Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p27722-cd

Safety or surveillance? The 'Safety exception' for neuroergonomic monitoring in the EU AI act

Abstract

The European Commission's "human-centric" approach to Artificial Intelligence (AI) encourages the development of "Collaborative Intelligence" to enhance human capabilities, especially in safety-critical industries. This has spurred innovations in neuroergonomics and physiology-recording wearable sensors, designed to monitor operator states like fatigue, stress, and cognitive load to improve workplace safety. However, this same technology presents a profound risk of enabling unprecedented, invasive workplace surveillance. This paper identifies a critical, unresolved ambiguity at the nexus of this conflict: the EU's landmark AI Act. While the Act commendably bans "inferring emotions in workplaces" as an "unacceptable risk," it provides a vague exception for "medical or safety reasons." This "safety exception" is dangerously undefined and threatens to create a legal loophole for ethics-washing, where invasive surveillance is rebranded as a safety feature. This paper performs a conceptual analysis, arguing that this ambiguity cannot be resolved by the AI Act alone. We contend that the exception must be interpreted through a multi-layered legal framework integrating the stricter principles of the General Data Protection Regulation (GDPR) concerning health data, the emerging rules on algorithmic management from the Platform Directive, and the powerful co-determination rights of national labor laws, such as the German Works Constitution Act. We conclude by proposing a five-point operational test for regulators, developers, and Works Councils to distinguish between legitimate, safety-enhancing neuroergonomic systems and prohibited, disproportionate surveillance. This framework operationalizes the Platform Work Directive (2024/2831) provisions on algorithmic management and demonstrates how the AI Act's "human-centric" goals can be enforced through multi-layered legal accountability.

Keywords: EU AI Act, Neuroergonomics, Workplace Surveillance, Human-Centric AI, Safety-Critical Systems, GDPR, Algorithmic Management, Platform Work Directive, Co-determination, Emotion Recognition

1. Introduction

The European Commission's 2019 guidelines on ethics in AI established a foundational mandate for a "human-centric" approach, one that is explicitly "respectful of European values and principles" (Tambiama, 2019).This vision is not merely philosophical; it is an active regulatory project aimed at ensuring AI systems collaborate with, rather than replace, humans (Tambiama, 2019).This "Collaborative Intelligence" paradigm is considered essential for safety-critical applications in manufacturing, IoT, and critical infrastructure, where human-in-the-loop (HITL) scenarios are common (Ramos et al. 2024)(Schleiger et al. 2024).

This policy has, in turn, fueled significant technological innovation at the intersection of Human Factors and AI, particularly in the field of neuroergonomics. The "fourth industrial revolution" has seen the adoption of wearable technologies as a means to "enhance workplace safety and well-being" (Naranjo et al. 2025). AI-integrated wearable devices, such as those monitoring EEG, eye-tracking, or other physiological signals, promise to revolutionize industrial ergonomics (Naranjo et al. 2025). These systems can monitor for cognitive and physical strain, "verify the long-term tolerability conditions of work" (Donisi et al. 2022), and theoretically intervene before a fatigue- or stress-induced error leads to a catastrophic accident in a control room or on a factory floor (Amazu et al. 2024).

However, this technology has a dual-use nature that presents a profound ethical and legal conflict. The very same systems that monitor for "fatigue" (for safety) can just as easily be used to monitor for "concentration levels" or "emotional responses" (for surveillance) (Muhl, 2024). This paper explores the "emerging practice of workplace surveillance by using neurotechnologies," a dystopian risk that challenges workers' fundamental rights, dignity, and privacy (Muhl, 2024). An AI that infers a worker's cognitive state is a powerful tool; whether it is a "safety device" or a "surveillance tool" depends entirely on its implementation, the flow of its data, and the power dynamics governing its use (Foeth, 2026).

This paper argues that the EU's flagship regulation, the AI Act, has inadvertently created a critical legal ambiguity at the precise center of this conflict. In its effort to ban the most dangerous AI practices, the Act prohibits "inferring emotions in workplaces" (Future of Life Institute, 2024). Yet, it provides a crucial, one-line exception: "…except for medical or safety reasons" (Future of Life Institute, 2024).

This "safety exception" is the core problem. It is vague, undefined, and lacks any criteria for its application. The *entire justification* for deploying neuroergonomic systems in safety-critical industries is *for safety*. This creates a loophole so large that it threatens to nullify the ban in the very contexts it is most needed. It provides a clear pathway for

"ethics-washing," where invasive, individual-level performance monitoring can be legally justified under the guise of "safety."

Recent regulatory guidance has begun to acknowledge this tension. The European Commission's February 2025 guidelines on prohibited AI practices confirm that the emotion recognition ban applies to both direct identification and indirect inference through biometric data analysis (European Comission, 2025). However, the guidelines provide minimal operational criteria for applying the medical or safety exceptions, stating only that such exceptions must be "interpreted strictly" (Pape, 2024). This regulatory ambiguity is particularly problematic given that neuroergonomic monitoring systems are now being commercially deployed in safety-critical industries including aviation maintenance, nuclear power operations, and chemical process control (Maillant et al. 2025),(Naranjo et al. 2025).

The aim of this paper is to resolve this critical ambiguity. We argue that the "safety exception" in the AI Act is unenforceable in isolation and must be interpreted through the lens of other, stronger, and more established legal frameworks. By performing a doctrinal analysis that synthesizes the AI Act with the General Data Protection Regulation (GDPR), the Platform Directive, and national labor laws, this paper proposes a concrete interpretive framework - a test, to distinguish legitimate, human-centric safety applications from prohibited surveillance.

2. Neuroergonomics as a Dual-Use System

To understand the legal ambiguity, one must first appreciate the technology it governs. Neuroergonomics, a key topic of the call for papers, applies neuroscience to understand and optimize cognitive and physical work. In practice, this involves using wearable sensors to gather "psycho-physiological" data from operators.

Publicly available datasets and experimental designs for safety-critical environments, such as process control rooms, show the richness of this data (Amazu et al. 2024). These experiments collect multi-modal data to assess cognitive states, including:

- **Psycho-physiological Data -** Electroencephalogram (EEG), eye-tracking, and health monitoring watches (e.g., heart rate variability) (Amazu et al. 2024).
- **Subjective Human Factors Data -** Standardized metrics like the NASA-TLX (for cognitive workload) and SART (Situational Awareness Rating Technique) (Amazu et al. 2024).

2.1 The "Human-Centric" Safety Case (The Promise)

From a Human Factors and System Safety perspective, the use-case for this technology is clear and compelling. The goal is to "enhance workplace safety and well-being" and "improve worker comfort" (Naranjo et al. 2025). In a high-stakes environment like a chemical plant control room (Amazu et al. 2024), an operator's undetected high fatigue or cognitive overload is a direct precursor to an accident.

An AI-driven neuroergonomic system could, in theory, act as a vital safety control.

1. It could detect physiological signals associated with high cognitive load [4] or fatigue.
2. It could feed this information back to the operator (e.g., "Your fatigue level is high, recommend a break") or to the system itself (e.g., automatically increasing the level of automation to compensate).
3. On a systemic level, anonymized, aggregated data could "verify the long-term tolerability conditions of work" (Donisi et al. 2022), identifying which workstations or processes are poorly designed and cause the most stress, allowing for redesign.

Recent research demonstrates the feasibility of these interventions. Vieira et al. developed EEG-based systems capable of anticipating operator actions 300-500 milliseconds before execution, enabling proactive safety interventions in human-robot collaboration (Vieira et al. 2024). Similarly, advanced mental workload classification systems achieve 87-92% accuracy in distinguishing between low, medium, and high cognitive load states using multimodal physiological data (Amazu et al. 2024). These technical capabilities make real-time, individualized cognitive state monitoring not merely theoretical, but commercially deployable.

In this benevolent model, the technology serves the worker, augmenting their capabilities and protecting their well-being in line with the EU's "human-centric" vision (Tambiama, 2019).

2.2 The Dystopian Surveillance Case (The Peril)

The peril lies in the fact that the *exact same data* can be used for entirely different purposes. The "emerging practice of workplace surveillance by using neurotechnologies" (Muhl, 2024) involves monitoring "cognitive abilities, concentration levels, and emotional responses" (Muhl, 2024). This is not a futuristic hypothetical; the technological capabilities are demonstrated in existing research (Vieira et al. 2024).

In this surveillance model, the data flow is inverted.

1. The EEG or eye-tracking data is not fed back to the worker for their benefit; it is fed to a manager's dashboard.
2. The data is not used to improve the *system*; it is used to *evaluate the individual*.

3. AI algorithms can create a "black box" phenomenon (Foeth, 2026), ranking workers by their "concentration level," flagging them for "low engagement," or algorithmically managing them (Halefom, 2025).

This model is fundamentally extractive and disempowering. It creates significant data protection risks (Foeth, 2026), introduces biases (Labkoff et al. 2024), and transforms a potential safety tool into a mechanism of "dystopian" control (Muhl, 2024) that challenges human dignity.

The legal and ethical challenge, therefore, is that a *single AI system* can be both. An EEG-based fatigue detection system is, by definition, a system that "infers emotions" (or, at minimum, cognitive states). Whether it is a safety tool or a surveillance weapon depends entirely on its governance, which is why the AI Act's "safety exception" is so critical.

3. Deconstructing the EU AI Act's "Safety Exception"

The EU AI Act, approved in 2024, is widely regarded as the world's first comprehensive AI law (ISACA, 2024). It employs a risk-based approach, and its most powerful measure is the "unacceptable risk" category, which bans certain AI practices outright (Future of Life Institute, 2024).

3.1 A Strong Stance on Workplace Monitoring

The AI Act's text explicitly identifies two practices relevant to neuroergonomic surveillance as "prohibited":

1. **Biometric Categorisation -** The use of AI systems "that are biometric categorisation systems" to infer "sensitive attributes" such as "political opinions, trade union membership, religious or philosophical beliefs," etc. (Future of Life Institute, 2024)
2. **Emotion Inference -** The use of AI systems to "infer emotions in workplaces or educational institutions" (Future of Life Institute, 2024).

This appears, on its face, to be a major victory for workers' rights. A system that analyzes a worker's brainwaves (EEG) or heart rate variability to infer their "stress," "fatigue," or "concentration" level is, by any reasonable definition, a biometric system inferring a sensitive cognitive or emotional state. The Act seems to ban precisely the dystopian scenario outlined in the previous section.

3.2 The Exceptions

The problem lies in the exceptions provided. The ban on "inferring emotions in workplaces or educational institutions" is immediately followed by a critical qualifier:

> "…except for medical or safety reasons." (Future of Life Institute, 2024)

This single clause is the central vulnerability of this paper. The AI Act provides no further definition, criteria, or guidance to interpret the scope of this "safety exception." This is not a minor oversight; it is a critical ambiguity that strikes at the heart of the call for papers.

The research in human-centric AI for safety-critical systems is *explicitly* focused on developing systems for "safety reasons." The "human-in-the-loop decision support in process control rooms" dataset, for example, is designed to evaluate AI support systems in a safety-critical plant (Amazu et al. 2024). The entire "promise" of neuroergonomics in Industry 4.0 is predicated on enhancing "workplace safety" (Naranjo et al. 2025).

This creates a direct conflict:

- The *technology* (neuroergonomics) is an "emotion/cognition inference system."
- The *stated purpose* (enhancing safety) places it *precisely* within the "safety exception."

Recent official interpretations have attempted to narrow this loophole. The European Commission's 2025 guidelines specify that the safety exception applies only when monitoring is "directly necessary to prevent imminent physical harm" and where "no less intrusive alternative exists" (Pape, 2024). However, these guidelines lack binding legal force and provide no operational test for "direct necessity" or procedural mechanisms for verifying that alternatives have been exhausted. An employer seeking to implement continuous concentration monitoring for control-room operators could still argue that the system prevents accidents and therefore qualifies for the exception, particularly if framed as detecting "cognitive overload" (a safety concern) rather than "low engagement" (a performance concern).

This loophole threatens to render the ban on workplace emotion inference meaningless in all safety-critical industries. Without a clear, enforceable interpretive framework grounded in existing legal mechanisms, the exception will swallow the rule, enabling "ethics-washing" where invasive surveillance (Muhl, 2024) is rebranded as innovative "safety technology."

4. A Multi-Layered Doctrinal Solution: Re-contextualizing the AI Act

The ambiguity of the AI Act's "safety exception" cannot be resolved by looking at the Act in isolation. The Act itself must be "read with" the existing, robust scaffolding of EU and national laws that already protect workers' data and dignity. We argue that any attempt to use the "safety exception" must *also* survive scrutiny under three other legal layers: the GDPR, the Platform Directive, and national co-determination laws.

4.1 Layer 1: The General Data Protection Regulation (GDPR)

A neuroergonomic system does not just process "data"; it processes "biometric data" and "data concerning health" (Article 9, GDPR). This is "special category data" and is subject to the GDPR's most stringent protections:

- Processing this data is prohibited *unless* a strict exception applies, such as "processing is necessary... for the purposes of... assessment of the working capacity of the employee" (Art 9(2)(h)).
- The GDPR mandates that data be "collected for specified, explicit and legitimate purposes" (Art 5(1)(b)) and be "adequate, relevant and limited to what is necessary" (Art 5(1)(c)) (Foeth, 2026).
- This is the key test. A proposed "safety" system must be *proportional*. Is continuous, individual-level EEG monitoring (Vieira et al., 2024) “EEG-Based Action Anticipation in Human-Robot Interaction.” a proportional response to the safety risk, or could the same risk be mitigated through less invasive means (e.g., better shift scheduling, redesigning the workstation, or simply providing the worker with an anonymous, voluntary feedback tool)?

The GDPR's principles of necessity and proportionality (UNESCO, 2021) must be used to *narrow* the AI Act's "safety exception." A system that is disproportionate, "black box" (Foeth, 2026), and collects more data than is strictly necessary for a specified safety purpose would fail the GDPR test, even if it claims a "safety" purpose under the AI Act.

4.2 Layer 2: Algorithmic Management (The Platform Directive)

The EU's Platform Work Directive (2024/2831), adopted in April 2024 and requiring member state transposition by December 2026, introduces groundbreaking rules for "algorithmic management" that have direct applicability to AI-driven neuroergonomic systems (European Union, 2025).

The Directive defines algorithmic management as "automated monitoring or automated decision-making systems used to support decisions" affecting working conditions

(European Union, 2025). Critically, a neuroergonomic system that processes cognitive state data to recommend or trigger interventions (e.g., "operator flagged for fatigue," "task reallocation recommended," "break mandated") constitutes algorithmic management under this definition (Halefom, 2025).

Transparency Obligation (Article 6)

Platforms must inform workers about (a) automated monitoring systems and (b) automated decision-making systems that affect work conditions. Information must include "the categories of data processed, the logic behind the system, and the significance and envisaged consequences" (European Union, 2025).

Data Minimization (Article 6(5))

Platforms "shall not process personal data concerning the emotional or psychological state of the person performing platform work" except where strictly necessary for contract performance. This directly prohibits generalized stress or concentration monitoring unless demonstrably necessary (European Union, 2025).

Human Oversight (Article 7)

"Decisions significantly affecting working conditions" including "dismissal, suspension, or penalization" must "not be taken solely by an automated decision-making system" but require "meaningful human review" (European Union, 2025), (Halefom, 2025)

Explanation Rights (Article 8)

Workers have the right to "obtain from the digital labour platform a clear and simple explanation" of any decision affecting them, including "the grounds on which the decision was based" (European Union, 2025).

These provisions must be applied to neuroergonomic systems in safety-critical contexts. An AI that processes EEG data to infer fatigue states and triggers task reassignment is engaging in algorithmic management requiring transparency, minimization review, human oversight, and explainability. The Platform Directive thus provides concrete, binding procedural requirements that narrow the AI Act's "safety exception" – even a legitimate safety system cannot operate as an opaque, fully automated black box (Foeth, 2026).

Importantly, while initially aimed at platform workers, member states may extend these principles to all employment relationships involving algorithmic management. Germany's draft transposition law explicitly considers broader application to "algorithm-controlled work" beyond platforms, suggesting these principles could become general employment law standards (Halefom, 2025).

4.3 Layer 3: National Labor Law & Co-Determination

Recent developments strengthen this co-determination framework. Germany's Works Council Modernization Act (Betriebsratsmodernisierungsgesetz), enacted in 2021 and amended in 2025, explicitly extends co-determination rights to "AI-based systems that monitor, evaluate, or influence employee behavior or performance" (BMAS, 2025). Works councils now have the right to engage external AI experts at the employer's expense when evaluating proposed AI systems (BMAS, 2025).

Austrian law goes further: the Arbeitsverfassungsgesetz (Labor Constitution Act) grants works councils veto power over monitoring technologies that "affect human dignity" – a standard that Austrian courts have interpreted to include systems creating "continuous psychological pressure" through performance tracking (Halefom, 2025). A 2023 Austrian Supreme Court decision held that real-time productivity monitoring via wearable sensors violated worker dignity even when implemented for ostensible "safety" purposes, because it created "a reasonable apprehension of constant evaluation" (Halefom, 2025).

This jurisprudence establishes a critical principle: a system's *subjective impact* on worker autonomy and dignity is legally relevant, not merely its *stated purpose*. Even genuine safety benefits cannot justify technologies that fundamentally alter the psychological contract of work by creating pervasive monitoring (Foeth, 2026).

5. A test for the Safety Exception

By synthesizing these three legal layers, we propose a five-point test for regulators, developers, employers, and Works Councils to determine if a proposed neuroergonomic system qualifies for the AI Act's "safety exception." A legitimate system should pass all five tests. The logic is visible in Table 1 and flow in Figure 1 below.

Table 1: A Proposed Test for the AI Act's "Safety Exception"

Test	Legitimate Safety System (Human-Centric)	Prohibited Surveillance (Human-in-Peril)	Concrete Example	Legal Basis
1. Locus of Control & Agency	Data is used for feedback directly to the worker for their own use (e.g., a private haptic alert). The worker retains agency.	Data is used for reporting to a manager or a third party, who then acts upon the worker.	**PASS:** Private haptic alert to operator when fatigue detected. **FAIL:** Dashboard showing each operator's "concentration score."	GDPR (Art. 5, 9); Platform Directive (Art. 7)
2. Data Aggregation & Target	Data is anonymized and aggregated to identify systemic, ergonomic risks (e.g., "Workstation B causes high stress").	Data is used at the individual level to track, rank, compare, or profile a specific worker.	**PASS:** "Workstation B shows 40% higher cognitive load than Workstation A." **FAIL:** "Operator Smith has 15% lower concentration than team average."	GDPR (Art. 5(1)(c), 9(2)(b)); Platform Directive (Art. 6)
3. Nature of Output & Consequence	The system's output is a non-punitive safety intervention (e.g., machine guard, team-wide break, option to rest).	The system's output is tied to an individual's performance record (e.g., "concentration score," disciplinary flag).	**PASS:** Automated increase in machine assistance level when overload detected. **FAIL:** Automated flag for "low engagement" in performance system.	Platform Directive (Art. 7, 8); AI Act (Art. 14)
4. Governance & Implement.	The system was co-designed, vetted, and approved by worker representatives (e.g., Works Council) in a binding agreement.	The system was unilaterally imposed by management without meaningful worker consultation or co-determination.	**PASS:** Binding works agreement specifying data flows and review rights. **FAIL:** Management-only decision to deploy after "informing" workers.	National Labor Law (BetrVG §87, §94; ArbVG §96)
5. Necessity & Proportion.	The invasive monitoring is the only feasible and least invasive means to mitigate a specific, proven, and significant safety risk.	The risk could be mitigated by simpler, non-invasive means (e.g., engineering controls, better staffing, shift redesign).	**PASS:** Nuclear control room where no other engineering solution exists. **FAIL:** General manufacturing where better shift scheduling would suffice.	GDPR (Art. 5, 9); Charter of Fund. Rights (Art. 8)

This framework provides a clear, defensible, and legally-grounded path for interpretation. It operationalizes vague ethical principles like "accountability" (UNESCO, 2021) and "fairness" (Microsoft, 2025) into concrete, testable propositions. A system that fails this test - for example, by sending individual concentration scores to a manager without a Works Council agreement - cannot and should not be able to use the "safety exception" as a legal defense.

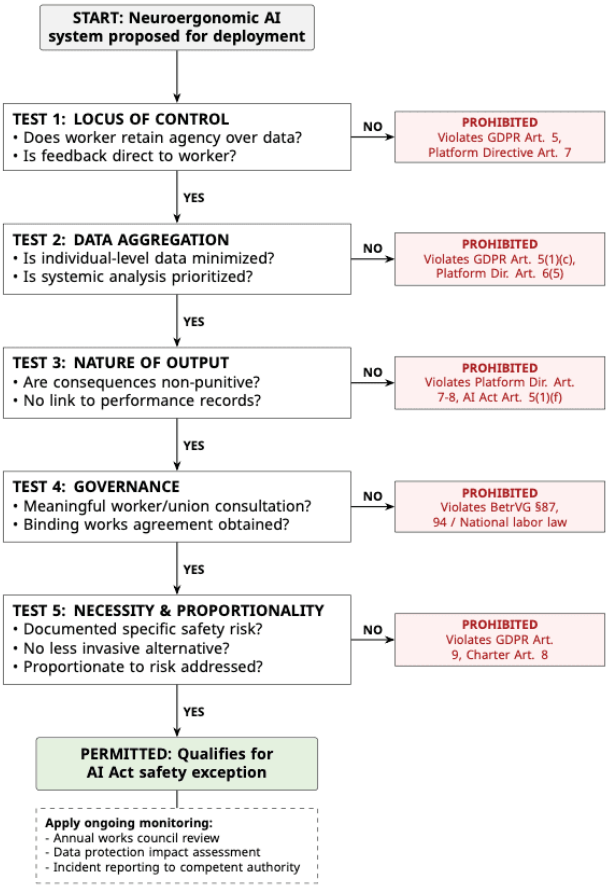

Figure 1: Decision tree for evaluating AI Act "safety exception" claims for neuroergonomic systems. A system must pass all five sequential tests to qualify; failure at any point renders the system prohibited.

5.1 Application of the Five-Point Test: Illustrative Scenarios

To demonstrate the practical utility of the five-point test, we apply it to three realistic deployment scenarios encountered in safety-critical industries:

A. Nuclear Power Plant Control Room - A nuclear facility proposes to deploy EEG-based fatigue detection for control room operators during 12-hour night shifts, following a 2024 incident where operator fatigue contributed to a near-miss safety event.

B. Manufacturing Floor "Productivity Optimization" - A manufacturing company proposes continuous EEG monitoring for assembly line workers, claiming it will "prevent repetitive strain injuries" by detecting concentration lapses before errors occur.

C. Aviation Maintenance Technician Monitoring (Hybrid Case) - An airline proposes eye-tracking systems for aircraft maintenance technicians during critical safety inspections, claiming it will ensure thoroughness and prevent oversight errors.

Table 2: Analysis of Neuroergonomic Deployment Scenarios

Scenario	Test 1: Agency	Test 2: Aggregation	Test 3: Consequence	Test 4: Governance	Test 5: Necessity	Final Verdict
A: Nuclear Power Plant	**PASS**: Private vibrotactile alerts; operator retains full decision-making agency.	**PASS**: Individual data deleted post-shift; only anonymized patterns retained.	**PASS**: Non-punitive; triggers backup rotation options rather than penalties.	**PASS**: Co-designed with works council; independent oversight by physician.	**PASS**: Demonstrated that staffing/redesign alone was insufficient.	**Legitimate Safety System**
B: Manufacturing Floor	**FAIL**: Data sent to management dashboard; individual "engagement scores" monitored.	**FAIL**: Individual-level tracking with comparative worker ranking.	**FAIL**: Linked to performance reviews and "corrective action plans."	**FAIL**: Unilateral management decision without meaningful consultation.	**FAIL**: Risks could be mitigated via less invasive ergonomic redesign.	**Prohibited Surveillance**
C: Aviation Maintenance	**PARTIAL PASS**: Provides helpful alerts, but logs data for quality assurance review.	**FAIL**: Individual identification retained for 24 months.	**FAIL**: Data used in annual competency assessments and training requirements.	**PARTIAL PASS**: Council was consulted, but management overrode key concerns.	**PARTIAL PASS**: Verification is needed, but 24-month retention is excessive.	**Hybrid (Requires Modification)**

These applications demonstrate that the five-point test provides clear, operationalizable distinctions. Scenario A shows that legitimate safety monitoring is possible under strict conditions. Scenario B shows that safety framing alone is insufficient - the actual data flows and power dynamics determine legitimacy. Scenario C shows that hybrid systems require modification to meet the standard, providing concrete guidance for compliance.

6. Discussion and Conclusion

This paper argues that "law" and "labor relations" must be considered core components of this interdisciplinary mix, as they form the non-technical governance structure that makes human-centricity possible. The field's current research is described as "limited, fragmented, and inconsistent," lacking even "consensus regarding terminology" (Bach et al. 2024). This legal and conceptual framework provides the codification and clarity that the field urgently needs.

While our five-point test provides a robust interpretive framework, several implementation challenges warrant acknowledgment. Technical complexity creates substantial verification challenges that extend beyond traditional regulatory capabilities. Determining whether a system truly operates with aggregated-only data or whether individual-level tracking is technically embedded requires expertise in data architecture and AI systems (Radanliev 2025). This necessity creates acute capacity-building demands among works councils, labor inspectorates, and data protection authorities who must evaluate these systems without necessarily possessing deep technical expertise. The Platform Directive's provision allowing works councils to engage external AI experts at employer expense provides a procedural mechanism for addressing this knowledge asymmetry (Directive 2024/2831), yet access to qualified AI auditors remains limited, particularly for smaller enterprises operating in specialized safety-critical domains. German works councils have begun to leverage this provision more aggressively following the 2025 amendments to the Works Council Modernization Act, which explicitly guarantees expert access for AI system evaluation (Betriebsratsmodernisierungsgesetz 2021, as amended 2025), but practical implementation reveals significant regional disparities in available expertise.

Cross-border harmonization presents a second layer of difficulty that threatens to undermine the framework's uniform application across the European Union. While GDPR and the AI Act apply EU-wide, national labor law governing co-determination varies significantly in both substance and enforcement culture. German and Austrian works councils enjoy strong veto powers rooted in constitutional protections of worker dignity and codetermination principles (Halefom, 2025), but countries like Ireland, France, and Poland maintain weaker employee representation structures with primarily consultative rather than determinative authority. This creates implementation asymmetry where the same neuroergonomic system might require binding works council agreement in Germany but face only informational consultation requirements in Ireland. Recent research examining Platform Directive implementation suggests that member states are taking divergent approaches to extending algorithmic management protections beyond platform workers, with Germany considering broad application to all "algorithm-controlled work" while other jurisdictions adopt narrow, sector-specific

interpretations. Whether Platform Directive implementation will narrow this co-determination gap or exacerbate existing fragmentation remains an open empirical question requiring longitudinal study as member states complete transposition by December 2026.

The necessity and proportionality test, identified as Test 5 in our framework, inherently involves contextual judgment that resists bright-line rule application. What constitutes "no less invasive alternative" may be legitimately contested by reasonable experts operating in good faith within the same factual scenario. Our framework requires evidence-based risk assessment and documented consideration of alternatives, establishing a procedural standard that shifts burden of proof to system deployers (Foeth, 2026), yet acknowledges that proportionality determinations involve value judgments about acceptable trade-offs between safety gains and privacy intrusions. This suggests the need for sector-specific guidance developed collaboratively by industry associations, labor unions, safety regulators, and technical standards bodies. The International Organization for Standardization's recent publication of ISO/IEC 42001:2023, providing requirements for artificial intelligence management systems, offers a promising foundation for such guidance (ISO, 2023), though the standard currently lacks specific provisions addressing workplace neuroergonomic monitoring in safety-critical contexts. Future standardization work should explicitly incorporate worker dignity protections and co-determination procedural requirements into technical AI management frameworks, creating normative convergence between international technical standards and European fundamental rights jurisprudence.

Enforcement capacity remains perhaps the most significant practical uncertainty surrounding this framework's real-world impact. The AI Act grants member states broad discretion in designating competent authorities and enforcement mechanisms (Regulation 2024/1689), creating potential for fragmented and under-resourced oversight. If AI Act enforcement is institutionally siloed within digital economy agencies without structural coordination mechanisms linking to labor inspectorates and data protection authorities, the multi-layered legal framework we propose may remain conceptually sound but operationally ineffective. Germany's draft AI Market Surveillance Act designates the Federal Network Agency (Bundesnetzagentur) as the primary competent authority but explicitly requires coordination protocols with data protection authorities and labor inspection bodies. Belgium has pioneered an alternative institutional model, creating an integrated AI oversight body with dedicated labor law expertise and formal consultation mechanisms with social partners, avoiding the coordination challenges inherent in multi-agency approaches. These divergent institutional choices will produce natural experiments in enforcement effectiveness that should be studied comparatively to identify best practices. The European AI Office, which holds exclusive enforcement authority over general-purpose AI models, faces

particular challenges balancing its supervisory role with impartiality concerns, as recent analysis has highlighted potential conflicts between its market-enabling and rights-protecting mandates (A&O Shearman, 2025).

This paper's primary contribution is the identification of a critical vulnerability in the EU AI Act and the proposal of a robust, multi-layered framework to resolve it. The "safety exception" for workplace emotion inference (Regulation 2024/1689, Art. 5(1)(f)) represents a well-intentioned but dangerously vague clause that, if left unaddressed, will undermine the very "human-centric" values the EU's 2019 guidelines sought to protect (Tambiama 2019). Left uninterpreted, this exception creates a two-tiered system where workers in "safe" industries receive protection from neuro-surveillance while those in "safety-critical" industries face the most invasive forms of monitoring precisely because they work in high-stakes environments. This outcome would be paradoxical and fundamentally unjust, subjecting workers who already bear significant occupational risks to additional dignitary harms through pervasive cognitive monitoring.

Our proposed test solves this problem by demonstrating that the safety exception need not be a loophole but rather can function as a narrow, rigorously-bounded authorization for legitimate safety interventions. The test balances the undeniable potential of neuroergonomics to improve worker safety (Naranjo et al. 2025) with the fundamental rights of workers to privacy, dignity, and autonomy guaranteed by the Charter of Fundamental Rights of the European Union. It provides a common conceptual language for the engineers designing these systems, the human factors experts evaluating their efficacy, the legal practitioners advising on compliance, and the labor representatives who must govern their workplace implementation. By grounding the AI Act's ambiguities in the established legal strengths of GDPR data protection principles and national co-determination traditions (Foeth, 2026; Halefom, 2025), this framework ensures that "human-centric AI" transcends aspirational rhetoric to become a legally-enforceable operational reality embedded in concrete procedural safeguards and substantive rights.

7. References

Abbas, Qaiser, Woonyoung Jeong, and Seung Won Lee. 2025. “Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges.” *Healthcare* 13 (17): 2154. https://doi.org/10.3390/healthcare13172154.

AI and Employee Data Protection in the European Union: 8 Key Takeaways for Multinational Businesses (2026). https://www.fisherphillips.com/en/news-insights/ai-employee-data-protection-european-union-takeaways-for-multinational-businesses.html.

Amazu, Chidera Winifred, Joseph Mietkiewicz, Ammar N. Abbas, et al. 2024. “Experiment Data: Human-in-the-Loop Decision Support in Process Control Rooms.” *Data in Brief* 53 (February): 110170. https://doi.org/10.1016/j.dib.2024.110170.

A&O Shearman. 2025. “Zooming in on AI – #13: EU AI Act – Focus on Fundamental Rights Impact Assessment for High-Risk AI Systems.” https://www.aoshearman.com/en/insights/ao-shearman-on-tech/zooming-in-on-ai-13-eu-ai-act-focus-on-fundamental-rights-impact-assessment-for-high-risk-ai-systems.

Bach, Tita A., Jenny K. Kristiansen, Aleksandar Babic, and Alon Jacovi. 2024. “Unpacking Human-AI Interaction in Safety-Critical Industries: A Systematic Literature Review.” *IEEE Access* 12: 106385–414. https://doi.org/10.1109/ACCESS.2024.3437190.

BMAS - Betriebsrätemodernisierungsgesetz. Accessed February 13, 2026. https://www.bmas.de/DE/Service/Gesetze-und-Gesetzesvorhaben/betriebsraetemodernisierungsgesetz.html.

Charter of Fundamental Rights of the European Union, 2012/C 326/02.

Directive (EU) 2024/2831 of the European Parliament and of the Council of 23 October 2024 on Improving Working Conditions in Platform Work (Text with EEA Relevance), CONSIL, EP (2024). http://data.europa.eu/eli/dir/2024/2831/oj.

Donisi, Leandro, Giuseppe Cesarelli, Noemi Pisani, Alfonso Maria Ponsiglione, Carlo Ricciardi, and Edda Capodaglio. 2022. “Wearable Sensors and Artificial Intelligence for Physical Ergonomics: A Systematic Review of Literature.” *Diagnostics* 12 (12): 3048. https://doi.org/10.3390/diagnostics12123048.

European Comission. 2025. “Commission Publishes the Guidelines on Prohibited Artificial Intelligence (AI) Practices, as Defined by the AI Act. | Shaping Europe’s Digital Future.” Guidelines. https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

Future of Life Institute. 2024. “High-Level Summary of the AI Act | EU Artificial Intelligence Act.” https://artificialintelligenceact.eu/high-level-summary/.

Google AI. n.d. “AI Principles.” Google AI. Accessed February 13, 2026. https://ai.google/principles/.

Halefom, Abraha. 2025. *Navigating Workers’ Data Rights in the Digital Age | International Labour Organization*. ILO Working Paper 149. https://doi.org/10.54394/MLUH5441.

ISACA. n.d. *White Papers 2024 Understanding the EU AI Act*. Accessed February 13, 2026. https://www.isaca.org/resources/white-papers/2024/understanding-the-eu-ai-act.

ISO. 2023. “ISO/IEC JTC 1/SC 42 - Artificial Intelligence.” ISO, September 11. https://www.iso.org/committee/6794475.html.

Labkoff, Steven, Bilikis Oladimeji, Joseph Kannry, et al. 2024. “Toward a Responsible Future: Recommendations for AI-Enabled Clinical Decision Support.” *Journal of the American Medical Informatics Association* 31 (11): 2730–39. https://doi.org/10.1093/jamia/ocae209.

Maillant, Jordan, Christophe Jallais, and Stéphanie Dabic. 2025. “Detecting Sources of Anger in Automated Driving: Driving-Related and External Factor.” *Frontiers in Neuroergonomics* 6 (May). https://doi.org/10.3389/fnrgo.2025.1548861.

Microsoft. n.d. “Responsible AI Principles and Approach | Microsoft AI.” Accessed February 13, 2026. https://www.microsoft.com/en-us/ai/principles-and-approach.

Muhl, Ekaterina. 2024. “The Challenge of Wearable Neurodevices for Workplace Monitoring: An EU Legal Perspective.” *Frontiers in Human Dynamics* 6 (October). https://doi.org/10.3389/fhumd.2024.1473893.

Naranjo, Jose E., Carlos A. Mora, Diego Fernando Bustamante Villagómez, María Gabriela Mancheno Falconi, and Marcelo V. Garcia. 2025. "Wearable Sensors in Industrial Ergonomics: Enhancing Safety and Productivity in Industry 4.0." *Sensors* 25 (5): 1526. https://doi.org/10.3390/s25051526.

Pape, Marketa. 2024. "Addressing AI Risks in the Workplace." European Parliamentary ResearchService.

Radanliev, Petar. 2025. "Privacy, Ethics, Transparency, and Accountability in AI Systems for Wearable Devices." *Frontiers in Digital Health* 7 (June). https://doi.org/10.3389/fdgth.2025.1431246.

Ramos, Inês F., Gabriele Gianini, Maria Chiara Leva, and Ernesto Damiani. 2024. "Collaborative Intelligence for Safety-Critical Industries: A Literature Review." *Information* 15 (11): 728. https://doi.org/10.3390/info15110728.

Schleiger, Emma, Claire Mason, Claire Naughtin, Andrew Reeson, and Cecile Paris. 2024. "Collaborative Intelligence: A Scoping Review Of Current Applications." *Applied Artificial Intelligence* 38 (1): 2327890. https://doi.org/10.1080/08839514.2024.2327890.

Tambiama, MADIEGA. 2019. *EU Guidelines on Ethics in Artificial Intelligence: Context and Implementation*. BRIEFING. European Parliamentary Research Service.

UNESCO. 2021. *Ethics of Artificial Intelligence*. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.

Vieira, Rodrigo, Plinio Moreno, and Athanasios Vourvopoulos. 2024. "EEG-Based Action Anticipation in Human-Robot Interaction: A Comparative Pilot Study." *Frontiers in Neurorobotics* 18 (December): 1491721. https://doi.org/10.3389/fnbot.2024.1491721.