is therefore not data elimination but data stratification, and Table 7.2 summarises how purpose limitation, minimisation, and storage controls should be differentiated across the three systems. Architecture should separate the different types of data, namely raw

sensor data, intermediate features, inferred states, final outputs, incident records, and aggregated process-improvement statistics. Each layer should serve a specific purpose and be associated with a defined retention period, access rule, and deletion [17], [41]. This layered approach allows the systems to remain useful without converting all operational data into permanent organisational memory.

Table 7.2: Purpose limitation, minimisation, and storage controls by system.

System	Primary legitimate purpose	Highest storage-risk objects	Industrial implementation safeguards
System A	Error correction during assembly	Raw video, full error logs, prompts sent to LLM	Use session IDs; remove name from prompts; store only verified correction event; delete logs after shift closure; aggregate errors by process, not by worker
System B	Cognitive-load mitigation and support	Raw physiological streams, cognitive-load scores, suitability/medical labels	Process locally; delete raw signals after inference; separate occupational-health alerts; restrict supervisor access; retain only support events or anonymised trends
System C	Intention recognition for safe HRC execution	Raw EEG, raw audio, raw video, fused command traces, cloud/API histories	Prefer less intrusive modalities; no default retention of EEG/audio/video; store validated commands and incident logs only; use on-prem or EU-hosted processing where possible

7.4 External services, APIs, and cross-border data-transfer implications

The use of external services changes the governance context because some processing may take place outside the factory’s direct organisational control. This is particularly important for GenAI and speech-processing components. The issue is not simply the use of an API, but the wider set of dependencies that may come with it, including additional service providers, different storage locations, service logs, possible use of data for model improvement, cybersecurity risks, and international data transfers.

Under the GDPR, this requires clarity on controller-processor allocation, Article 28 processor terms, security measures, retention by the provider, sub-processors, and, where relevant, Chapter V transfer mechanisms, including the framework set out in Articles 44-50 of the GDPR [18].

System A illustrates this issue clearly. In the paper it is stated that error logs were fed into OpenAI ChatGPT 4.0 via a custom Python interface. The legal significance of this design depends on what the error log contains. The examples reported in the paper include information such as date and time, the person reporting the event, location, workstation, expected and detected components or tools, and the corrective action taken. If these details are sent to an external LLM service, that transmission becomes a separate part of the processing chain and should be governed accordingly. A practical starting point is to reduce the information exposed to the model. Operator names, unnecessary workstation details, shift identifiers, and historical performance records should remain outside the prompt unless they serve a clearly defined function. Where external processing is involved, contractual and technical controls should address retention and deletion, encryption in transit, access logging, visibility over sub-processors, and any provider-side reuse of prompts or outputs for model improvement. Such reuse should not be assumed as part of the original processing purpose.

System B has less direct dependence on external GenAI services, although third-party data flows may still arise elsewhere in the architecture. Wearable platforms, smartwatch ecosystems, camera-analytics services, cloud dashboards, and robot-vendor infrastructure can all involve additional processors or processing environments without any LLM being present. The DPIA should therefore map how data enter, move through, and leave each sensor and software component, with particular attention to physiological information. Routing raw signals through consumer-device cloud services would be difficult to justify unless the organisation can show that processing terms, access controls, storage locations, and restrictions on secondary use are compatible with the workplace purpose. For raw physiological streams, local or edge processing would generally provide a more controlled option. System C raises the strongest concerns in this area because several external dependencies appear across different parts of the architecture.

The study refers to OpenAI Whisper for speech recognition, GPT API support for intent interpretation, Emotiv Pro software for EEG acquisition, a service layer connected to cloud services, and robot execution through ROS. Spoken commands or transcripts may therefore leave the local environment through the audio pipeline, while EEG acquisition depends on proprietary software and the service layer introduces further cloud-enabled functions. The governance issue is not tied to one provider alone, but to the combined effect of several external components across the same processing chain. The legal assessment must therefore be granular, as it is not enough to say that the system uses a cloud service. The organisation

must identify which data leave the local environment, whether they are raw data or features, whether they are stored by the provider, whether they are accessible for model improvement, where they are hosted, and whether they can be linked back to an individual worker.

Third-country transfers deserve a specific, although limited, comment. If personal data are transferred outside the European Economic Area, the GDPR transfer framework applies. The analysis may involve adequacy decisions, standard contractual clauses, transfer impact assessments, supplementary measures, and practical restrictions on access.

This thesis does not focus on international data transfers as its core object, but Systems A and C show why the issue cannot be ignored, since external LLM and speech-recognition services may be globally distributed unless configured otherwise. In a production deployment, the safest industrial design is to keep high-sensitivity data, especially physiological and EEG data, within the organisation or within EU-based processing environments, and to transmit only minimal, pseudonymised, task-specific prompts when external processing is unavoidable [18].

From a governance perspective, API dependency also affects traceability and contestability. If an instruction or command interpretation is generated by an external model, the deployer must still be able to reconstruct what input was sent, what output was received, which model or version was used, and which safety or validation layer accepted the output.

This is not only a GDPR accountability issue, but it is also aligned with the AI Act’s emphasis on documentation, logging, transparency, accuracy, robustness, and human oversight for high-risk systems [19]. External processing, therefore, requires both legal contracts and technical observability.

7.5 Transparency, human oversight, and contestability requirements

Transparency in the three systems should be understood at two levels. The first one is worker-facing transparency, and it means that an operator should know what the system sees, why it sees it, what inferences are made, and how these inferences influence the support that the operator gets. The second one is organisational transparency: the organisation should be able to audit a system’s data flows, model behaviour, logs, and decision points.

These two forms of transparency are connected, but they should not be treated as the same thing. A system may be well documented for engineers and still be difficult for workers to understand. Conversely, a simple user interface does not necessarily mean that the underlying processing is sufficiently traceable [30], [38]. System A offers the clearest basis for worker-facing transparency because its output

can be linked directly to observable task events. For example, the operator could be informed that the system expected component C26, detected C8 instead, and therefore generated corrective guidance for step S6. This supports contestability because the operator can compare the expected object, the detected object, and the physical state of the workstation.

A good implementation should make this explicit in the interface, since it should not simply play a voice prompt, but it should show the expected part or tool, the detected part or tool, the step rule that triggered the alert, and a confidence or uncertainty indicator. This would transform the LLM output from an authoritative instruction into a reviewable operational explanation. The main contestability risk in System A is the fluency of GenAI output [30], [38]. A plausible spoken instruction may still originate from an upstream perception error.

The LLM should therefore operate within validated procedural boundaries rather than act as an independent source of truth. Operators need simple ways to pause, reject, or flag incorrect guidance, while escalation to a supervisor should support recovery instead of assigning blame. When a worker corrects a false positive, the event should be logged as a system failure rather than interpreted as operator non-compliance.

System B presents a different transparency problem because the basis for an intervention may remain hidden from the operator. A support action could result from a combination of heart rate, task duration, environmental conditions, previous errors, and fuzzy inference rules, even though only the final intervention is visible. Meaningful transparency does not require exposing every sensor value. It requires an understandable and non-stigmatising explanation of the factors that led the system to trigger support. For example, the interface should say that the system is offering additional step guidance because the task has become complex or because the sequence has been repeated, rather than labelling the worker as cognitively overloaded. Internally, however, the system must maintain auditable documentation of which variables contributed to the classification, which thresholds were used, and whether the decision was confirmed by a human. Human oversight in System B must also be role sensitive.

Access to information should reflect the role of the recipient. A line supervisor may need to know that additional support was provided, without seeing the underlying physiological signals or cognitive-load score. Health-related escalation information may instead require review by qualified occupational-health personnel under predefined conditions. A split-visibility model can preserve this separation by providing operational support information for shop-floor assistance, restricted health-related information for occupational-health review, and aggregated statistics for process improvement. Workers should also be able to challenge or request review of a support classification, particularly where it may influence task allocation or perceptions of suitability [28], [47].

System C raises a more immediate oversight issue because an inferred command can be translated into robot action. The interface should make the interpreted command visible before execution when the action is non-trivial or safety-relevant. Additional confirmation may be appropriate where confidence is low, the command is ambiguous, the modalities disagree, or the resulting movement could create physical risk.

Because vision, audio, and EEG contribute jointly to the decision, disagreement between modalities should not disappear behind a single fused output. If speech indicates ‘move’, the gesture indicates ‘stop’, and the EEG signal remains uncertain, the system should request clarification before allowing the robot to act. Contestability in System C must also include immediate override and post-hoc review, where immediate override includes emergency stop, manual robot stops, and clear command cancellation. Post-hoc review permits retracing the command process starting from the audio transcript, gesture class, EEG classification result, fusion confidence, through to the final command and robot action. Such a detailed logging is indispensable for safety and accountability. However, it should be balanced with the need to minimise the logs: they should be centred on events and only kept for a long period in case an incident or a dispute arises.

The comparative conclusion is that oversight must be proportional to both opacity and consequence. System A needs operational explanation and false-alert correction, while System B needs non-stigmatising explanations, role-based review, and separation from managerial evaluation. System C, instead, needs real-time confirmation, modality disagreement handling, emergency override, and incident auditability. In all three systems, transparency should be designed as a practical condition for trust and contestability, not as a generic privacy notice.

7.6 Function creep, workplace monitoring, and proportionality safeguards

The principal governance risk across the three systems is function creep: a system introduced to support inclusion and reduce cognitive burden may gradually become a tool for monitoring, comparison, evaluation, or discipline [48]. This risk is structurally present because all three systems generate data about worker behaviour. The degree of risk, however, differs across the configurations.

In System A, function creep is most likely to occur through error-log accumulation. The system records deviations from assembly procedure and links them to time, location, and potentially operator identity. A manager could use these logs to identify the ‘most error-prone’ worker, compare recovery times, or infer reading difficulties. These secondary uses would be different from real-time assistance and should therefore be assessed separately in terms of purpose, legal basis, worker transparency, and proportionality. A proportionate approach would avoid using

assistance logs to generate individual productivity scores. Process-improvement analysis could instead rely on aggregated information at workstation, component, or task-step level.

The risk of function creep is greater in System B because the system directly classifies aspects of the worker's cognitive or functional state. The categories are intended to guide support strategy, but they could also be misused to rank workers as more or less suitable, to exclude them from tasks, or to justify closer supervision [14], [17], [47]. This would contradict the inclusive rationale of the system. Proportionality requires that cognitive-load assessment be used to adapt the environment or support modality, not to define the worker as deficient. The appropriate organisational question is not 'Which worker has poor cognitive capacity?' but 'Which task, interface, or environmental condition is producing avoidable cognitive burden, and what support is justified?'

In System C, instead, function creep may occur through the combination of multimodal monitoring and service-layer optimisation. EEG, speech, and gestures could be used not only to understand commands but also to profile attention, stress, compliance, or engagement. If the virtual layer or service layer stores and analyses these data for optimisation, the system could evolve into a continuous behavioural and neurophysiological monitoring platform [15], [17], [28]. The risk is especially high because the data are rich, synchronised, and action linked. A proportionate design must therefore define an explicit negative boundary, since EEG and audio data collected for command inference must not be used for productivity assessment, emotional analysis, disciplinary decisions, or general behavioural profiling.

The DPIA is the appropriate instrument to make these limits concrete, and it has a different focus for each analysed system, as shown in Table 7.3. System A already presents circumstances in which a DPIA would be appropriate and may become necessary, particularly where camera-based observation is systematic or extends across the workplace. The assessment should establish whether video is genuinely required for error detection, whether processing can remain local without retaining recordings, which information enters the prompts, and how long error logs are kept. It should also address the possibility that operational records are later reused for performance evaluation. Where comparable support can be achieved through less intrusive means, alternatives such as step-confirmation buttons, pick-to-light systems, or component scanners should be examined.

The need for ex-ante assessment is stronger in System B because physiological data, behavioural monitoring, cognitive-state inference, and potentially health-related classifications are brought into the same architecture. The applicable legal basis should be identified for each processing activity and, where Article 9 of the GDPR applies [16], the relevant condition for processing special-category data must also be established.

Sensor necessity, calibration error, and incorrect classification deserve particular scrutiny. The DPIA should define who can access cognitive-load outputs, how disputed results are reviewed, and whether such information could contribute to stigma or employment-related decisions. A clear boundary is also needed between operational assistance and occupational-health functions. Worker consultation or

representative involvement may be appropriate where the deployment materially affects monitoring, trust, or perceived autonomy in the workplace.

System C requires the broadest assessment because its data flows and operational consequences are more extensive. Relevant issues include EEG acquisition, audio and speech processing, computer vision, external APIs, cloud-based service layers, cross-border transfers, modality disagreement, and the possibility that an inferred command leads directly to robot action. The assessment should test whether EEG adds sufficient value for the intended task or whether less intrusive modalities could achieve an acceptable level of safety and accuracy.

Oversight procedures are equally important. The organisation should define when command confirmation is required, how emergency override operates, how incident records are retained, and how secondary uses such as emotion recognition or productivity profiling are excluded. A predefined response process is also needed for misinterpreted commands that lead to near misses or other safety-relevant events. The AI Act introduces a related layer of scrutiny centred on intended use and system function. AI used for employment, worker management, or decisions affecting task allocation may fall within high-risk categories depending on the specific deployment context. The Regulation also prohibits certain practices, including emotion recognition in workplace settings subject to the limited exceptions it provides [19]. The systems examined here are not presented as emotion-recognition tools: their stated purposes concern error correction, cognitive-load support, and intention recognition. Systems B and C nevertheless process physiological or neurophysiological signals that could support broader inferences about stress, fatigue, or affective state if their function or downstream use expands.

The DPIA and AI-governance documentation should therefore explicitly state what the system does not infer and should prevent later expansion into prohibited or disproportionate forms of worker monitoring.

Table 7.3: DPIA focus by system.

System	DPIA position	Main triggers	Core DPIA questions
System A	Likely strongly advisable; possibly mandatory depending on scale and national DPIA lists	Systematic video-based task monitoring; operator-linked error logs; GenAI/API transfer	Can the system work without storing video? Are operator names excluded from prompts? Are logs separated from performance management?

System	DPIA position	Main triggers	Core DPIA questions
System B	Essential and likely mandatory in practice	Physiological data; cognitive-load inference; suitability/medical categories; vulnerable employment context	Which physiological inputs are necessary? Who can see cognitive scores? How are health-related alerts separated from management?
System C	Essential and highest priority	EEG, audio, vision, APIs/cloud, robot-action consequences, multimodal profiling risk	Is EEG necessary? Are external services controlled? How are ambiguous commands handled? What logs are retained after incidents?

7.7 GDPR and AI Act implications across the three systems

This section applies the GDPR and AI Act implications to the three cases. The analysis does not provide a final compliance determination for any real deployment. Such a conclusion would depend on additional factors, including national labour law, the scale of deployment, contractual arrangements, sensor configuration, retention practices, and the way workers are consulted. The aim here is narrower: to identify the legal issues most likely to arise in each configuration and to highlight where the design may need to change before industrial use.

Figure 7.2 summarises this screening approach. It shows how the three configurations raise different concerns under the GDPR and the AI Act depending on the data involved, the extent of inference, reliance on external services, and the practical consequences of system outputs.

Seen in this light, the FRIA becomes particularly relevant because it requires attention to more than technical performance or data protection alone. It also asks how the AI system may affect fundamental rights within the specific organisational setting in which it is deployed. The FRIA can be understood as a structured assessment tool that connects the AI Act’s risk-based logic with a broader analysis of affected persons, contextual vulnerabilities, rights at stake, severity of impacts, and mitigation measures [22].

Under the GDPR, System A requires a lawful basis for processing visual and operational data. Consent is generally weak in employment contexts because of the imbalance between employer and employee. The more plausible basis would usually be legitimate interest or legal obligation or public-interest grounds where safety or quality obligations apply, but this depends on the concrete employer

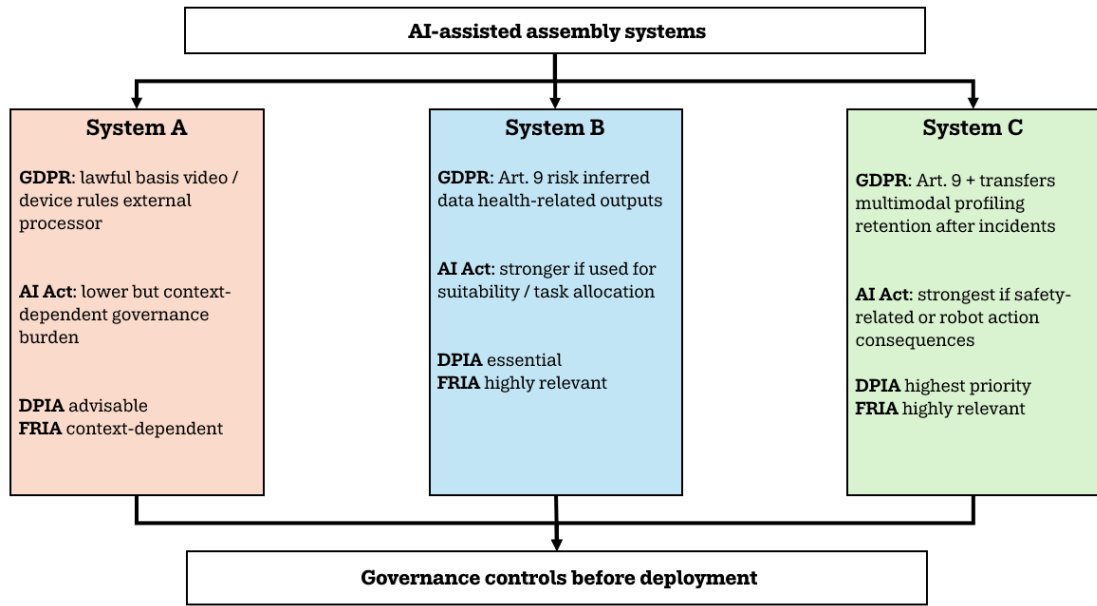


Figure 7.2: GDPR and AI Act assessment flow across the three system configurations.

and national law. In any case, legitimate interest would require a balancing test and a demonstration that the camera-based system is necessary for the assistive purpose and not overridden by worker rights. The EDPB Guidelines 3/2019 on video devices are relevant here because they require detailed purpose specification for video surveillance and stress that less intrusive alternatives must be considered [20].

System A also raises broader GDPR requirements. Workers would need appropriate information under Articles 13 and 14, while Article 25 requires data protection to be considered in the design of the system and Article 32 addresses security. Where the LLM relies on an external provider, processor arrangements and possible data transfers also need to be considered [18]. Article 22 does not appear to apply automatically, since the system mainly provides corrective guidance rather than making a solely automated decision with legal or similarly significant effects [18]. The assessment could change, however, if accumulated error logs were later reused for performance evaluation, disciplinary measures, or decisions about task allocation. Therefore, the GDPR design requirement for System A is to keep the system within assistance and quality-control boundaries, with clear limits on individualised reuse. Under the AI Act, System A may not necessarily be high-risk if it is used only as a real-time assistive instruction tool and its outputs do not determine work allocation, performance evaluation, promotion, termination, or access to employment [19]. This distinction is important because the workplace context alone should not be

treated as sufficient to classify the system as high-risk. What matters most is the purpose for which the system is introduced and how its outputs are actually used within the organisation.

A system that supports task execution or provides real-time corrective guidance has a different legal profile from one whose outputs influence task allocation, worker monitoring, performance assessment, suitability decisions, or other aspects of working conditions. If the system is used for monitoring or evaluating workers, or if its logs contribute to decisions about individuals, the employment and worker-management categories listed in Annex III may become relevant [19]. Even where the system does not fall within the high-risk category, basic governance measures would still be needed, including transparency, logging of prompts and outputs, human oversight, accuracy monitoring, and robustness testing. GenAI should also remain constrained by the underlying assembly procedure rather than generating corrective instructions without a controlled retrieval and validation process.

System B raises more significant concerns under the GDPR. The use of physiological data and inferences about cognitive load requires a clear justification in terms of necessity and proportionality. Where the data reveal information about physical or mental health, Article 9 of the GDPR must be considered. The system's classification into 'Suitability', 'Technological Support', and 'Medical Aid' is particularly sensitive because it may influence employment-related treatment, and the lawful basis should not be framed as worker consent unless genuine freedom of choice is demonstrable [18].

In some circumstances, processing may be justified by obligations connected with occupational health and safety, occupational medicine, or employment law, provided that the relevant condition under Article 9 and the applicable national rules are satisfied. Whether those conditions are met would depend on a context-specific legal assessment that falls outside the scope of this technical study.

System B also illustrates why the analysis cannot stop at the original inputs. A heart-rate measurement may be sensitive in itself, but the inference derived from it can have greater consequences than the underlying signal.

A cognitive load estimate or a final support category may have greater consequences for the worker than the original physiological signal itself. In similar workplace-technology literature, the Article 29 Working Party has emphasised proportionality before deploying new technologies and the need to minimise both the data collected and the information made visible to the employer [42]. Greco and Mantelero's analysis [41] are also useful for this case because their privacy-by-design analysis of smart industrial environments supports database separation, pseudonymisation, decentralised storage, exceptional re-linking, and automatic deletion. These safeguards are directly relevant to System B. Where feasible, raw physiological records should remain separated from operator identity, cognitive-load scores should not be routinely available to line management, and any re-identification mechanism

should be subject to access controls and logging.

System B may also warrant closer assessment under the AI Act than System A when its classifications are used for task allocation, suitability decisions, worker monitoring, or performance evaluation. The decisive factor is how those outputs are used in the employment context. Physiological sensing alone does not determine the regulatory position; greater concern arises when inferred classifications affect decisions about workers.

Discrimination risks become more tangible in this setting. Indicators derived from physiological signals may function as proxies for health conditions, disability, neurodivergence, stress tolerance, or presumed productivity. The consequences can therefore reach beyond data protection, particularly where workers are treated differently because of bodily signals or conclusions inferred from them. For this reason, the system should be designed so that physiologically derived indicators are used to adapt the work environment and the assistance strategy, not to rank, exclude, discipline, or compare workers. A FRIA would be particularly useful at this stage, because it would require the deployer to assess whether these inferences may affect non-discrimination, dignity, autonomy, access to work, and power asymmetries in the concrete workplace context [22]. If cognitive load descriptors remain confined to immediate, non-stigmatising assistance, the governance burden remains significant, but the high-risk assessment is more context dependent. If, instead, those descriptors influence allocation, suitability, escalation, or evaluation, the system moves much closer to the employment and worker-management logic of the AI Act. If, however, 'Suitability' determines whether a worker can perform a task, whether they are escalated for medical aid, or whether their performance is evaluated, high-risk worker-management implications become much stronger.

In such a case, the relevant AI Act requirements would need to be translated into concrete organisational and technical measures. These would include risk management, data governance, technical documentation, logging, appropriate information for deployers and workers, human oversight, accuracy and robustness controls, cybersecurity, and clear deployer responsibilities.

System C raises the most complex questions under the GDPR because it combines vision, speech and audio, EEG, multimodal fusion, external APIs, and direct robot action. Particular caution is needed for EEG data and for the conclusions that may be drawn from them. Their use in ordinary production should not be justified through broad or generic consent, nor should EEG be introduced simply because it improves classification accuracy in a laboratory experiment. The controller would need to show why this modality is necessary for a defined purpose, identify the relevant lawful basis and, where applicable, an Article 9 condition, limit the number of modalities used, set clear retention periods, and prevent incompatible secondary uses [18].

Audio data raise a different set of concerns. Spoken input may contain personal

information beyond the command itself, and the involvement of external services in processing transcripts or command interpretations may bring processor arrangements and international data transfers into the analysis. In those cases, Article 28 and Chapter V of the GDPR would need to be considered [18].

System C also raises additional questions under the AI Act. One concerns the role played by the system within the wider industrial environment. Depending on its intended use and integration into the product or machinery context, high-risk governance may become relevant beyond the employment categories alone. A second issue is the direct connection between inference and robot behaviour. Since an interpreted command may lead to physical action, human oversight and robustness are especially important. The assessment therefore depends on how the system is actually used and integrated, rather than simply on the fact that it operates in a workplace.

In this setting, requirements concerning accuracy, robustness, cybersecurity, logging, and human oversight have direct engineering consequences. They may translate into confirmation steps before execution, explicit handling of disagreement between modalities, fail-safe robot behaviour, version control for deployed models, incident logging, emergency-stop mechanisms, and monitoring of performance after deployment.

The AI Act's restrictions on workplace emotion recognition also deserve attention in relation to Systems B and C. The systems analysed in the selected papers are not presented as emotion-recognition tools. However, both process physiological or neurophysiological signals that could potentially support later inferences about affective states if the data were reused or the systems extended. This boundary should therefore be made explicit in the DPIA and in the broader AI-governance documentation. Internal uses that fall outside the assistive purpose should also be clearly excluded, including emotion scoring, engagement scoring, productivity-linked fatigue rankings, or other forms of affective profiling, unless a specific and legally assessed exception applies. This is an example of how legal compliance should be expressed as a design boundary, not merely as a policy statement.

The comparative GDPR and AI Act conclusion is therefore differentiated. System A is mainly a purpose-limitation, video-monitoring, logging, and processor-governance case. System B is a physiological-data, inferred-health, worker-classification, and anti-stigma case. System C is a multimodal special-category, API and cloud, high-risk HRC, and action-safety case. The three systems require different compliance architectures, even though they share the same inclusive-manufacturing narrative.

7.8 Design and governance requirements derived from the comparison

The comparison leads to a set of design and governance requirements that can be derived from the analysis.

These requirements are not generic ethical principles, they are concrete constraints that should shape the architecture of AI-assisted assembly systems before deployment. They also show that legal compliance and industrial implementability are not opposed, as depicted in Table 7.4. On the contrary, a highly intrusive, or opaque system will lead to organisational resistance, worker distrust, and regulatory exposure, all of which will make the system less implementable.

- A first step is to separate the different purposes of the system from the outset. Real-time assistance, safety intervention, quality control, model improvement, and collection of training data should not be treated as a single activity. Each of them may involve different data, access rules, retention periods, legal grounds, and permitted outputs. This distinction is particularly important for System A, where error logs created to support correction could later be reused for performance analysis. In System B, data collected to adapt support may gradually become the basis for assessing worker suitability. System C raises a similar concern because a system introduced for command recognition could later be extended toward behavioural or neurophysiological profiling.
- A second step is to minimise data at each stage of the system. Rather than treating all information in the same way, the architecture should distinguish between raw inputs, extracted features, inferred states, final outputs, and retained logs, and use the least intrusive data needed at each level. Raw video, audio, physiological, or EEG signals should not be stored routinely. Where retention is necessary for incident analysis or model validation, it should be limited in time and supported by a specific justification. For System A, this would favour real-time processing without keeping raw video as the default. In System B, physiological streams could be processed locally and removed once the relevant inference has been produced. In System C, the use of EEG should remain exceptional and should not become part of ordinary production unless less intrusive modalities prove insufficient for the intended task.
- A third principle is to keep identity separate from the rest of the processing wherever possible. Operator names should not be included in prompts or model inputs unless they are genuinely needed. Session identifiers, workstation codes, or rotating pseudonyms may be enough to support system operation without exposing a person's identity across every layer of the architecture. Where re-identification is necessary, for example after a safety incident, it should

follow a restricted procedure available only to authorised roles. This becomes especially important in Systems B and C, where physiological and EEG data make any link back to a specific worker more sensitive.

- A fourth principle concerns how system outputs are handled once they are produced. Corrective prompts, support recommendations, health-related alerts, and robot commands can have very different consequences and should not be managed in the same way. Corrective guidance can usually be delivered directly to the operator, provided that there is a clear way to question or correct it. Cognitive-support outputs should be framed around the needs of the task rather than presented as personal shortcomings. Any medical or health-related alert should follow an appropriate professional channel, while robot commands may require confirmation or additional constraints when the system is uncertain or the input is ambiguous.
- A fifth principle concerns meaningful human oversight. Meaningful human oversight requires intervention to remain possible at the stage where an incorrect output could produce material consequences. The appropriate safeguard therefore depends on the function of each system. In System A, operators should be able to reject unsuitable guidance and request human assistance. System B needs human review where a classification may influence task allocation or initiate health-related escalation. System C requires more immediate controls because inferred commands can affect robot behaviour directly. A real-time stop or cancel function should be available, and additional confirmation may be needed when modalities disagree or confidence falls below an acceptable level.
- A sixth principle concerns explanation and the possibility of challenging system outputs. Explanation and contestability should be matched to the function of each system. In System A, the operator should be able to identify the expected step and the mismatch detected by the assistance logic. System B should make the reason for a support intervention understandable without unnecessarily exposing or reinforcing sensitive health-related information. Where appropriate, System C should display the interpreted command and a confidence indication before robot execution. All three systems also need a clear mechanism for questioning incorrect outputs. Such interventions should be treated as evidence about system performance and recorded accordingly, rather than being interpreted as resistance or operator non-compliance.
- A seventh principle concerns the use of external providers and API-based services. External providers and API-based services introduce an additional governance requirement. Identifiable or sensitive information should leave the local environment only where external processing is genuinely necessary

for the intended function. The organisation should maintain visibility over processor arrangements, sub-processors, retention practices, data locations, and the security controls applied across the service chain. For Systems A and C, higher-sensitivity data may be more appropriately handled through on-premise infrastructure, meaning systems hosted and controlled within the organisation's own IT environment, or through suitable EU-hosted alternatives. Where an external LLM remains part of the architecture, prompts should be limited to task-relevant information, generated outputs should be validated before operational use, and the deployed model version should remain traceable.

- The eight principle is related to anti-function-creep governance. Assistive data should remain separate from unrelated managerial uses. Disciplinary action, individual productivity scoring, or covert worker monitoring should not emerge as secondary uses of information collected for support unless they are independently justified, assessed for proportionality, and made transparent to workers. Where possible, data generated by an assistance system should contribute to improving the technology or the working environment rather than ranking individual employees. Preserving this distinction is particularly important in inclusive manufacturing, where a system introduced to reduce barriers could otherwise evolve into a mechanism for algorithmic control.
- The ninth and final principle concerns monitoring throughout the system lifecycle. When product variants change, cameras are repositioned, new sensors are introduced, a different LLM version is deployed, support categories are revised, or data logs start being used for new analytical purposes, reassessment should be needed. This necessity for continued review follows both from engineering concerns around drift and robustness and from the requirement that data processing remain necessary, proportionate, and transparent as the system evolves. The DPIA should therefore be treated as a living document and updated whenever significant technical, functional, or organisational changes occur.

Table 7.4: Design and governance requirements derived from the comparison.

Requirement	General meaning	System A	System B	System C
Purpose decomposition	Separate assistance, safety, quality, training and analytics purposes	Error logs not reused as performance records	Support scores not reused as suitability rankings	Command/EEG data not reused as behavioural profiling

Requirement	General meaning	System A	System B	System C
Data-layer minimisation	Retain only the intrusive layer needed for each function	No default raw-video storage	Delete raw physiological streams after inference	No default EEG/audio/video retention
Identity separation	Prevent identity from flowing through all modules	Session ID in prompts	Separate physiology from identity	Exceptional re-linking only after incidents
Output governance	Classify outputs by consequence and restrict who sees them	Prompt visible to operator	Health-related alerts routed to qualified roles	Robot action requires confidence/-confirmation
Human oversight	Embed oversight where system error would matter	Reject/correct alert option	Review task-affecting classifications	Emergency stop and command cancellation
Contestability	Make system outputs challengeable and auditable	Expected vs detected object shown	Support reason explained without stigma	Fusion result and command trace reviewable
Third-party governance	Control external processors, APIs and transfers	Minimised LLM prompt	Local sensor processing	On-premise/EU processing for speech/EEG where possible
Anti-function-creep rules	Prohibit unsupported reuse for monitoring or discipline	Aggregate by process, not worker	No supervisor dashboard of cognitive scores	No emotion/productivity profiling
Lifecycle monitoring	Update DPIA and AI governance after changes	Re-test after camera/process changes	Re-calibrate thresholds and rules	Re-validate after model/API/fusion updates

Another important and more general requirement can be the participatory deployment. Because these systems operate in the workplace and are justified by inclusion, workers and, where relevant, their representatives should be involved in the deployment process. This is not only a labour-relations safeguard, but it is also an engineering safeguard: operators are often best placed to identify false alerts, intrusive interaction patterns, unworkable audio instructions, unacceptable sensor burdens, or ambiguous robot commands. Participatory governance can strengthen both legitimacy and system reliability [31], [33].

The three cases nevertheless create distinct governance demands. System A is comparatively easier to manage from a data-protection perspective, particularly where video is processed locally, retention is limited, prompts contain only necessary information, and assistive records remain separate from performance evaluation.

System B requires closer control because adaptation depends on physiological signals and classifications of cognitive state. In a workplace setting, these outputs may acquire health-related, sensitive, or evaluative significance, especially when they become linked to decisions about the individual worker.

System C presents the most demanding governance configuration. EEG, audio, vision, external processing services, and direct robot execution are combined within the same architecture, creating a setting in which sensitive data, multiple processing dependencies, and immediate physical consequences can interact.

The broader implication is that legal legitimacy depends on architectural restraint. Technology is not an excuse for systems to gather more detailed data or to keep more detailed records just because storing information is cheaper. Inclusion is a legitimate and important design objective, but it does not remove the need for necessity, proportionality, transparency, and contestability. On the contrary, it makes those safeguards even more critical because systems that are designed for workers with diverse cognitive profiles should not, at the same time, expose those workers to hidden profiling or excessive surveillance.

Therefore, the legal analysis does not simply dismiss AI-supported assembly, but it points out the scenarios in which such support can still be in line with human-centred and inclusive manufacturing. The final design goal is not a system that is merely efficient, but a system that is useful, bounded, explainable, contestable, and governable throughout its lifecycle.

Chapter 8

Discussion of findings

This chapter discusses the main findings emerging from the engineering and legal comparison. Rather than treating technical usefulness and legal legitimacy as separate dimensions, it examines the tensions that arise between inclusion, personalisation, sensing intensity, GenAI, and governance. The chapter also formulates the main contribution of the thesis in terms of proportionate and human-centred deployment of AI-assisted assembly systems.

8.1 Inclusion versus data intrusiveness

One of the first questions a company should ask is not which sensor or AI model to adopt, but which specific barrier the system is intended to address. In assembly work, barriers may concern reading instructions, maintaining the correct sequence, distinguishing similar components, recovering from errors, coordinating with a robot, or managing cognitive overload. Each of these barriers may justify a different level of technical support, but not necessarily the same level of data collection.

The literature on inclusive assembly and cognitive support shows that instruction modality and interface design can influence task execution, learning, and workload, especially where operators present heterogeneous abilities or support needs [3], [4], [29]. Therefore, companies should begin with low-intrusion interventions before introducing more intensive sensing architectures.

Through better work-instruction design, audio or visual cues, stepwise guidance, simplified language, and task-level verification, companies can already reduce many cognitive barriers without requiring the system to infer the worker’s physiological or cognitive state.

This leads to a general guideline, where inclusion should be pursued through the least intrusive effective support. A system does not become human-centric simply because it provides more adaptation, instead it becomes human-centric when the support is useful, proportionate, limited to the declared purpose, and protected from secondary organisational uses [41], [42].

A useful distinction for companies is between task-state data and worker-state data. Task-state data concern the assembly process, where it is necessary to highlight which component is present, which tool was selected, which step is active, or whether a sequence deviation occurred. Worker-state data concern the operator, with parameters like cognitive load, fatigue, physiological signals, behavioural patterns, or inferred support needs. Task-state data are usually easier to justify because they are directly connected to the assembly operation. Worker-state data, on the contrary, require a stronger justification because they can be interpreted as information about the worker's capacity, health, performance, or behaviour. In practical terms, an AI-assisted assembly system should first monitor the task rather than the worker [17], [28].

In a system similar to System A, for example, the camera should be configured to observe the task area rather than the operator as a person. Its field of view could focus on components, tools, hands, and the work surface, while excluding the face and the rest of the body as far as possible. The same principle should guide data retention. Raw video should not normally be stored, whereas verified correction events or aggregated process statistics may be kept where they are genuinely needed for quality control or system improvement and only for a clearly defined period.

Companies may consider behavioural or process proxies only when task-state data are insufficient: repeated re-checking, long pauses, interruption recovery time, recurring errors at a specific step, or environmental conditions could be a valid alternative in those cases. However, these signals are still potentially sensitive, but they are generally less intrusive than physiological monitoring. Nevertheless, physiological or neurophysiological data should be treated as an escalation layer, not as the starting point of inclusive design. Table 8.1 summarises this gradual approach. It suggests that companies should move from interface-level assistance toward worker-state sensing only when less intrusive forms of support are not sufficient for the intended purpose.

In practice, the fact that data are available does not by itself justify their use. Each additional data source should be linked to a clear assistive need, with evidence that less intrusive alternatives would not provide adequate support.

8.2 Personalisation versus proportionality

Personalisation is often presented as a core benefit of AI-assisted work systems. In assembly, personalised support may concern language, instruction format, audio speed, level of detail, visual layout, task pacing, or the type of corrective feedback. This can be useful because standardised instructions often assume a relatively uniform user, whereas inclusive manufacturing needs to account for different cognitive and practical needs [3], [29]. Companies should nevertheless keep a clear distinction between adapting assistance and profiling the worker. Personalisation can mean changing the interface, level of guidance, or communication modality so that the task is easier to perform. Profiling goes further by creating a persistent picture of the worker's abilities, limitations, cognitive state, or medical characteristics. The first can support inclusive design; the second may become intrusive or stigmatising and is harder to justify unless there is a clear need and strong governance in place [17], [28], [47].

Table 8.1: Staged implementation logic for proportionate support.

Layer	Main question	Preferred response
Interface adaptation	Can support be provided by changing the instruction format?	Use audio guidance, visual cues, simplified language, pacing options, or multilingual support
Task-state adaptation	Can the system support the task without modelling the worker?	Detect component, tool, sequence, or operation errors and provide step-level corrective guidance
Behavioural or process proxies	Are low-sensitivity indicators sufficient?	Use hesitation, re-checking, pauses, rework, interruption recovery time, or environmental conditions
Physiological sensing	Is worker-state sensing strictly necessary?	Use local processing, short retention, no raw-data storage, and task-support outputs only
Neurophysiological sensing	Is EEG indispensable for safety or accessibility?	Treat as exceptional and avoid default production use unless less intrusive modalities fail

A proportionate implementation should therefore begin with preference-based personalisation. The system may allow the operator to choose audio-first guidance, visual-first guidance, larger text, simplified language, slower audio pace, or multilingual output. These preferences should not require storing diagnostic labels, because a worker should be able to use an audio-first mode without the system recording that the person is dyslexic or has reading difficulties [31], [47].

This is particularly important in workplace contexts, where diagnostic or functional labels may affect how the worker is perceived. Where more intensive adaptation is required, system outputs should describe the support needed for the task rather than assign a status to the worker. Labels such as ‘suitable’, ‘unsuitable’, or ‘medical aid’ are difficult to justify as routine outputs on a production dashboard because they can attach evaluative or health-related meaning to the individual.

A less intrusive design would express the outcome in operational terms, using categories such as ‘standard guidance sufficient’, ‘additional instruction recommended’, ‘human review required’, or ‘task paused for safety check’. This shifts attention away from judging the worker and toward deciding how the work system should respond, reducing the risk of problematic individual classification [14], [17], [47].

This is especially relevant for systems similar to System B, where cognitive load estimates may influence the type of support provided. Such architectures can be technically valuable, but their outputs should not become capacity judgements. In fact, the purpose of personalisation should be to adapt the task environment, not to rank the worker [14], [17], [47].

The progressive architecture represented in Figure 8.1 can be used by companies as a design rule. Interface preferences and task-state adaptations should be prioritised. Behavioural and physiological data should be introduced only when less intrusive forms of assistance are insufficient and when their use is linked to a clear support objective. From a company perspective, proportional personalisation should satisfy four conditions. It should be useful for the task, limited to the specific support function, reversible by design, and separated from performance evaluation [14], [17], [41]. If these conditions are not met, personalisation may become a pathway to workplace profiling rather than inclusion.

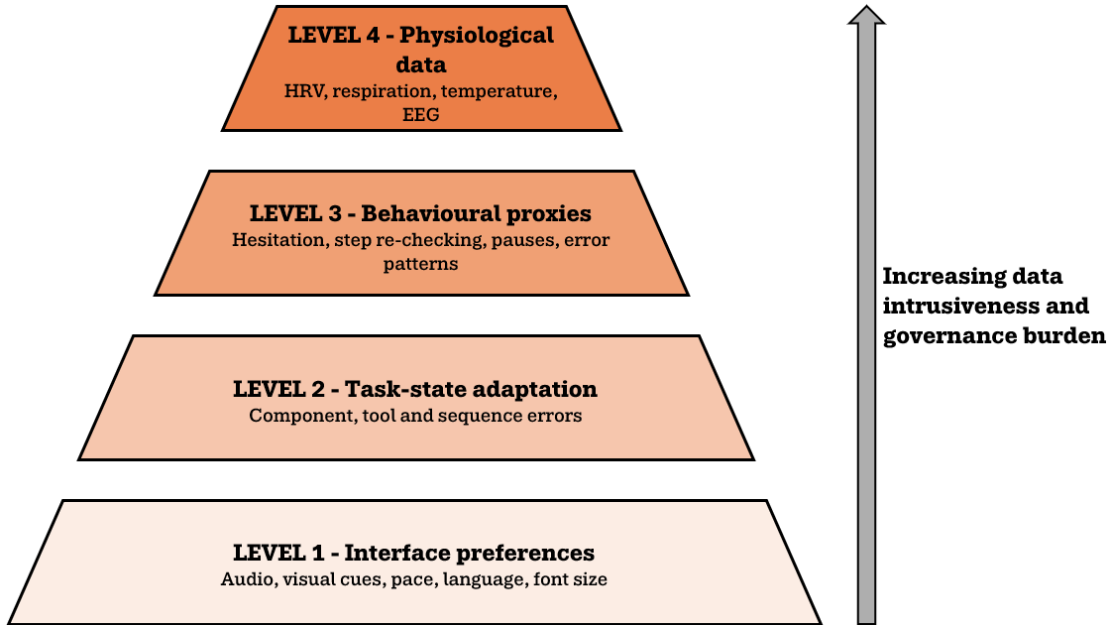


Figure 8.1: Progressive support architecture for proportionate implementation.

8.3 GenAI benefits versus reliability constraints

GenAI can support assembly work by turning technical information into explanations that are easier to understand and better suited to the immediate context. In a System A-type architecture, for example, it can convert structured error logs and assembly rules into simple corrective guidance. In a System C-type setting, it can instead help interpret imperfect speech input and link it to a predefined set of robot commands.

More generally, LLMs are increasingly discussed as tools for manufacturing knowledge extraction, instruction generation, and human-machine interaction [35], [43].

For companies, however, the main guideline is that GenAI should be used as a controlled interaction component, not as an autonomous authority over the assembly process. The GenAI component is most useful when it reformulates verified process information into guidance that is easier for the operator to follow. Determining the actual workstation state should remain grounded in the upstream perception and process logic. This separation is important because fluent language can make an incorrect instruction appear credible. A misidentified component or incomplete error log may still lead to a response that sounds coherent while being unsuitable for the task.

Generation should therefore remain anchored to validated process information and explicit control rules. Approved work instructions, bills of materials, step-specific constraints, and established recovery procedures can provide the basis for corrective guidance, limiting the need for open-ended generation. Output structure can also be constrained through numbered steps, verification points, and relevant safety warnings. Where the available evidence is insufficient, guidance should be withheld or sent for review instead of being delivered directly to the operator. One possible implementation is to link the level of generated assistance to confidence in the upstream information. A complete corrective instruction would be produced only when detection reliability is adequate, the inferred task state is consistent with the expected process, and the available context is sufficiently complete. Lower-confidence situations should trigger a different response, such as asking the operator to confirm the detected object, displaying the expected and observed states, or escalating the case for human review.

The governance issue becomes even stronger when GenAI contributes to robot command interpretation. If a generated or interpreted command can trigger physical robot movement, the company should require additional safety controls to prevent potential injuries or problems due to unexpected errors. Speech input should also be mapped only to predefined commands that have already been validated for safety. Where an action could have serious consequences, the robot should not execute it on the basis of speech alone. A second confirmation channel, such as a button, foot pedal, additional gesture, or visual prompt, could be required before the action proceeds. This reflects established human-automation research, which emphasises the importance of matching automation with appropriate human involvement and oversight rather than treating it as a substitute for human judgement [27]. Another important requirement for GenAI governance is traceability. Companies should document which model version was used, which input was sent, which output was generated, and which safety or human-oversight layer accepted the output. This is not only useful for debugging, instead it is also aligned with broader requirements for transparency, logging, robustness, and human oversight under the AI Act [19]. Moreover, when external APIs are used, companies should avoid sending operator names, full error histories, raw video frames, or unnecessary workplace identifiers to the model. In this context, prompt minimisation should be treated as part of data minimisation. The general rule is therefore that GenAI should remain bounded. It can be useful and relevant for reformulation, explanation, translation, and structured support, but it should not independently diagnose the worker, decide the physical state of the workstation, or trigger robot action without validation.

8.4 Physiological monitoring benefits versus data sensitivity

Physiological monitoring may provide valuable information when the company needs to detect excessive cognitive strain, fatigue, or overload before errors occur. Reviews of cognitive workload in Industry 5.0 highlight the growing interest in multimodal and physiological approaches, including cardiovascular, respiratory, ocular, neural, subjective, and performance-based indicators [12], [13].

However, physiological signals are not neutral industrial data, because when they are used to infer workload, stress, fatigue, or the need for support, they become part of a sensitive representation of the worker [28], [44].

For this reason, physiological sensing should not be introduced simply because it is technically available. Companies should first ask whether the same support objective can be achieved through less intrusive methods [17], [41], [42]. In many assembly contexts, cognitive burden can be reduced through improved instruction design, visual or multimodal guidance, stepwise presentation of information, task-level error-prevention logic, cobot support, and environmental adjustments [3], [4], [52]. These forms of support do not depend on physiological monitoring and should generally be explored before introducing worker-state sensing.

When physiological data are genuinely needed, their use should remain closely tied to immediate assistance. Raw signals such as heart rate, respiratory rate, or body temperature could be processed locally, converted into short-term support indicators, and then removed once the relevant inference has been made [17], [41], [44]. Routine production management does not require access to individual physiological traces or detailed cognitive-load profiles. Shop-floor interfaces can remain focused on task-related outcomes, using messages such as ‘additional guidance recommended’, ‘pause suggested’, or ‘human review required’ without revealing the underlying information about the operator [14], [15].

A separate handling process is needed where an alert may have health-related implications. Such information should not enter ordinary production workflows or be treated in the same way as operational support data. It should be subject to a separate protocol, with restricted access, human review, and separation from performance management [42], [44]. This distinction is important because physiological data can easily move from assistance to evaluation if the organisational boundary is not clear and well defined.

Neurophysiological data, such as EEG, require an even higher threshold. In a research setting, EEG may be useful to explore multimodal interaction or intention recognition, but, in ordinary production, however, its use should be exceptional and limited. If gesture and speech commands, combined with confirmation mechanisms, can provide adequate safety and usability, EEG should not be used as a default modality. A system similar to System C could retain much of its collaborative functionality through depth-camera gesture recognition, a limited set of speech commands, and suitable confirmation mechanisms, while leaving EEG out of routine deployment. This would be especially reasonable given its weak standalone performance in the reported study [49].

More generally, physiological and neurophysiological sensing should be used to improve

how the work system responds, rather than to build increasingly detailed classifications of the worker. Keeping this distinction clear helps separate inclusive adaptation from intrusive forms of profiling [28], [47].

8.5 Technical usefulness versus legal legitimacy

A system may be technically valuable because it reduces errors, shortens recovery time, supports collaboration with robots, or adapts instructions to the user. That alone, however, does not make its deployment legitimate. In practice, legitimacy also depends on whether the way data are collected, interpreted, stored, and used within the organisation remains consistent with the system's stated assistive purpose.

The GDPR and the AI Act should therefore influence system design from the beginning rather than being considered only at the compliance stage. Principles such as purpose limitation, data minimisation, storage limitation, transparency, integrity and confidentiality, and data protection by design can shape concrete technical choices, including sensor selection, camera framing, separation of databases, retention periods, prompt content, access permissions, and deletion rules [18]. The AI Act adds a further layer of governance by bringing attention, where applicable, to risk management, documentation, logging, transparency, human oversight, accuracy, robustness, cybersecurity, and monitoring after deployment [19]. These elements are fundamental for a system that lasts over time and allows for the correct use of the data and information generated.

For businesses, deploying AI-assisted assembly systems is an important decision that requires a structured governance process broken down into several stages. By setting a clear system purpose, mapping data flows, identifying if personal or special-category data are processed, exploring less intrusive alternatives, and carrying out a DPIA in case the processing of workers is likely to lead to high risks, the company can be ready before deployment. During deployment, the company implements role-based access, retention rules, human oversight, output contestability, and transparency to workers. After deployment, the company should monitor accuracy, false alarms, usability, workload, worker acceptance, and possible function creep, since this last one is one of the main organisational risks. Data collected to support the worker may later appear useful for productivity scoring, discipline, benchmarking, training, or workforce planning, and this secondary use should not be allowed by default [14], [15], [17].

Assistance logs should be kept separate from performance-management systems. Any later reuse should be assessed independently, considering its purpose, necessity, compatibility with the original use, and the safeguards required. This is especially important in employment settings, where the Article 29 Working Party has emphasised proportionality and data minimisation in light of the power imbalance between employers and workers and the particular sensitivity of consent and monitoring [42].

Table 8.2 turns these principles into a practical deployment checklist for companies introducing data-intensive assistance systems in the workplace.

The practical rule is that technical usefulness becomes legally legitimate only when the architecture prevents the misuse that its data structure makes possible. If camera data

Table 8.2: Company checklist for AI-assisted inclusive assembly systems.

Dimension	Company action
Purpose	Define the assistive function before selecting sensors or AI models
Data minimisation	Justify each data stream and prefer task-space data over worker-state data
Personalisation	Store interface preferences, not diagnostic labels or permanent cognitive profiles
GenAI governance	Use bound prompts, validated process knowledge, confidence gates, and output checks
Physiological data	Use staged sensing, local processing, short retention, and task-support outputs
Human oversight	Require confirmation for uncertain outputs, high-risk corrections, or robot actions
Access and retention	Separate assistance logs from performance management and apply role-based access
External services	Minimise prompts, assess API or cloud processing, and prefer controlled or EU/on-premise options
Validation	Test accuracy, usability, workload, false alarms, worker acceptance, and failure modes

can become surveillance, the camera must be constrained. If error logs can become performance records, identity must be separated and retention shortened. Where physiological indicators may influence judgements about worker capacity, the resulting outputs should be expressed as task-support recommendations rather than evaluations of the individual. Where inferred commands can trigger physical robot action, confirmation and oversight need to be embedded in the control logic.

Legal legitimacy is therefore better understood as a condition for sustainable innovation than as a constraint opposed to it. A system is unlikely to earn lasting worker trust or remain viable in practice if it cannot account for its data flows, justify the modalities it uses, limit retention, and provide meaningful ways to question its outputs. The most promising approach is therefore not to maximise AI capability at all costs, but to keep it within clear boundaries. The system should provide enough support to be genuinely useful, while remaining proportionate in its use of data, open to challenge, and reliable enough for safe integration into everyday work.

8.6 Main contributions of the thesis

The main contribution of this thesis is the development of a socio-technical evaluation logic for AI-assisted assembly systems. The analysis shows that these systems cannot be assessed only through engineering performance, nor only through legal compliance. Their real significance emerges from the interaction between sensing choices, AI paradigm, data sensitivity, inferential depth, human-machine interface, organisational use, and regulatory safeguards. The thesis contributes to the discussion on human-centric and inclusive manufacturing by showing that inclusion is not achieved by maximising adaptation, but by making adaptation proportionate and governable. Such systems can support workers by simplifying instructions, reducing unnecessary cognitive load, preventing error propagation, or enabling safer HRC. However, the same systems may become problematic if they require excessive monitoring, generate sensitive inferences, or retain support-related information that may later be used for evaluation or profiling. The proposed guideline is therefore to adapt the task and the support modality before modelling the worker.

Another contribution is the identification of a staged implementation architecture. Companies should first rely on interface adaptation and task-state support, then on behavioural or process proxies, and only later on physiological or neurophysiological sensing when less intrusive solutions are insufficient. This staged model translates proportionality into an engineering decision process, while it also helps companies decide not only whether a technology is useful but also whether its data requirements are necessary for the specific assistive function. The thesis also clarifies the appropriate role of GenAI in inclusive assembly. LLMs can be valuable when they transform verified procedural knowledge into accessible instructions, support multilingual or dyslexic-friendly communication, or help map imperfect speech inputs onto predefined commands. However, it is much more difficult to justify their use when they are considered independent thinking layers without a basis, validation, or human supervision. Within this approach, GenAI serves as a constrained interaction layer rather than an autonomous authority over the assembly process. The thesis also develops a more cautious position on physiological and neurophysiological sensing. Such data are not excluded from assistive systems, but their use should depend on a demonstrated need and appropriate safeguards. Physiological indicators should inform task support without becoming routine labels of worker capability. EEG and comparable neurophysiological signals require an even stronger justification in production settings, particularly where less intrusive modalities could provide adequate support.

The analysis is translated into a set of practical design guidelines that connect technical choices with relevant GDPR and AI Act considerations. These guidelines address camera framing, sensor selection, prompt minimisation, local processing, retention limits, role-based access, constrained command vocabularies, confidence thresholds, human confirmation, and the separation of assistance records from performance-management systems. The broader contribution of the thesis lies in this decision-orientated perspective. Rather than presenting one technology as universally preferable, it offers a structured way to assess when an AI-assisted assembly system is useful for the task, proportionate in its use of data, manageable from a governance perspective, and aligned with the wider goals of Industry 5.0.

Chapter 9

Conclusions and future research

This thesis has examined AI-assisted assembly systems for inclusive manufacturing as socio-technical architectures in which engineering performance, human-centred design, and legal-governance requirements are mutually interdependent and cannot be separated. The long-term validity of these systems cannot be judged by error reduction, task performance, or automation gains alone. It also depends on whether operator support remains technically reliable, proportionate to the actual need, and embedded in a data architecture that is understandable, purpose-limited, and manageable throughout the system lifecycle.

Across the three cases, AI-assisted assembly emerges as a credible route towards more inclusive and human-centric manufacturing, provided that specific technical and governance conditions are present into the design. The systems analysed demonstrate different ways of reducing cognitive friction in assembly work: through accessible digital instructions and voice-based corrective guidance, through cognitive load support and RL, or through multimodal intention recognition in HRC. However, the comparison also shows that increasing technical capability often corresponds to increasing complexity, data intensity, and governance burden. For this reason, the most defensible implementation strategy is not to maximise sensing or personalisation at any cost, but to adopt the least intrusive technical configuration capable of delivering a demonstrable assistive benefit.

From a technical perspective, the analysis suggests that relatively bounded systems may be more suitable for near-term industrial adoption. A system that monitors task progress, detects objects, and provides spoken corrective guidance can offer practical support to operators who face reading or procedural difficulties without requiring deeper inferences about cognitive difficulties or physiological state. System A therefore appears particularly promising for structured assembly settings, where components, tools, and task sequences are already well defined. At the same time, its reported benefits must be interpreted as proof-of-concept evidence rather than as definitive industrial validation, since the original study recognises limitations linked to lighting, camera positioning, deep-learning training, limited semantics, and the need for further experimental work on error types and cognitive-load effects [10]. System B shows a different and more adaptive logic. Its value lies in the attempt to connect cognitive load estimation, support selection, and

reciprocal learning within a HRC framework. This makes it relevant where inclusion requires not only clearer instructions, but also dynamic adaptation to the operator's cognitive demands. Nevertheless, this stronger adaptability depends on more intrusive and complex data processing, including physiological and behavioural information.

The original study recognises that future work should refine RL algorithms, improve adaptability across neurodiverse conditions, and address the difficulty of applying the system to diverse assembly environments and real data from neurodiverse operators [8]. System C shows the potential of multimodal fusion for recognising operator intention in collaborative manufacturing, with the reported fused model performing better than the individual modalities [49]. Its transfer to real industrial settings is less linear, however, because the architecture brings together vision, audio, EEG, large-model processing, and direct robot action. This combination places greater demands on robustness, safety validation, management of modality failures, and effective human oversight.

The legal and governance analysis leads to a convergent conclusion. AI-assisted assembly systems should not be assessed only by asking whether they process personal data but also by examining what they infer, how the output affects the worker, and whether the same data could later be reused for monitoring, profiling, or performance evaluation. In this respect the GDPR and the AI Act are better viewed as design constraints rather than compliance layers added from outside [18], [19]. These principles should shape the architecture from the implementation stage rather than be addressed only after the main technical choices have been made. Each data source should have a clear purpose, and less intrusive alternatives should be preferred where they can provide comparable support. Retention, access, and external processing arrangements also need to be defined explicitly. Workers, meanwhile, should be able to understand the system's role and challenge outputs they consider incorrect.

Function creep is a recurring concern across the three cases. Error logs, cognitive load indicators, or intention-recognition outputs may initially serve an assistive purpose but later become attractive for evaluation, monitoring, or other managerial uses. The risk is especially significant in employment settings because workers and employers do not hold equal organisational power. Assistance data should therefore remain clearly separated from performance-management systems, with additional boundaries around any later reuse. Physiological and neurophysiological signals should be used only when there is a demonstrated need for them, while cognitive or behavioural inferences should not become permanent labels attached to individual workers. In practical terms, the priority should be to improve the task and the working environment before building increasingly detailed models of the person.

The findings of this thesis should also be read in light of its methodological limitations. The analysis is based on published case studies and reconstructed system architectures, without direct implementation, experimental replication, or validation in real operational industrial settings. The three cases also differ in purpose, maturity, sensing strategy, and evaluation context. They can therefore be compared as techno-legal configurations, although they should not be treated as equivalent industrial products or expected to produce directly comparable outcomes. The underlying evidence is still limited in maturity, as most of the reported validation comes from prototype-level and short-duration studies. Consequently, aspects such as long-term effectiveness, worker acceptance, organisational

integration, trust calibration, and the durability of governance safeguards remain insufficiently demonstrated [12], [13], [33], [40].

The contribution of this thesis should therefore be read as a comparative and design-orientated assessment, rather than as a final empirical validation of the three systems.

Future research should therefore move in two parallel directions. On the empirical side, AI-assisted assembly systems should be tested in more realistic industrial environments and with more heterogeneous groups of users. Evaluation should combine task-performance metrics, such as time, error rate, recovery time, and false detections, with human-centred indicators such as workload, usability, trust, perceived autonomy, acceptance, and learning over time [6], [40]. Where ethically and practically feasible, manufacturing systems developed for inclusive support should be evaluated with the people they are intended to assist [3], [4], [31]. Otherwise, inclusion risks remaining a design assumption rather than a benefit demonstrated through empirical evidence.

Further technical research is particularly needed on system robustness, constrained uses of GenAI, and safe multimodal interaction. Camera-based systems, for example, need to be tested under changing lighting conditions, occlusion, visually similar components, and variations in the process itself [20]. Voice-based systems should be tested under noise, accents, PPE constraints, and multilingual settings, while physiological and EEG-based systems should be assessed against confounding factors, such as physical effort, stress unrelated to the task, fatigue, and individual variability [36], [51]. For GenAI-based assistance, future work should develop controlled pipelines based on validated process documentation, structured prompts, constrained outputs, traceability to source instructions, and human confirmation in uncertain or safety-critical cases [35], [43]. On the governance side, future research should translate the findings of this thesis into operational tools for companies. A useful development would be the creation of DPIA templates specifically tailored to AI-assisted assembly systems [22], [23], [41]. These templates should distinguish between task-state monitoring systems, cognitive load adaptation systems, and multimodal HRC configurations. They ought to help companies figure out data types, usages of data, types of output, data retention periods, levels of access, external service providers, human intervention and safeguards against misuse or secondary use. Also, design guidelines should guide organisations decide when simply changing their interfaces is enough, when behavioural or process data may be justified and when physiological or neurophysiological data are such highly intrusive options that they should also come with stronger safeguards.

The thesis ultimately argues for a cautious and constructive approach to AI-assisted assembly. These technologies can contribute to more inclusive manufacturing when they reduce avoidable cognitive barriers, support recovery from errors, and improve HRC without expanding into disproportionate forms of worker monitoring. Their value lies in creating work environments in which a wider range of people can perform effectively and safely, rather than in replacing workers or increasing the amount of detailed information collected about them. Hence, the main idea behind AI-assisted inclusive manufacturing would be very simple: the help should be so good that it makes technology worth introducing, and technology should be kept constrained in order to maintain the human-centred reason why it was brought in.

References

- [1] Maija Breque, Lars De Nul, and Athanasios Petridis. *Industry 5.0: towards a sustainable, human-centric and resilient European industry*. Tech. rep. Directorate General for Research and Innovation (DG RTD) of the European . . . , 2021 (cit. on pp. 1, 2, 25, 29).
- [2] Yuqian Lu, Hao Zheng, Saahil Chand, Wanqing Xia, Zengkun Liu, Xun Xu, Lihui Wang, Zhaojun Qin, and Jinsong Bao. «Outlook on human-centric manufacturing towards Industry 5.0». In: *Journal of manufacturing systems* 62 (2022), pp. 612–627 (cit. on pp. 2, 3, 12).
- [3] Jaakko Peltokorpi, Steven Hoedt, Thomas Colman, Kim Rutten, El-Houssaine Aghezzaf, and Johannes Cottyn. «Manual assembly learning, disability, and instructions: an industrial experiment». In: *International Journal of Production Research* 61.22 (2023), pp. 7903–7921 (cit. on pp. 4, 13, 15, 21, 28, 29, 31, 52, 134, 135, 139, 145).
- [4] Pieter Vanneste, Yi Huang, Jung Yeon Park, Frederik Cornillie, Bart Decloedt, and Wim Van den Noortgate. «Cognitive support for assembly operations by means of augmented reality: an exploratory study». In: *International Journal of Human-Computer Studies* 143 (2020), p. 102480 (cit. on pp. 4, 13, 15, 20, 29, 31, 47, 134, 139, 145).
- [5] John Sweller. «Cognitive load during problem solving: Effects on learning». In: *Cognitive science* 12.2 (1988), pp. 257–285 (cit. on pp. 5, 17).
- [6] Sandra G Hart and Lowell E Staveland. «Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research». In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 139–183 (cit. on pp. 5, 17, 19, 28, 145).
- [7] Faisal M Alessa, Mohammed H Alhaag, Ibrahim M Al-Harkan, Mohamed Z Ramadan, and Fahad M Alqahtani. «A neurophysiological evaluation of cognitive load during augmented reality interactions in various industrial maintenance and assembly tasks». In: *Sensors* 23.18 (2023), p. 7698 (cit. on p. 5).

- [8] Alessandro Simeone, Yuchen Fan, Dario Antonelli, Angioletta R Catalano, Paolo C Priarone, and Luca Settineri. «Inclusive manufacturing: A contribution to assembly processes with human-machine reciprocal learning». In: *CIRP Annals* 73.1 (2024), pp. 5–8 (cit. on pp. 6, 7, 19, 22, 25, 50, 54, 55, 74, 91, 92, 94, 98, 110, 144).
- [9] Alessandro Simeone, Yuchen Fan, Dario Antonelli, Paolo C Priarone, and Luca Settineri. «Conceptualisation of a multimodal, non-intrusive, generative AI-based assistive system for assembly». In: *CIRP Annals* 74.1 (2025), pp. 37–41 (cit. on pp. 6, 7, 21–23).
- [10] Yuchen Fan, Alessandro Simeone, Dario Antonelli, Alessandra Caggiano, Paolo C Priarone, and Luca Settineri. «A framework for digital assembly instructions as a step towards manufacturing inclusiveness». In: *Procedia CIRP* 132 (2025), pp. 116–121 (cit. on pp. 6, 21, 23, 25, 28, 49, 55, 60, 62, 91, 94, 98, 109, 143).
- [11] Yuchen Fan, Alessandro Simeone, Dario Antonelli, Angioletta R Catalano, Paolo C Priarone, and Luca Settineri. «A logic-assisted assembly support system to promote the inclusion of neurodiverse operators in manufacturing». In: *Procedia CIRP* 135 (2025), pp. 1119–1124 (cit. on pp. 6, 21, 23, 25, 28, 55).
- [12] Micaela Vitti, Antonio Padovano, and Francesco Facchini. «A review on cognitive workload for industry 5.0». In: *Computers & Industrial Engineering* 207 (2025), p. 111350 (cit. on pp. 6, 18, 19, 25, 28, 30, 139, 145).
- [13] Andrea Lucchese, Antonio Padovano, and Francesco Facchini. «Comprehensive systematic literature review on cognitive workload: trends on methods, technologies and case studies». In: *IET Collaborative Intelligent Manufacturing* 7.1 (2025), e70025 (cit. on pp. 6, 28, 30, 139, 145).
- [14] Valerio De Stefano. «‘Negotiating the algorithm’: Automation, artificial intelligence and labour protection». In: *Artificial Intelligence and Labour Protection (May 16, 2018). Comparative Labor Law & Policy Journal* 41.1 (2019) (cit. on pp. 7, 122, 136, 137, 139, 140).
- [15] Oliver G Kayas. «Workplace surveillance: A systematic review, integrative framework, and research agenda». In: *Journal of Business Research* 168 (2023), p. 114212 (cit. on pp. 7, 42, 78, 122, 139, 140).
- [16] Luciano Floridi et al. «AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations». In: *Minds and machines* 28.4 (2018), pp. 689–707 (cit. on pp. 7, 17, 44).

-
- [17] Michèle Finck and Asia J Biega. «Reviving purpose limitation and data minimisation in data-driven systems». In: *Technology and Regulation* 2021 (2021), pp. 44–61 (cit. on pp. 7, 30, 31, 42–44, 113, 115, 117, 122, 135–137, 139, 140).
- [18] European Union. *Regulation (EU) 2016/679 (General Data Protection Regulation)*. Official Journal of the European Union, L 119. CELEX: 32016R0679. 2016. URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (cit. on pp. 8, 31, 33, 35, 36, 43, 109, 110, 112, 118, 119, 125–128, 140, 144).
- [19] European Union. *Regulation (EU) 2024/1689 (Artificial Intelligence Act)*. Official Journal of the European Union, L 2024/1689. 12 July 2024. 2024. URL: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (cit. on pp. 8, 9, 31, 33, 42–44, 110, 119, 123, 125, 126, 138, 140, 144).
- [20] European Data Protection Board. *Guidelines 3/2019 on processing of personal data through video devices*. Guidelines. Final version, 30 January 2020. European Data Protection Board, 2020. URL: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201903_video_devices_en_0.pdf (cit. on pp. 9, 66, 78, 125, 145).
- [21] Michael Veale and Frederik Zuiderveen Borgesius. «Demystifying the draft EU artificial intelligence act». In: *arXiv preprint arXiv:2107.03721* (2021) (cit. on p. 9).
- [22] Alessandro Mantelero. «The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template». In: *Computer Law & Security Review* 54 (2024), p. 106020 (cit. on pp. 9, 30, 33, 37, 42–44, 124, 127, 145).
- [23] Paul Voigt and Axel Von dem Bussche. «The eu general data protection regulation (gdpr)». In: *A practical guide, 1st ed., Cham: Springer International Publishing* 10.3152676 (2017), pp. 10–5555 (cit. on pp. 11, 31, 33, 38, 41–44, 145).
- [24] Ewa Rollnik-Sadowska and Violetta Grabińska. «Managing neurodiversity in workplaces: A review and future research agenda for sustainable human resource management». In: *Sustainability* 16.15 (2024), p. 6594 (cit. on p. 13).
- [25] Elisa Castaldi, Manuela Piazza, and Teresa Iuculano. «Learning disabilities: Developmental dyscalculia». In: *Handbook of clinical neurology* 174 (2020), pp. 61–75 (cit. on p. 14).
- [26] David Romero, Peter Bernus, Ovidiu Noran, Johan Stahre, and Åsa Fast-Berglund. «The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems». In: *IFIP international conference on advances in production management systems*. Springer. 2016, pp. 677–686 (cit. on pp. 16, 22).

-
- [27] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. «A model for types and levels of human interaction with automation». In: *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 30.3 (2000), pp. 286–297 (cit. on pp. 16, 26, 27, 29, 138).
 - [28] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. «Why a right to explanation of automated decision-making does not exist in the general data protection regulation». In: *International data privacy law* 7.2 (2017), pp. 76–99 (cit. on pp. 17, 27, 64, 87, 110, 113, 120, 122, 135, 139, 140).
 - [29] Zhuo Wang, Xiaoliang Bai, Shusheng Zhang, Mark Billingham, Weiping He, Peng Wang, Weiqi Lan, Haitao Min, and Yu Chen. «A comprehensive review of augmented reality-based instruction in manual assembly, training and repair». In: *Robotics and Computer-Integrated Manufacturing* 78 (2022), p. 102407 (cit. on pp. 20, 28, 30, 52, 134, 135).
 - [30] Alejandro Barredo Arrieta et al. «Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI». In: *Information fusion* 58 (2020), pp. 82–115 (cit. on pp. 21, 26, 27, 64, 87, 119, 120).
 - [31] Vivian Genaro Motti and Anna Evmenova. «Designing technologies for neurodiverse users: Considerations from research practice». In: *International Conference on Human Interaction and Emerging Technologies*. Springer. 2019, pp. 268–274 (cit. on pp. 21, 22, 25, 29, 30, 76, 132, 136, 145).
 - [32] Ali Keshvarparast, Daria Battini, Olga Battaia, and Amir Pirayesh. «Collaborative robots in manufacturing and assembly systems: literature review and future research agenda». In: *Journal of Intelligent Manufacturing* 35.5 (2024), pp. 2065–2118 (cit. on pp. 22, 88).
 - [33] Margherita Peruzzini, Elisa Prati, and Marcello Pellicciari. «A framework to design smart manufacturing systems for Industry 5.0 based on the human-automation symbiosis». In: *International journal of computer integrated manufacturing* 37.10-11 (2024), pp. 1426–1443 (cit. on pp. 22, 25, 30, 73, 76, 132, 145).
 - [34] Hleg Ai. «High-level expert group on artificial intelligence». In: *Ethics guidelines for trustworthy AI* 6.15.3 (2019), p. 2024 (cit. on pp. 23, 30).
 - [35] Yiwei Li et al. «Large language models for manufacturing». In: *arXiv preprint arXiv:2410.21418* (2024) (cit. on pp. 24, 30, 48, 65, 138, 145).
 - [36] Maximilian Rosilius, Martin Spiertz, Benedikt Wirsing, Manuel Geuen, Volker Bräutigam, and Bernd Ludwig. «Impact of industrial noise on speech interaction performance and user acceptance when using the ms hololens 2». In: *Multimodal Technologies and Interaction* 8.2 (2024), p. 8 (cit. on pp. 25, 63, 86, 145).

- [37] Marta Lagomarsino, Marta Lorenzini, Elena De Momi, and Arash Ajoudani. «An online framework for cognitive load assessment in industrial tasks». In: *Robotics and Computer-Integrated Manufacturing* 78 (2022), p. 102380 (cit. on pp. 25, 29, 30, 76).
- [38] Margot E Kaminski. «The right to explanation, explained». In: *Research handbook on information law and governance*. Edward Elgar Publishing, 2021, pp. 278–299 (cit. on pp. 26, 27, 29, 64, 87, 119, 120).
- [39] Leon Eversberg and Jens Lambrecht. «Evaluating digital work instructions with augmented reality versus paper-based documents for manual, object-specific repair tasks in a case study with experienced workers». In: *The International Journal of Advanced Manufacturing Technology* 127.3 (2023), pp. 1859–1871 (cit. on pp. 28, 30).
- [40] Fred D Davis. «Perceived usefulness, perceived ease of use, and user acceptance of information technology». In: *MIS quarterly* 13.3 (1989), pp. 319–340 (cit. on pp. 28, 29, 145).
- [41] Laura Greco, Alessandro Mantelero, et al. «Industria 4.0, robotica e privacy-by-design». In: *IL DIRITTO DELL'INFORMAZIONE E DELL'INFORMATICA* 39.6 (2018), pp. 875–900 (cit. on pp. 30–35, 37–39, 41–44, 115, 117, 126, 134, 137, 139, 145).
- [42] PROCESSING OF PERSONAL DATA. «ARTICLE 29 Data Protection Working Party». In: (2012) (cit. on pp. 30, 35, 37–39, 41, 43, 115, 126, 134, 139, 140).
- [43] Christos Gkournelos, Christos Konstantinou, and Sotiris Makris. «An LLM-based approach for enabling seamless Human-Robot collaboration in assembly». In: *CIRP Annals* 73.1 (2024), pp. 9–12 (cit. on pp. 30, 138, 145).
- [44] European Data Protection Supervisor. *Health data at the workplace*. 2026. URL: https://www.edps.europa.eu/data-protection/data-protection/reference-library/health-data-workplace_en (cit. on pp. 34, 43, 44, 116, 139).
- [45] Els J Kindt. «Having yes, using no? About the new legal regime for biometric data». In: *Computer law & security review* 34.3 (2018), pp. 523–538 (cit. on p. 35).
- [46] Sandra Wachter and Brent Mittelstadt. «A right to reasonable inferences: re-thinking data protection law in the age of big data and AI». In: *Colum. Bus. L. Rev.* (2019), p. 494 (cit. on p. 35).
- [47] Solon Barocas and Andrew D Selbst. «Big data's disparate impact». In: *Calif. L. Rev.* 104 (2016), p. 671 (cit. on pp. 35, 110, 113, 120, 122, 135–137, 140).

- [48] Ifeoma Ajunwa, Kate Crawford, and Jason Schultz. «Limitless worker surveillance». In: *Calif. L. Rev.* 105 (2017), p. 735 (cit. on pp. 42, 121).
- [49] Juntao Li, Junyan Xiong, Zimo Zhang, and Daqiang Guo. «A Multimodal Large Model to Enhance Robot Understanding of Human Intentions for Accurate Human Robot Collaborative Manufacturing». In: *IFAC-PapersOnLine* 59.10 (2025), pp. 2820–2825 (cit. on pp. 51, 55, 87, 91, 92, 98, 110, 139, 144).
- [50] Sandra Mattsson, Dan Li, and Åsa Fast-Berglund. «Application of design principles for assembly instructions—evaluation of practitioner use». In: *Procedia CIRP* 76 (2018), pp. 42–47 (cit. on p. 63).
- [51] Patrik Gustavsson, Anna Syberfeldt, Rodney Brewster, and Lihui Wang. «Human-robot collaboration demonstrator combining speech recognition and haptic control». In: *Procedia CIRP* 63 (2017), pp. 396–401 (cit. on pp. 86, 145).
- [52] Rieke Leder, Hendrik Stern, and Michael Freitag. «Towards design guidance for the digitalisation of work instructions by focusing on technological possibilities and industrial requirements». In: *Procedia CIRP* 109 (2022), pp. 466–471 (cit. on p. 139).

Appendix

Table A.1: Reported evaluation evidence and methodological limitations of System A.

Evaluation dimension	What is reported	Structural evidence	Main limitation
Detection model training	YOLOv7 trained for proof-of-concept object and error detection	500+ epochs; batch size 16; learning rate 0.01; 2.3 hours	No standard metrics such as precision or recall
Training coverage	Reduced task coverage compared with full assembly	5 components; 2 tools; 3 operations	Only partial domain coverage
Classification quality	Mostly correct classification, minor confusion between similar classes	Qualitative statement only	No full confusion matrix values shown
Experimental setting	Comparison between baseline text instructions and digital voice guidance	1 operator; 10 repetitions	Extremely small sample
Baseline condition	Simulated dyslexia-related reading difficulty	Reading-based timing scenario	Simulated, not a real neurodiverse participant group
Error scenario E1	Component-selection error in step S6	Total saving 94 s; 61% reduction	Only one induced error type in one step
Error scenario E2	Sequence-related error in step S9	Total saving 36 s; 41% reduction	Only one induced error type in one step

Evaluation dimension	What is reported	Structural evidence	Main limitation
Instruction input time	Reduced when using digital audio instructions	E1: 27.8 s $\rightarrow$ 14.8 s; E2: 14.8 s $\rightarrow$ 9.8 s	Scenario-specific timing only
Error detection time	Strong reduction with digital system	E1: 45 s $\rightarrow$ 2 s; E2: 25 s $\rightarrow$ 2 s	Depends on setup assumptions
Error correction time	Reduced in at least one scenario	E1: 33 s $\rightarrow$ 3 s	Not consistently generalisable
Cognitive load	Mentioned as a likely benefit	No direct measure reported	No NASA-TLX or physiological measurement
Full framework validation	Not all modules were tested	Semantic analysis and image support excluded	Architecture broader than evaluation

Table A.2: Data and governance profile of System A.

Data item	Source	Technical function	Possible sensitivity	Governance-sensitive point
Workstation images	Camera or intelligent vision system	Detect components, tools, task state, operator action	Potentially yes, if operator is identifiable	Workplace monitoring
Detected component labels	YOLOv7 inference	Compare actual vs expected assembly state	Usually indirect	Operational data
Detected tool labels	YOLOv7 inference	Error detection and correction support	Usually indirect	Operational data
Detected operation state	Vision and process comparison	Identify incorrect action or sequence	Potentially linked to worker performance	Performance-related inference potential
Assembly instructions	Internal process documentation	Reference model for correct execution	No, unless linked to operator profile	Low sensitivity by itself

REFERENCES

Data item	Source	Technical function	Possible sensitivity	Governance-sensitive point
Manual context description	System designer or process engineer	Contextualise LLM prompt and NLP interpretation	Normally no	Knowledge-engineering asset
Error log: date and time	System log	Event traceability	Yes, when linked to an operator	Traceability of worker activity
Error log: operator name	System log or manual input	Personalise and identify event	Yes	Direct identifier
Error log: workstation/location	System log	Operational localisation	Yes, if linked to worker	Contextual identifier
Error description	Derived from detection and instructions	Explain deviation	Potentially yes	Can become performance-related record
Corrective action text	Template and system logic	Define recovery procedure	Indirectly yes	Contains process and event information
Prompt to LLM	Error log, instructions and context	Generate corrective instructions	Potentially yes	Important governance-sensitive transfer point
Generated textual instructions	ChatGPT 4.0 output	Human guidance	Possibly personalised	AI-generated operational content
Audio instructions	T2S rendering	Deliver guidance to operator	Low by itself	Final delivery layer
Repeated-error escalation record	Rule-based monitoring	Trigger supervisor intervention	Yes	Sensitive because it may support managerial intervention

Table A.3: Reported evaluation evidence and methodological limitations of System B.

Evaluation dimension	What is reported	Structural evidence	Main limitation
Overall evaluation design	Proof-of-concept comparison across three scenarios	Baseline (a), manual correction (b), RL correction (c)	Narrow experimental setting
Targeted CLT dimension	Analysis restricted to memory-related cognitive load	NMCL used as a focal indicator	Other CLT dimensions not empirically assessed
Neurodiversity modelling	Neurodiversity-related condition simulated rather than broadly validated	Forced warning values for memory, delayed execution to simulate memory problems	No substantial validation on real neurodiverse workforce
Data channels in experiment	Behavioural and workplace data controlled to non-warning levels	Multimodal framework reduced in practice	Full multimodal architecture not stress-tested
Assembly setting	Collaborative assembly case study with ten steps	Nominal step times reported for 10 tasks	Small and highly controlled task domain
Manual correction scenario	Error detected by operator and checked by supervisor	$t_{ed} = 5$ s, $t_{si} = 10$ s	Assumed timing structure
RL scenario timing	Robot-assisted and detection-assisted correction	$t_{od} = 1$ s, $t_r = 3$ s	Not fully comparable with baseline due to changed task configuration
Cognitive load result	RL associated with lower increase and faster recovery in NMCL	Evidence shown through scenario comparison chart	No inferential statistics or broad participant study
Object detection Phase 1	High recognition with one relevant confusion case	100% for most components, 9% error for gear unit #40, confidence 0.837	Small object set and controlled setting
Object detection Phase 2	Very high recognition	100% recognition, confidence 0.985	Phase-specific and setup-specific result

Evaluation dimension	What is reported	Structural evidence	Main limitation
HITL improvement	Operator relabelling improves model performance	Error for gear unit #40 reduced from 9% to 3%, 6 re-training epochs, about 2.5 minutes	Very local demonstration of reciprocal learning
Training setup	YOLOv7 trained on small datasets	Phase 1: 270 images; Phase 2: 300 images across 6 categories; 300 epochs, batch size 16, image size 640×640 , about 2 h total	Limited scale and uncertain external generalisability
Performance scope	Evidence concerns NMCL dynamics, task timing, and recognition performance	Mixed outcome structure	No long-term learning, acceptance, or broad productivity validation
Decision layer validation	Support strategy selected as reciprocal learning	Technological Support classification leads to RL	Classifier details not fully specified in the implemented case study

Table A.4: Data and governance profile of System B.

Data item	Source	Technical function	Possible sensitivity	Governance-sensitive point
Heart rate	Smartwatch or physiological sensing	Input for cognitive-load inference	Physiological / health-related	Sensitive bodily data used for worker-state inference
Respiratory rate	Wearable and nearable sensing framework level	Input for cognitive-load inference	Physiological / health-related	Sensitive bodily data, even if not fully instantiated in the case study
Body temperature	Physiological sensing	Input for cognitive-load inference	Physiological / health-related	Bodily monitoring linked to suitability or support decisions

Data item	Source	Technical function	Possible sensitivity	Governance-sensitive point
Blood pressure	Physiological sensing in the case study	Input for NMCL assessment	Physiological / health-related	Particularly sensitive inference base
Behavioural data	Non-intrusive camera tracking	Behavioural component of CLT assessment	Behavioural monitoring / surveillance risk	Interpretive monitoring of operator conduct
Real-time activity video	Camera stream in collaborative cell	Task analysis and object detection support	Potential personal-data relevance	Workplace monitoring and action traceability
Workplace temperature	Environment sensing	Context variable for CLT assessment	Usually low alone	Becomes sensitive when linked to operator-state inference
Noise	Context variable for CLT assessment	Context variable for CLT assessment	Usually low alone	Can contribute to worker-state classification
Lighting	Workplace or environment setup	Context variable and vision support	Usually low alone	Supports monitoring and intervention logic
Task duration nominal and actual	Process monitoring	Input for CLT assessment and scenario construction	Performance-related	Links worker execution to inferred cognitive state
Number of sub-tasks	Process model	Input for CLT assessment	Low alone	Structural process data used to infer operator burden
Task difficulty	Process model	Input for CLT assessment	Medium when linked to worker state	Embedded assumptions about cognitive demand
Number of items or components	Process model	Input for CLT assessment	Low alone	Part of inferred workload profile
CLT descriptors	Fuzzy inference output	Intermediate representation of operator state	High interpretive sensitivity	Converts heterogeneous proxies into cognitive-state descriptors

Data item	Source	Technical function	Possible sensitivity	Governance-sensitive point
NMCL	Fuzzy-based assessment output	Main case-study cognitive indicator	High interpretive sensitivity	Central inferred indicator driving support logic
Suitability class	ML-based decision output	Decision on operator status	High organisational sensitivity	Can affect access to tasks or perceived fitness for work
Technological Support class	ML-based decision output	Trigger for support-strategy selection	High organisational sensitivity	Can amount to profiling of support need
Medical Aid class	ML-based decision output	Escalated decision category	Very high organisational sensitivity	Strong fairness, dignity, and exclusion risk if misclassified
Selected support strategy	Decision module	Operational intervention choice	Medium to high	Links inferred state to concrete treatment of worker
Object images	Camera + training dataset	YOLOv7 training and inference	Context-dependent visual-data sensitivity	Visual monitoring and dataset governance issues
Object labels	HITL operator feedback	Retraining the object detector	Medium	Human feedback becomes part of system optimisation
Retrained model output	YOLOv7 + HITL loop	Improved object classification	Medium	Human performance influences machine adaptation
Robot action and assistance state	Collaborative robot system	Task support in RL scenario	Medium	Links operator state to machine intervention

Table A.5: Reported evaluation evidence and methodological limitations of System C.

Evaluation dimension	What is reported	Structural evidence	Main limitation
Overall evaluation design	Proof-of-concept multimodal HRC experiment in a drone disassembly setting	Laboratory simulation with UR5e cobot and human operator	Narrow task domain and controlled setup
Modalities used	Vision, audio, and EEG collected in the same spatiotemporal context	Three synchronised modalities explicitly described	Strong dependence on synchronisation quality
Target interaction space	Closed command-intention recognition task	Five predefined commands associated with hand gestures	Limited expressive scope and narrow generalisability
Dataset size	Synchronised multimodal dataset	1,000 total samples	Modest dataset scale for multimodal HRC
Single-modal training	Separate setup for unimodal models	Samples 1–500 for training; 501–700 for testing	Limited detail on variability and subject diversity
Fusion-modal training	Separate setup for attention fusion	Samples 501–850 for fusion training; 851–1000 for fusion testing	Overlap with overall dataset regime may limit broader interpretability
Vision performance	YOLOv5 used for vision-based gesture and object-related recognition	88.8% accuracy	Vulnerable to occlusion, view dependence, and scene constraints
Audio performance	Whisper-based speech recognition enhanced with keyword matching and GPT API support	66.6% accuracy; initial performance below 40%	Audio result depends on layered repair logic, not raw ASR robustness alone
EEG performance	EEG classification through pre-processing and SVM	30.4% accuracy	Weak standalone robustness and practical instability

Evaluation dimension	What is reported	Structural evidence	Main limitation
Multimodal fusion performance	Attention-based fusion improves intention prediction	96.7% accuracy	Accuracy reported without broader deployment metrics or uncertainty analysis
Attention mechanism	Dynamic weighting across modalities	Query-key-value fusion and backpropagation described	Limited explainability on which modality drives wrong predictions
Command execution	Fused command transmitted via ROS to UR5e	Demonstration of 'remove the object' execution	Demonstration-level evidence, not wide operational validation
Virtual and service layers	Included in architecture	Digital twin, safety monitoring, natural language interaction, cloud services described	Much less directly validated than the fusion-and-execution core
Industrial validity	Framed as support for complex manufacturing environments	Demonstrated in one drone-disassembly lab scene	Transferability to noisy, variable, real factories remains untested

Table A.6: Data and governance profile of System C.

Data item	Source	Technical function	Possible sensitivity	Governance-sensitive point
Hand gesture data	Depth camera	Vision-based intention recognition	Behavioural and personal data relevance	Continuous visual monitoring of operator action
Vision frames feed	ORBEC depth camera	Input for gesture recognition and environmental interpretation	Behavioural and contextual visual data	Action traceability and workplace monitoring
Audio recordings	Laptop microphone	Speech recognition and command interpretation	Voice data and possible identifiability	Potential external processing and command-level profiling

Data item	Source	Technical function	Possible sensitivity	Governance-sensitive point
Speech transcripts	Whisper and audio pipeline	Intermediate representation for command inference	Derived personal and behavioural data	Sensitive because transformed into actionable intent
Keyword-matched command candidates	Audio post-processing	Constrain speech output to predefined commands	Medium interpretive sensitivity	Encodes operator expression into bounded machine-readable actions
GPT-assisted interpreted command	OpenAI GPT API support	Intent inference from speech and predefined commands	High inferential sensitivity	External AI service involvement and opaque interpretation logic
EEG signals	Emotiv EPOX headset and Emotiv Pro	Physiological intention-related inference	Highly sensitive neuro-physiological data	Strong proportionality, privacy, and governance concerns
Intermediate EEG classifications	EEG pre-processing and SVM	Match EEG patterns to command classes	Derived physiological inference	Converts bodily data into command-related inference
Single-modal predictions	Vision, audio, EEG models	Inputs to multi-modal fusion	Inferred data	May already constitute meaningful operator-state representations
Attention weights	Fusion model	Assign modality relevance in integrated prediction	High interpretive sensitivity	Difficult to explain which modality drove the decision
Fused representation	Attention-based fusion	Integrated multi-modal command inference	High inferential sensitivity	Central opaque transformation point
Integrated command output	Fusion model	Final operator-intention inference	Operationally critical derived output	Directly triggers robot action
ROS control message	ROS execution pipeline	Transfer inferred command to cobot	Medium to high	Bridges inference to physical execution
Robot actuation state	UR5e cobot	Perform physical task based on inferred command	High operational significance	Wrong inference may become wrong physical action

REFERENCES

Data item	Source	Technical function	Possible sensitivity	Governance-sensitive point
Virtual-layer model data	Digital twin or simulation layer	System optimisation and virtual representation	Context-dependent	Potential secondary use and extended monitoring scope
Service-layer monitoring data	Safety and optimisation services	Operator safety, interaction, resource monitoring	High organisational sensitivity	Function creep toward broader monitoring or evaluation
Cloud-service interaction data	Service layer and external services	Support interaction and monitoring functions	Medium to high	Access control, retention, and possible transfer concerns