Responsibility, Habit, and Control: Digital Humanism and the Delegation of Critical Functions to Intelligent Autonomous Systems

Gordana Dodig-Crnkovic *

Posted Date: 14 January 2026

doi: 10.20944/preprints202601.1095.v1

Keywords: autonomous systems; responsibility; digital humanism

 doi:10.20944/preprints202601.1095.v1

Article

Responsibility, Habit, and Control: Digital Humanism and the Delegation of Critical Functions to Intelligent Autonomous Systems

Gordana Dodig-Crnkovic [1,2]

[1] Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden; dodig@chalmers.se

[2] Mälardalen University, Västerås, Sweden

Abstract

As intelligent autonomous systems (IAS) continue to assume increasingly central roles in safety- and mission-critical domains such as transportation, healthcare, finance, and infrastructure management, humans are becoming unable to monitor or intervene in real time. This shift is driven by the speed, data-processing capacity, and adaptivity of IAS. To manage this complexity, a new paradigm is emerging: IAS controlling and monitoring other IAS, a development that introduces at the same time practical efficiency and profound practical and ethical challenges. This article explores the multi-layered delegation of responsibilities within IAS ecosystems, where decisions influencing human lives and well-being are made with minimal human intervention. One often-overlooked consequence of this delegation is the capacity of AI systems to shape and create new human habits, whether through personalized persuasion, behavioral feedback loops, or autonomous decision enforcement. As humans increasingly adapt their behaviors to machine-optimized environments, questions arise about autonomy, agency, and responsibility for resulting behavior changes. Drawing on insights from recent research on responsibility delegation in IAS and on AI-driven habit formation, the article critically examines how responsibility should be distributed across human actors, autonomous systems, and institutions. Framed within the principles of Digital Humanism, I argue for a value-sensitive governance model that ensures transparency, explainability and human oversight even in complex IAS-to-IAS control scenarios. I propose a normative framework for responsibility attribution that accounts for both the technical architecture of IAS networks and the behavioral effects these systems have on human users. The article concludes by addressing the ethical risks of diminished human agency, manipulation through behavioral design, and the need for institutional mechanisms that align IAS operations with fundamental human values.

Keywords: autonomous systems; responsibility; digital humanism

1. Introduction

In recent years, the delegation of critical functions to Intelligent Autonomous Systems (IAS) has expanded rapidly across sectors such as healthcare, transportation, finance, and infrastructure management. Driven by their speed, data-processing capacity, and decision-making capabilities, IAS increasingly operate in regimes where human monitoring and intervention are no longer possible in real time (Dodig-Crnkovic, Basti, & Holstein 2025; Stahl 2021). This leads to growing reliance on autonomous decision-making and poses new challenges for ethical governance, responsibility attribution, and human agency.

A particularly striking development is the emergence of IAS-to-IAS control mechanisms, where autonomous systems are tasked with monitoring, regulating, and making decisions about the behavior of other IAS. While this layered control architecture improves efficiency and reduces error rates in complex environments (Gawer & Cusumano 2014), it introduces profound ethical and legal

 doi:10.20944/preprints202601.1095.v1

questions regarding accountability, transparency, and value alignment (Floridi & Cowls 2019; Morley, Floridi, Kinsey, & Elhalal 2020).

Simultaneously, the behavioral consequences of human interaction with IAS deserve closer scrutiny. Intelligent systems are not only performing delegated tasks but also actively shaping human behavior and creating new habits through personalized feedback and algorithmic recommendation loops (Zuboff 2019). These mechanisms can subtly alter users' cognitive patterns, routines, and decision-making strategies—raising concerns about possibility of manipulation, behavioral lock-in, and the erosion of human autonomy (Binns 2018; Habermas 1984).

In this context, Digital Humanism offers an ethically grounded framework for addressing these challenges. Digital Humanism emphasizes human dignity, agency, and the social embedding of technology development (Vienna Manifesto on Digital Humanism 2019). Its principles advocate for value-sensitive design approaches that prioritize human well-being, informed consent, and institutional accountability, even within highly automated socio-technical systems (Spiekerman 2023).

This article brings together insights from recent work on responsibility delegation in IAS (Dodig-Crnkovic et al. 2025), the behavioral impacts of AI-driven habit formation, and the ethical governance of autonomous ecosystems (Stahl 2021; Floridi & Cowls 2019). By doing so, it aims to answer the following research questions:

1. How should responsibility for critical decisions be distributed among designers, operators, IAS themselves, and institutional frameworks?
2. What are the ethical implications of IAS shaping and creating human habits?
3. How can Digital Humanism guide the design, governance, and oversight of IAS-to-IAS control structures in ways that safeguard human agency and social values?

Through an interdisciplinary analysis, the article proposes a normative framework for multi-level responsibility attribution that accounts for both technical system complexity and behavioral effects on human users.

2. Background and Context

2.1. The Rise of Intelligent Autonomous Systems in Critical Infrastructure

The deployment of Intelligent Autonomous Systems (IAS) in critical technological domains is transforming how decision-making, and control functions are managed. IAS are now central actors in fields as diverse as autonomous transportation, medical diagnostics, industrial process control, and financial trading (Stahl 2021). Their ability to process vast amounts of data at speeds and scales beyond information processing capacities makes them essential for managing real-time, high-stakes environments (Dodig-Crnkovic et al. 2025).

However, this shift is not without significant trade-offs. Human cognitive limitations mean that operators cannot meaningfully intervene in many IAS-driven processes, when decisions are made within milliseconds or based on multi-dimensional sensor input (Gawer & Cusumano 2014). As a result, humans are forced to delegate both operational control and decision-making responsibility to IAS, trusting in the system's capacity for accuracy, reliability, and ethical behavior.

2.2. IAS-to-IAS Control: From Human Oversight to Machine Governance

A direct consequence of the growing complexity of intelligent autonomous system (IAS) networks is the emergence of IAS-to-IAS control structures, in which one autonomous system monitors, constrains, or corrects another. For example, in traffic domains, vehicle "driver agents" request reservations from an AI "intersection manager" that grants or denies passage to ensure safety—an explicit AI-over-AI coordination regime (Au, Zhang, and Stone 2010). In finance, regulatory and supervisory AI increasingly oversees algorithmic trading by detecting and

 doi:10.20944/preprints202601.1095.v1

responding to market-abuse patterns, thereby placing AI-driven surveillance in a controlling relationship to AI-driven trading (IOSCO 2025). Importantly, machine learning is now embedded in a large share of trading algorithms themselves, sharpening the sense in which AI supervises AI (AFM 2023).

While these architectures introduce significant efficiencies and reduce latency, they also create new accountability challenges. When a cascade of decisions is made by multiple interlinked IAS, identifying causal responsibility in the event of system failure becomes non-trivial (Calo 2015; Latour 1992). This is particularly problematic when decisions involve outcomes—such as life-or-death judgments (Dodig-Crnkovic et al. 2025).

Moreover, explainability and transparency remain technical challenges. Many IAS employ opaque machine learning models, "black boxes", making it difficult for external stakeholders (users, regulators, or designers) to reconstruct how or why a specific decision was made (Morley et al. 2020).

2.3. Behavioral Consequences: Intelligent Autonomous Systems Shaping Human Habits

Beyond operational control and decision-making, IAS more and more shape human behaviors and create new user habits. AI-driven systems embedded in everyday apps and digital services engage users through continuous feedback loops, personalized nudging, and adaptive behavior reinforcement mechanisms. These systems alter users' cognitive and behavioral patterns over time, sometimes supporting positive change (e.g., health apps encouraging exercise) but also influencing users for commercial or operational goals (Zuboff 2019).

The ethical problem intensifies when IAS-to-IAS control systems enforce behavioral adjustments at scale, such as recommendation engines influencing user choices across millions of individuals simultaneously. The concern is not only the loss of individual autonomy but also the emergence of new social norms and collective habits shaped by AI-driven governance mechanisms (Binns 2018).

2.4. The Need for a Digital Humanism Framework

Given these developments, Digital Humanism emerges as a central normative approach for understanding and addressing the challenges of IAS control and human behavioral impact. The Vienna Manifesto on Digital Humanism (2019) calls for technology design and governance that foregrounds human agency, dignity, and societal values.

Digital Humanism insists on embedding ethical principles directly into system architectures while ensuring that accountability remains with human and institutional actors. This is particularly critical when humans are excluded from operational decision loops, and when IAS increasingly function as behavior-shaping agents.

In the following sections, we will explore how responsibility can be ethically distributed across human and non-human agents in IAS ecosystems, and how Digital Humanism can guide the design of systems that both enhance operational efficiency and protect human values.

3. Responsibility Delegation in Intelligent Autonomous Systems: Conceptual Foundations

Responsibility in complex socio-technical systems like IAS is a multi-dimensional concept. It includes causal, moral, legal, and functional domains (Dodig-Crnkovic, Basti, & Holstein 2025). Traditional ethical models often assume clear lines of human control and agency, where responsibility is attributed based on direct causality and intentionality (Habermas 1984). However, the operational reality of IAS, especially in high-speed, machine-to-machine environments, challenges these assumptions. Key forms of responsibility include:

Causal responsibility: Who or what caused an event or outcome?

Legal responsibility: Who is legally liable for the outcome?

 doi:10.20944/preprints202601.1095.v1

Functional responsibility: Which system component or agent was functionally tasked with the relevant action?

In IAS contexts, functional responsibility becomes central. While IAS lack moral agency, they carry task-based responsibilities within the operational system architecture (Dodig-Crnkovic et al. 2025). This framing helps to track, document, and analyze machine behaviors for oversight and accountability purposes.

3.1. Delegation Chains and Responsibility Distribution

Modern IAS operate within delegation chains, where responsibilities move across multiple layers: from designers, programmers, and operators to the IAS themselves—and now increasingly from IAS to other IAS.

For example, in autonomous traffic control, human engineers design algorithms that regulate vehicle behavior, but during real-time operations, decision authority passes from vehicles to traffic management IAS, all without direct human oversight. Each link in this chain carries different types of responsibility:

Design Responsibility includes program developers and architects,

Operational Responsibility concerns system operators and supervisors,

Functional Execution Responsibility includes the IAS itself,

Oversight Responsibility concerns institutional bodies (e.g., regulators).

Recognizing and mapping these responsibility layers is essential for legal liability, ethical oversight, and user trust.

3.2. Intelligent Autonomous Systems as Functional Agents

A central argument by Dodig-Crnkovic et al. (2025) is that IAS should be viewed as autonomous functional agents—entities with task-based roles, but not moral agents in the human sense. This distinction is critical.

While IAS can autonomously select actions, they do so based on programmed goals, learned patterns, and predefined constraints. They do not possess moral reasoning, intentionality, or the human-like capacity for ethical deliberation (Latour 1992).

However, their causal impact on morally relevant outcomes makes it necessary to design accountability structures around their actions. This includes tools for logging decision pathways, implementing explainability mechanisms (Morley et al. 2020), and embedding ethical constraints and value alignment protocols (Floridi & Cowls 2019).

3.3. Emerging Challenges in IAS-to-IAS Responsibility Flow

When IAS delegate control to other IAS responsibility chains become non-linear and layered, making it difficult to reconstruct decision causality after system failures or unexpected behaviors (Calo 2015).

Key emerging challenges include opacity (decisions made through multiple opaque AI models), traceability loss (difficulty in reconstructing who or what "caused" a specific outcome), responsibility dilution (when multiple IAS are involved, human actors may feel less accountable as a "many hands" problem), and autonomous behavior shaping (when IAS shape human habits, responsibility for behavioral outcomes becomes even more diffuse).

These issues necessitate new governance models and design methodologies that explicitly account for both technical causality and moral responsibility, while respecting human dignity and autonomy in line with Digital Humanism (Vienna Manifesto on Digital Humanism 2019).

3.4. AI as a Habit-Creating Force: Behavioral Implications

 doi:10.20944/preprints202601.1095.v1

While much of the ethical discourse on Intelligent Autonomous Systems (IAS) focuses on decision-making, control, and accountability at the system level, an important but often unnoticed dimension is the role of IAS in shaping human behavior.

IAS are no longer just passive tools or decision-support systems. They actively influence human habits and behavioral patterns, typically without explicit user awareness or consent (Zuboff 2019). Through reinforcement mechanisms, adaptive feedback, and personalized nudging, IAS systems can condition users into forming new habits, whether beneficial, neutral, or harmful for themselves and for society.

3.5. Mechanisms of AI-Driven Habit Formation

AI systems engage in habit shaping using multiple techniques, such as personalized nudging (adaptive prompts designed to steer user behavior toward desired outcomes) (Binns 2018); feedback loops (real-time performance tracking and behavioral reinforcement); behavioral lock-in (repetitive exposure to AI-curated choices that limit behavioral diversity) (Zuboff 2019) and choice architecture manipulation (structuring options in ways that subtly direct user decisions) (Floridi & Cowls 2019).

These technologically sophisticated mechanisms raise concerns about autonomy, consent and manipulation, especially when deployed at scale without user knowledge.

3.6. Ethical Risks: Manipulation and Erosion of Autonomy

The behavioral influence of intelligent autonomous systems poses significant ethical risks, particularly when users are unaware of how their habits are being shaped or when the system's goals do not align with users' best interests. Potential risks include:

Loss of autonomy. Users may find themselves engaging in behaviors they did not consciously choose (Habermas 1984).

Informed consent violations. Behavioral nudges are often embedded invisibly, bypassing traditional consent mechanisms (Morley et al. 2020).

Behavioral lock-in. Algorithmic reinforcement may lead users into rigid behavioral loops, limiting freedom of choice (Zuboff 2019).

Ethical drift. When IAS-to-IAS control structures propagate behavioral nudging at scale, small design biases may amplify across populations (Stahl 2021).

These issues become even more complex when IAS-to-IAS governance layers themselves reinforce or correct other IAS-generated behavioral nudges, introducing a cascade of behavioral influence mechanisms with limited human oversight.

3.7. Potential Opportunities: Supporting Beneficial Habits

Despite these risks, IAS also hold promise for positive behavior change, especially in health, education, and sustainability contexts. Examples include health behavior change apps encouraging exercise, healthy eating, and medication adherence; sustainability nudges: promoting energy-saving habits through smart home IAS and learning support systems, personalized educational tools that reinforce effective study habits.

The ethical imperative, however, is to ensure that habit-shaping interventions respect human agency, are transparent, and align with user values, in keeping with Digital Humanism principles (Vienna Manifesto on Digital Humanism 2019).

To reconcile the behavioral power of IAS with the ethical commitments of Digital Humanism, we propose the following key design and governance requirements:

Transparency (users should know when and how they are being influenced).

 doi:10.20944/preprints202601.1095.v1

Consent and control (users should have meaningful ways to opt out).

Value Alignment (human dignity, autonomy, and well-being are prioritized).

Oversight in IAS-to-IAS chains (understanding IAS decision-makers and behavioral agents necessitates responsibility frameworks that address both systemic control and individual behavioral impact, a theme developed further in the next section).

4. Digital Humanism as an Ethical Framework

Digital Humanism emerges as a critical and timely response to the ethical, social, and political challenges posed by the growing autonomy of Intelligent Autonomous Systems (IAS). Rooted in the belief that technology must serve human dignity, agency, and societal well-being, Digital Humanism emphasizes the centrality of human values in the design, deployment, and governance of digital technologies (Vienna Manifesto on Digital Humanism 2019) and calls for:

Human-centric design (human needs and rights over technological efficiency)

Value-sensitive innovation (embedding societal, cultural, and ethical values into technical systems)

Transparency and accountability (ensuring that decision-making processes, both human and machine-based, remain understandable and traceable)

Democratic control (advocating for collective societal oversight over the development and use of digital technologies).

4.1. Digital Humanism and Responsibility Distribution

The delegation of responsibilities to IAS, especially in IAS-to-IAS control structures, presents a direct challenge to these principles. As (Dodig-Crnkovic et al. 2025) point out, responsibility becomes distributed, diffuse, and difficult to track in multi-layered, autonomous systems.

From a Digital Humanism perspective, this delegation must not undermine human accountability (ultimate moral and legal responsibility must remain with human actors and institutions) (Stahl 2021); human agency and control (even when IAS operate autonomously, long-term mechanisms for human oversight and intervention must exist); and transparency (the logic, goals, and value trade-offs embedded in IAS operations must be explainable and auditable) (Morley et al. 2020).

In practice, this requires not only technical design interventions (such as explainable AI and audit trails) but also institutional arrangements for oversight, governance, and legal redress (Calo 2015).

4.2. Behavioral Impacts Within Digital Humanism

Importantly, Digital Humanism also extends to the behavioral dimension of IAS operations. AI systems shaping human habits cannot be ethically neutral. Their capacity to influence cognition and behavior means that behavioral design becomes a site of ethical responsibility.

Applying Digital Humanism here means ensuring that behavioral interventions are aligned with human autonomy and well-being, requiring informed consent for behavior-shaping processes, preventing covert manipulation or commercial exploitation of user behavior patterns (Zuboff 2019), and promoting reflection and user empowerment, rather than behavioral lock-in (Habermas 1984).

This perspective aligns with calls for value-sensitive design and ethical AI governance frameworks emerging in the broader AI ethics community (Floridi & Cowls 2019; Morley et al. 2020; Spiekermann 2023).

4.3. Digital Humanism-Informed Governance Model

 doi:10.20944/preprints202601.1095.v1

To operationalize these principles in the context of IAS responsibility delegation and habit formation, multi-level governance models are needed. These should integrate:

Design-time ethics - technical safeguards,

Run-time oversight - organizational policies,

Post-hoc accountability - regulatory and legal frameworks,

Informed consent for use - user-centric behavioral governance.

A robust socio-technical system is needed supporting the above (education, infrastructure, etc.)

5. Toward a Normative Framework for Responsibility Sharing

5.1. Multi-Level Responsibility Attribution

Given the technical complexity and behavioral impact of IAS, particularly in IAS-to-IAS control environments, a single-layered concept of responsibility is no longer sufficient. Instead, responsibility must be understood as distributed across multiple human and institutional actors, alongside the functional roles assigned to IAS themselves (Dodig-Crnkovic, Basti, & Holstein 2025).

Building on Digital Humanism (Werthner, Prem, Lee, & Ghezzi 2022; Nida-Rümelin & Weidenfeld 2018), we propose a multi-level responsibility model, which includes:
Design responsibility: Engineers, software developers, and data scientists hold responsibility for embedding ethical constraints, explainability mechanisms, and safety layers within IAS.

Operational responsibility: System operators and organizational users are responsible for monitoring system performance and intervening where necessary.

Institutional responsibility: Organizations and regulatory bodies must provide oversight, establish clear accountability structures, and enforce compliance with ethical standards.

Functional responsibility: IAS are assigned specific operational roles, tracked through audit trails and decision logs, but remain non-moral agents (Latour 1992) (Dodig-Crnkovic et al. 2025).

Behavioral Responsibility: Designers and deployers of habit-shaping AI bear responsibility for assessing and mitigating potential manipulation and autonomy loss (Zuboff 2019).

This layered model helps address responsibility gaps inherent in complex socio-technical systems (Calo 2015).

5.2. Ethical Design Recommendations

To operationalize responsibility sharing, the following value-sensitive design principles, aligned with Digital Humanism (Werthner et al. 2022; Werthner, et al. 2024; Spiekermann 2023) can be used:
Transparency. AI systems, especially those in IAS-to-IAS governance structures, must include mechanisms that allow decision-making processes to be inspected, audited, and explained (Floridi & Cowls 2019; Morley et al. 2020).

Ethical constraints and value alignment. IAS should be designed, architected and programmed with ethical boundaries, ensuring that autonomous decisions cannot violate human rights or ethical norms (Nida-Rümelin & Weidenfeld 2018; Spiekermann 2023).

Behavioral safeguards. AI-driven habit formation processes should include clear behavioral intention disclosures, user opt-outs, and mechanisms for user self-reflection and agency preservation.

 doi:10.20944/preprints202601.1095.v1

Oversight of IAS-to-IAS interactions. Institutional bodies and human supervisors must maintain systemic oversight over how IAS monitor and control each other. This includes meta-governance systems, where dedicated AI governance layers track and regulate IAS-to-IAS interactions (Stahl 2021).

5.3. Managing Responsibility in Behavior-Shaping Systems

A particularly novel responsibility challenge emerges from IAS-driven behavioral shaping. Given that IAS can create or reinforce human habits, responsibility for user behavior is no longer limited to individual free will but becomes co-produced between human users and machine agents (Binns 2018; Dodig-Crnkovic 2025).

Digital Humanism insists that habit-forming interventions must respect human autonomy and dignity. This requires explicit user awareness so that users know when behavior-shaping algorithms are active; ethical goal alignment where behavioral nudges align with users' self-identified goals and societal values (Spiekermann 2023), as well as institutional accountability implying that organizations deploying behavioral AI must be auditable and legally accountable for long-term behavioral impacts (Nida-Rümelin & Weidenfeld 2018).

5.4. Systemic Governance and Institutional Roles

Finally, responsibility sharing cannot stop at the level of system design or individual user interaction. It requires robust institutional and legal frameworks (Werthner et al., 2022; Stahl, 2021). Governments, regulatory bodies, and international organizations must develop clear liability models for harm caused by IAS decisions, ethical AI certification systems, behavioral impact assessments for habit-shaping technologies and oversight mechanisms for IAS-to-IAS governance architectures.

These governance structures must reflect the core Digital Humanism commitment: technology must remain subservient to human values, democratic principles, and societal well-being.

6. Discussion

The growing delegation of critical functions to IAS—and the increasing reliance on IAS-to-IAS control structures—presents an inevitable trade-off between operational efficiency and ethical control.

On one hand, IAS bring undeniable benefits: improved decision speed, consistency, scalability, and error reduction in complex systems (Stahl 2021). On the other hand, as functional control shifts away from humans, traditional ethical oversight mechanisms struggle to keep pace (Dodig-Crnkovic et al. 2025).

A central tension is how human ethical responsibility can remain meaningful when human intervention is technologically impossible in real time? Digital Humanism helps resolve this tension by insisting that responsibility is not a function of proximity to decision points but of design-time and institutional control structures (Werthner, Prem, Lee, & Ghezzi 2022; Nida-Rümelin & Weidenfeld 2018).

6.1. The Complexity of Distributed Responsibility

IAS environments are characterized by distributed, layered, and non-linear responsibility flows (Floridi & Cowls 2019). Responsibility is not concentrated in any single actor but spread across designers, developers, software architects, system operators and users, IAS components performing functional roles, and institutional regulators.

This distribution makes post-hoc accountability attribution difficult, especially when system failures emerge from unforeseen interactions between diverse human and non-human autonomous agents (Calo 2015).

 doi:10.20944/preprints202601.1095.v1

One emerging solution is the implementation of responsibility mapping tools during system design and deployment phases, which can document who holds which form of responsibility at each layer (Morley et al. 2020).

A particularly urgent context for addressing responsibility delegation is the use of IAS in military operations, especially the deployment of autonomous weapons systems and drone warfare. Here, decision cycles are compressed to milliseconds, often necessitating IAS-to-IAS control hierarchies for threat detection, targeting, and engagement (Calo 2015; Dodig-Crnkovic et al. 2025).

The ethical risks are profound: lethal autonomous decisions, lack of human oversight, unclear responsibility attribution in case of civilian harm, and violations of international humanitarian law. Critics argue that the use of autonomous weapons fundamentally undermines human dignity and moral accountability, contradicting core principles of Digital Humanism (Nida-Rümelin & Weidenfeld 2018; Werthner et al. 2022).

Emerging research on "meaningful human control" and calls for a global ban on fully autonomous lethal systems highlight the need for institutional responsibility frameworks that prevent ethical erosion in this domain (Floridi & Cowls 2019). Warfare thus presents a stress-test case for any proposed model of distributed responsibility and ethical AI governance.

6.2. The Ethical Significance of AI-Driven Habit Formation

A critical and under-examined dimension of responsibility is the behavioral influence IAS exert on human users. Habit formation, driven by AI feedback loops and personalized nudging, is not ethically neutral. These systems can enhance well-being (e.g., by promoting healthy behaviors) or undermine autonomy (e.g., by locking users into behavior patterns that serve system owners' interests) (Zuboff 2019).

Digital Humanism provides a normative stance here: human autonomy, informed consent, and dignity must be preserved, even in systems designed for large-scale behavior shaping (Werthner, Prem, Lee & Ghezzi 2022). This raises variety of new questions, among others:

Who is responsible when a user adopts a harmful habit due to IAS nudging?

How can unintended behavioral side-effects be ethically managed?

Should there be behavioral impact assessments for habit-shaping systems (analogous to environmental or data protection impact assessments)?

6.3. Toward Institutional and Legal Innovation

Given the complexity of IAS ecosystems, ethical responsibility frameworks must extend beyond individual actors to include institutions and legal systems (Stahl 2021; Nida-Rümelin & Weidenfeld 2018). Possible approaches include AI-specific regulatory bodies with the authority to audit IAS behaviors, mandatory transparency and explainability standards and behavioral ethics regulations, focusing on user autonomy, manipulation risks, and consent in habit-formation technologies.

These institutional mechanisms should be explicitly designed to align with Digital Humanism principles, ensuring that technological progress does not erode core human rights and democratic values (Vienna Manifesto on Digital Humanism 2019).

6.4. Future Research Directions

This article identifies the following key areas for future research:

Multi-level multi-agent responsibility modeling. Development of tools for mapping and tracking responsibility flows in IAS networks

Behavioral governance. Empirical research on how AI-driven habit formation affects user autonomy over time

 doi:10.20944/preprints202601.1095.v1

IAS-to-IAS oversight mechanisms. Technical research into self-regulating and meta-governance architectures within IAS ecosystems

Legal innovation. New legal frameworks capable of addressing distributed and non-human agency in IAS operations

Ethics of machine-to-machine behavior shaping. Deeper exploration of how IAS-to-IAS interactions can indirectly shape human behavior at scale.

7. Conclusion

The accelerating delegation of critical functions to Intelligent Autonomous Systems (IAS), combined with the emergence of IAS-to-IAS control structures, presents a fundamental shift in how responsibility, agency, and decision-making are distributed in socio-technical systems. This transformation challenges long-standing ethical, legal, and philosophical frameworks that assume direct human control over all morally significant decisions.

This article has highlighted two interconnected dimensions of this transformation: first, the technical and legal complexities of responsibility delegation within increasingly autonomous IAS ecosystems; and second, the neglected behavioral impacts resulting from AI-driven habit formation in human users. Both dimensions pose serious ethical risks—including opacity, accountability gaps, loss of human autonomy, and manipulation through behavioral design.

Framed within the principles of Digital Humanism, we argue that technological efficiency cannot come at the cost of human dignity, agency, and societal well-being. Responsibility must remain traceable, distributed across human actors and institutions, and embedded within system design through value-sensitive architectures, transparent decision-making processes, and ethical oversight mechanisms (Spiekermann 2023).

Particularly in high-stakes domains like military operations and autonomous warfare, the need for ethical governance becomes even more urgent. The deployment of autonomous weapons systems and decision-making drones exemplifies the extreme risks of responsibility erosion, necessitating strong international legal frameworks, institutional accountability, and a global ethical consensus guided by Digital Humanism.

Ultimately, the rise of IAS capable of influencing both system-level operations and individual human habits demands a new normative framework for responsibility sharing, one that is both technically realistic and ethically grounded. This requires interdisciplinary collaboration between technologists, ethicists, behavioral scientists, policymakers, and affected communities.

As IAS continue to reshape not only our technologies but also our behaviors and social norms, we must ensure that technology serves humanity and not the other way around. Digital Humanism provides the ethical roadmap for navigating this transformation with care, responsibility, and foresight.

Acknowledgments. The author would like to thank the reviewers for their constructive feedback and valuable comments.

Disclosure of Interests. The author has no competing interest to declare that is relevant to the content of this article.

 doi:10.20944/preprints202601.1095.v1

References

1. AFM (The Netherlands Authority for the Financial Markets) (2023) *Machine learning in algorithmic trading: Application by Dutch proprietary trading firms and possible risks* (Report). Available at: https://www.afm.nl/~/profmedia/files/rapporten/2023/report-machine-learning-trading-algorithms.pdf
2. Au, T.-C., Zhang, S., & Stone, P. (2010) Motion planning algorithms for autonomous intersection management. In *AAAI Workshop on Bridging the Gap Between Task and Motion Planning*. Available at: https://cdn.aaai.org/ocs/ws/ws0611/2053-8456-1-PB.pdf
3. Binns, R. (2018) Fairness in machine learning: Lessons from political philosophy. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, vol. 81, pp. 149–159. Available at: https://proceedings.mlr.press/v81/binns18a.html
4. Calo, R. (2015) Robotics and the lessons of cyberlaw. *California Law Review*, 103(3), 513–563. http://dx.doi.org/10.2139/ssrn.2402972
5. Çürüklü, B., Dodig-Crnkovic, G., & Akan, B. (2010) Towards industrial robots with human-like moral responsibilities. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction* (p. 85). Osaka, Japan, March 2–4, 2010. http://dx.doi.org/10.1145/1734454.1734484
6. Dennett, D.C. (1973) Mechanism and responsibility. In Honderich, T. (Ed.), *Essays on Freedom of Action* (pp. 143–163). Boston: Routledge & Kegan Paul.
7. Dodig-Crnkovic, G. (2006) Professional ethics in computing and intelligent systems. In *Proceedings of the Ninth Scandinavian Conference on Artificial Intelligence (SCAI 2006)* (pp. 11–18). Espoo, Finland, October 25–27, 2006.
8. Dodig-Crnkovic, G., & Persson, D. (2008) Sharing moral responsibility with robots: A pragmatic approach. In Holst, A., Kreuger, P., & Funk, P. (Eds.), *Tenth Scandinavian Conference on Artificial Intelligence (SCAI 2008). Frontiers in Artificial Intelligence and Applications*, vol. 173, pp. 165–168. Amsterdam: IOS Press.
9. Dodig-Crnkovic, G., Basti, G., & Holstein, T. (2025) Delegating responsibilities to intelligent autonomous systems: Challenges and benefits. *Journal of Bioethical Inquiry*. Advance online publication. http://dx.doi.org/10.1007/s11673-025-10428-5
10. Floridi, L., & Cowls, J. (2019) A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). http://dx.doi.org/10.1162/99608f92.8cd550d1
11. Floridi, L., & Sanders, J.W. (2004) On the morality of artificial agents. *Minds and Machines*, 14, 349–379. http://dx.doi.org/10.1023/B:MIND.0000035461.63578.9d
12. Gawer, A., & Cusumano, M.A. (2014) Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, 31(3), 417–433. http://dx.doi.org/10.1111/jpim.12105
13. Habermas, J. (1984) The theory of communicative action: Reason and the rationalization of society, vol. 1. Boston, MA: Beacon Press.
14. IOSCO (International Organization of Securities Commissions) (2025) *Artificial Intelligence in Capital Markets: Use Cases, Risks, and Challenges* (CR/01/2025). Available at: https://www.iosco.org/library/pubdocs/pdf/IOSCOPD788.pdf
15. Latour, B. (1992) Where are the missing masses? The sociology of a few mundane artifacts. In Bijker, W., & Law, J. (Eds.), *Shaping Technology / Building Society* (pp. 225–258). Cambridge, MA: MIT Press.
16. Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020) From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), 2141–2168. http://dx.doi.org/10.1007/s11948-019-00165-5
17. Nida-Rümelin, J., & Weidenfeld, N. (2023) *Digital Humanism: For a humane transformation of democracy, economy and culture in the digital age*. Cham: Springer. http://dx.doi.org/10.1007/978-3-031-12482-2
18. Spiekermann, S. (2023) Value-Based Engineering: A Guide to Building Ethical Technology for Humanity. Berlin: Walter de Gruyter.
19. Stahl, B.C. (2021) *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*. Cham: Springer. http://dx.doi.org/10.1007/978-3-030-69978-9
20. Vienna Manifesto on Digital Humanism. (2019). *Digital Humanism Initiative*. Vienna: Vienna University of Technology. Available at: https://caiml.org/dighum/

 doi:10.20944/preprints202601.1095.v1

21. Werthner, H., Ghezzi, C., Kramer, J., Nida-Rümelin, J., Nuseibeh, B., Prem, E., & Stanger, A. (Eds.). (2024). *Introduction to Digital Humanism: A Textbook*. Cham: Springer. http://dx.doi.org/10.1007/978-3-031-45304-5
22. Werthner, H., Prem, E., Lee, E.A., & Ghezzi, C. (Eds.). (2022). *Perspectives on Digital Humanism*. Cham: Springer. http://dx.doi.org/10.1007/978-3-030-86144-5
23. Zuboff, S. (2019) The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs.