Received: 28.08.2025 | Revised: 31.10.2025 | Accepted: 09.01.2026

Artificial Intelligence as a Socio-Economic Dilemma: Ordonomic Diagnosis–Reflection–Design for Education, Work and Governance

Patrick Hedfeld

Abstract: This paper analyses artificial intelligence (AI) through the lens of *ordonomics*, a normative-institutional approach that connects economic rationality with ethical reflection. While most discussions in AI ethics focus on principles such as fairness, transparency, and accountability, fewer studies address how these principles can be institutionalized through incentive-compatible rules. We therefore conceptualize AI not as a primarily technological challenge but as a social order problem that requires institutional design and governance. The paper explicitly maps the classical ordonomic three-level schema—actor, institutional order, and market/discourse—onto an applied heuristic of Diagnosis–Reflection–Design, demonstrating how this triad operationalizes ordonomic reasoning for the AI context. Building on this foundation, we identify and categorize key AI-related social dilemmas (economic, epistemic, ethical, and educational). The analysis develops differentiated responsibilities across levels of coordination and proposes rule-based cooperation solutions that align individual incentives with collective welfare. By linking ordonomics to current frameworks such as *Responsible AI*, *algorithmic accountability*, and the *EU AI Act*, the paper positions ordonomics as a design-oriented ethics that bridges normative ideals and institutional economics. The result is a framework for diagnosing conflicts, reflecting responsibilities, and designing cooperative solutions that reconcile innovation with social responsibility.

Keywords: Ordonomics; social dilemmas; artificial intelligence; Diagnosis–Reflection–Design; moral paradox; institutional design; AI governance

Künstliche Intelligenz als sozioökonomisches Dilemma: Eine ordonomische Diagnose–Reflexion–Design Struktur für Bildung, Arbeit und Governance

Zusammenfassung: Dieser Artikel analysiert Künstliche Intelligenz (KI) aus der Perspektive der Ordonomik, einem normativ-institutionellen Ansatz, der ökonomische Rationalität mit ethischer Reflexion verbindet. Während sich die meisten Diskussionen in der KI-Ethik auf Prinzipien wie Fairness, Transparenz und Rechenschaftspflicht konzentrieren, befassen sich weniger Studien mit der Frage, wie diese Prinzipien durch anreizkompatible Regeln institutionalisiert werden können. Wir konzeptualisieren KI daher nicht als primär technologische Herausforderung, sondern als ein Problem der sozialen Ordnung, das institutionelle Gestaltung und Governance erfordert. Der Artikel bildet das klassische ordonomische Drei-Ebenen-Schema – Akteur, institutionelle Ordnung und Markt/Diskurs – explizit auf eine angewandte Heuristik aus Diagnose, Reflexion und Design ab und zeigt, wie diese Triade ordonomisches Denken im KI-Kontext operationalisiert. Aufbauend auf

dieser Grundlage identifizieren und kategorisieren wir zentrale KI-bezogene soziale Dilemmata (wirtschaftliche, epistemische, ethische und pädagogische). Die Analyse entwickelt differenzierte Verantwortlichkeiten auf verschiedenen Koordinationsebenen und schlägt regelbasierte Kooperationslösungen vor, die individuelle Anreize mit dem Gemeinwohl in Einklang bringen. Durch die Verknüpfung der Ordonomik mit aktuellen Rahmenwerken wie Responsible AI, algorithmische Rechenschaftspflicht und dem EU-AI-Act positioniert das Papier die Ordonomik als designorientierte Ethik, die normative Ideale und institutionelle Ökonomie verbindet. Das Ergebnis ist ein Rahmenwerk zur Diagnose von Konflikten, zur Reflexion von Verantwortlichkeiten und zur Entwicklung kooperativer Lösungen, die Innovation mit sozialer Verantwortung in Einklang bringen.

Stichwörter: Ordonomik; soziale Dilemmata; Künstliche Intelligenz; Diagnose–Reflexion–Design; Moralparadox; institutionelles Design; KI-Governance

1. Introduction

The rapid development of artificial intelligence (AI) represents one of the most profound transformation processes of the twenty-first century. Machine learning, generative language models, and algorithmic decision systems not only alter production and service processes but also intervene deeply in education, labor markets, and democratic communication. These developments raise questions that extend far beyond technology: What happens to employment, to equal opportunity, and to the legitimacy of decisions when learning and reasoning are delegated to machines? How can societies ensure that innovation strengthens, rather than erodes, the social foundations on which it depends?

This paper starts from the assumption that AI is not primarily a technical phenomenon, but a problem of institutional order—a problem of how rules, incentives, and norms coordinate the relationship between individual benefit and collective well-being. Approaching AI from this institutional perspective allows us to go beyond moral appeals to responsible behavior and instead analyse the structural incentive constellations that shape action. Many current debates in AI ethics still address dilemmas as if they could be solved through awareness or virtue. Ordonomics offers a complementary approach: it understands moral conflicts as coordination failures that must be resolved through the redesign of social rules.

1.1 Research gap and motivation

Existing research on AI ethics and governance provides a rich set of principles and guidelines—for example, transparency, fairness, accountability, and human oversight (Floridi & Cowls, 2019; Jobin, Ienca & Vayena, 2019). Yet these frameworks often remain declarative rather than institutionalized. They articulate what ought to be valued, but not how competing incentives can be realigned to make responsible behavior rational for the actors involved. Ordonomics uniquely links moral reasoning to incentive design (Pies, 2017a; Pies, 2017b).

This lack of institutional-ethical analysis constitutes a gap in the current debate. The present paper addresses this gap by applying ordonomic reasoning—a framework within normative institutional economics that studies how moral norms and economic incentives can be integrated through rule design. Ordonomics thereby offers a grammar of argumen-

tation that explains social dilemmas as unintended consequences of rational action and seeks cooperative reforms rather than moral condemnation.

1.2 Conceptual orientation: ordonomics as applied framework

While ordonomics is traditionally formulated in terms of three levels of social coordination—

(1) the actor level (rule-following),
(2) the institutional order (rule-setting), and
(3) the market or discourse (rule-finding)—

This triad does not replace the classical scheme but serves as its applied heuristic translation:

Diagnosis corresponds to analysing incentive structures and identifying coordination failures.

Reflection assigns differentiated responsibilities across the actor, institutional, and market-discourse levels.

Design develops rule-based cooperation solutions that transform moral conflicts into win–win constellations through outbidding arguments (Pies, 2017a, Pies 2017b, Pies 2025).

Through this mapping, ordonomics becomes practically applicable for AI governance, offering both theoretical rigor and analytical clarity.

1.3 Positioning within AI ethics and governance

This paper situates ordonomics among established frameworks of Responsible AI, socio-technical systems theory, and algorithmic accountability.

Responsible AI emphasizes ethical principles and value alignment. Socio-technical approaches highlight the interplay of technical artefacts and social structures. Algorithmic accountability introduces procedural mechanisms like audits and documentation.

Ordonomics complements these approaches by focusing on the institutional preconditions under which moral principles become sustainable in practice. It treats fairness, transparency, and accountability not as behavioral exhortations but as rule properties that can be designed, enforced, and competitively rewarded.

1.4 Aim and contribution

The overall aim of this paper is to develop a systematic ordonomic analysis of artificial intelligence as a socio-economic dilemma. Specifically, the contribution is fourfold:

It clarifies the theoretical foundations of ordonomics and maps the classical three-level schema onto the applied triad Diagnosis–Reflection–Design. It provides a structured typology of AI-related social dilemmas—economic, epistemic, ethical, and educational—illustrating how individually rational strategies lead to collective inefficiencies. It elaborates differentiated responsibilities across actor, institutional, and market levels, addressing the moral paradox of modernity, in which individuals are morally overburdened and citizens politically underchallenged. It develops cooperative design strategies—regulatory, organizational, and educational—that align self-interest with the common good, illustrated through case studies on AI in employment, education, and governance.

1.5 Normative orientation and structure

The study is normative and design-oriented. It does not merely describe dilemmas but formulates institutional recommendations that can mitigate them. Accordingly, the paper is structured as follows: Section 2 explains the ordonomic framework and its conceptual mapping. Section 3 diagnoses and categorizes AI-related dilemmas. Section 4 applies the triad in depth, combining theoretical reasoning with practical illustration. Sections 5–7 develop case studies, discussion, and implications for governance, education, and policy.

Through this structure, the paper aims to demonstrate that ordonomics provides a coherent and operational method for linking moral reasoning with institutional design—an approach particularly needed in the governance of artificial intelligence.

2. Theoretical Framework: Ordonomics and the Moral Paradox of Modernity

Ordonomics, as developed primarily by Ingo Pies and colleagues since the 1990s, belongs to the broader tradition of normative institutional economics. It builds on Walter Eucken's classic dictum that "people must not be required to do what the economic order alone can achieve" (Eucken, 1952/1990, p. 368). In this sense, ordonomics conceives of morality not as an alternative to economics, but as a coordination resource that becomes productive through appropriately designed rules, procedures, and institutions (Homann & Pies, 2000). It thereby continues the ordoliberal tradition of German economic thought, combining ethical reasoning with regulatory policy (Ordnungspolitik).

2.1 The ordonomic logic of coordination

Analytically, ordonomics distinguishes three levels of social coordination:

(1) the actor level, where individuals and organizations pursue their interests under given rules;
(2) the institutional level, where rules and procedures are formulated and modified; and
(3) the market or discourse level, where ideas and competition generate incentives for further rule evolution.

This three-level schema provides a systematic architecture for analyzing moral and social conflicts. It reveals how individually rational strategies, when aggregated, can generate collectively inefficient outcomes—a pattern that ordonomics interprets as a social dilemma (Pies, 2000; Beckmann & Pies, 2016; Buchanan, 2000). In such cases, moral exhortations to individuals often fail, because the incentive structures that shape behavior remain unchanged. Instead, ordonomics calls for rule reforms that realign incentives so that self-interest and the common good coincide.

2.2 The moral paradox of modernity

This insight underlies what Pies (2022) describes as the moral paradox of modernity: modern societies tend simultaneously to morally overburden individuals and to politically underchallenge citizens. In other words, systemic coordination failures are often reinterpreted as moral deficits of individuals. For example, the public debate might blame users, consumers, or workers for outcomes that are structurally determined by flawed institu-

tional incentives. The result is a category error: problems of rule design are treated as problems of personal virtue.

Ordonomics resolves this paradox by analytically separating the level of rules (where solutions are designed) from the level of actions (where outcomes occur). This separation allows moral expectations to be rechanneled into institutional arrangements that reward cooperation instead of presupposing it. By doing so, ordonomics transforms moral appeals into rationally enforceable mechanisms of cooperation.

2.3 Social dilemmas as analytical core

Methodologically, ordonomics interprets many social problems as dilemmas of cooperation. In these situations, individually rational strategies generate collectively inferior outcomes—such as environmental degradation, social inequality, or algorithmic bias. The ordonomic response is to generate "outbidding arguments" (Pies, 2017a, Pies 2017b): arguments that make cooperative rule change not only morally desirable but privately attractive to the actors involved. This mechanism ensures that reforms are incentive-compatible rather than coercive or moralistic. The analytic procedure thus proceeds in three interrelated steps—Diagnosis, Reflection, and Design—which correspond directly to the classical three levels of coordination: Diagnosis clarifies the conflict structure and identifies perverse incentives (actor level). Reflection allocates differentiated responsibilities for reform across individual, institutional, and market-discourse levels. Design develops new rules, procedures, and incentives that transform destructive conflicts into mutually beneficial cooperation.

2.4 Ordonomics as normative institutional economics

In its philosophical foundations, ordonomics shares assumptions with the constitutional economics of Buchanan (2000) and the discourse ethics of Habermas (1981): both seek to reconcile rational self-interest with moral legitimacy through rule-based cooperation. Yet ordonomics distinguishes itself by treating moral norms explicitly as resources for coordination, not merely as constraints. Its focus lies on institutional learning processes—how societies improve their rules by identifying and correcting structural inefficiencies (Minnameier, 2016).

In this way, ordonomics transcends the traditional opposition between morality and economics. Rather than judging behavior from a moral standpoint, it examines how rules can make moral behavior rational. This shift from appeal to design marks ordonomics as a practical, action-guiding framework for ethics in modern, complex societies.

2.5 Application to artificial intelligence

Applied to artificial intelligence, the ordonomic perspective produces two key insights.

First, it highlights where the real coordination failures occur: in the rules, incentives, and information structures that govern data use, algorithmic decision-making, and accountability (Selbst et al., 2019; Raji et al., 2020). Second, it cautions against the moralization of AI discourse—for instance, when public debates focus on "ethical AI" as a matter of individual responsibility, while neglecting the institutional and regulatory dimensions.

Ordonomics provides a lens to differentiate responsibilities:

- What competences are required at the individual level?
- What rules and procedures must institutions establish?
- What market and discourse structures ensure that private interest aligns with the public good?

Through these guiding questions, the framework prevents the reduction of complex socio-technical systems to individual virtue ethics and instead treats AI as a governance problem. In this sense, ordonomics bridges normative theory, institutional design, and economic reasoning—making it particularly well suited to the ethical challenges of artificial intelligence.

Ordonomic level	Analytical focus	Guiding questions (examples)	Applied Triad
Actor level	Individual and organizational behaviour	What actions are rational under existing rules?	Diagnosis
Institutional level	Rules, procedures, governance structures	How are incentives shaped and responsibilities allocated?	Reflection
Market / discourse level	Public debate, norm evolution	How can rules be reformed through cooperation and new technology?	Design

Table 1. Mapping the classical ordonomic levels of coordination onto the applied heuristic of Diagnosis–Reflection–Design.

The schematic in this table visualizes how the classical ordonomic levels of coordination (actor, institutional, and market/discourse) correspond to the applied triad of Diagnosis–Reflection–Design. It illustrates the process logic by which AI-related social dilemmas—economic, epistemic, ethical, and educational—are diagnosed, reflected, and institutionally redesigned toward cooperative outcomes.

A conceptual framework mapping the classical ordonomic three-level schema (actor, institutional, and market–discourse levels) onto the applied heuristic of Diagnosis–Reflection–Design in the analysis of AI-related social dilemmas.

3. Artificial Intelligence as a Transformation Phenomenon and Source of Socio-Economic Dilemmas

Artificial intelligence represents not only a technological innovation but a profound transformation of social coordination. As with earlier general-purpose technologies such as electricity or the Internet, AI reconfigures the way information, labor, and responsibility are distributed within society. Yet unlike previous industrial revolutions, AI operates at a cognitive level: it automates not only physical tasks but also judgment, evaluation, and decision-making. This dual nature—technological and epistemic—makes AI a paradigmatic case for ordonomic analysis. It does not create entirely new moral questions but intensifies existing social dilemmas by scaling them through automation, data, and network effects.

3.1 Defining the transformation phenomenon

AI can be described as a transformation phenomenon in two interconnected senses.

First, it transforms productive structures by automating and optimizing tasks that were previously human. Second, it transforms institutional structures, altering how knowledge, authority, and legitimacy are distributed. The technology therefore acts simultaneously on the economic and the normative infrastructures of society. This dual impact leads to a constellation of social dilemmas—conflicts between rational individual optimization and collective welfare—that require not moral exhortation but institutional realignment.

Ordonomics conceptualizes such transformation processes as rule dynamics: new technologies challenge the adequacy of existing rules, which then triggers discursive and institutional adaptation. The challenge for AI governance is to ensure that this adaptation process does not lag behind technological progress but evolves in parallel, maintaining the compatibility between innovation and cooperation.

3.2 Typology of AI-related social dilemmas

To analyze these coordination challenges, AI-related dilemmas can be grouped into four interrelated categories: economic, epistemic, ethical, and educational. Each category captures a distinctive misalignment of incentives that arises when private rationality diverges from collective welfare.

(a) Economic dilemmas: efficiency and employment

At the firm level, AI promises efficiency gains, cost reduction, and competitive advantage. Rational companies therefore invest in automation, data analytics, and algorithmic management. Yet when aggregated across the economy, these micro-level decisions may generate macro-level side effects: the polarization of skill profiles, displacement of middle-income jobs, and rising inequality (Acemoglu & Restrepo, 2020). This tension between individual competitiveness and collective stability constitutes a classical social dilemma.

From an ordonomic viewpoint, the question is not whether automation is good or bad, but under which institutional conditions its benefits can be distributed without eroding social cohesion. Rules for retraining, social insurance, and innovation incentives determine whether the dilemma remains functional (promoting progress) or becomes dysfunctional (undermining solidarity). Written in other terms:

- Actors: Firms competing in product and labor markets invest in AI-driven automation to improve productivity and reduce costs.
- Incentives: Competitive pressure rewards early adoption, efficiency gains, and short-term cost reduction, making automation a rational firm-level strategy.
- Collective outcome: When aggregated, these individually rational decisions contribute to job polarization, the displacement of middle-income work, and rising inequality.
- Ordonomic insight: The resulting tension between firm-level efficiency and macro-level social stability constitutes a social dilemma that can only be addressed through institutional rules for reskilling, social insurance, and labor market adaptation.

(b) Epistemic dilemmas: performance and transparency

AI systems often trade accuracy for explainability. Proprietary algorithms, trained on vast datasets, outperform humans in prediction but resist scrutiny. The resulting opacity generates what Burrell (2016) calls the black-box problem. Companies have an incentive to protect intellectual property; regulators and citizens demand transparency and accountability. The conflict lies between the private incentive to conceal and the public need to understand.

Ordonomically, this dilemma is not solved by appealing to corporate ethics but by designing institutions—such as audit rights, model documentation (Gebru et al., 2018), and disclosure standards—that make transparency compatible with competitive interest.

- Actors: AI developers and deploying organizations on the one hand, regulators and affected publics on the other.
- Incentives: Firms seek to protect proprietary models and competitive advantage, while regulators and citizens require transparency to ensure accountability.
- Collective outcome: The resulting opacity undermines trust and democratic oversight, despite high system performance.
- Ordonomic insight: Transparency must be institutionalized through audit rights and documentation standards that make accountability incentive-compatible.

(c) Ethical dilemmas: optimization and fairness

Algorithmic optimization frequently reproduces structural bias present in data (Binns, 2018; Raji et al., 2020). Employers, insurers, or credit institutions pursue efficiency and risk minimization; in aggregate, such optimization can undermine fairness and social legitimacy. The dilemma arises when private utility maximization erodes the collective basis of trust.

The ordonomic lens interprets fairness as a property of rules, not of isolated actions. The ethical challenge is therefore to reform the rule systems that govern data collection, model training, and evaluation—turning fairness from a moral aspiration into an institutional standard.

- Actors: Organizations deploying AI systems for hiring, credit scoring, insurance, or public administration, as well as individuals affected by algorithmic decisions.
- Incentives: Organizations are incentivized to optimize decision accuracy, efficiency, and risk minimization, often relying on historical data and automated optimization criteria.
- Collective outcome: When scaled across systems and institutions, such optimization can reproduce or amplify existing social biases, undermining fairness, equal opportunity, and the legitimacy of algorithmic decision-making.
- Ordonomic insight: Fairness deficits are not primarily the result of unethical individual behavior but of rule systems that reward efficiency without internalizing distributive effects; addressing this dilemma therefore requires institutional standards for data governance, bias auditing, and accountability that make fairness incentive-compatible.

(d) Educational dilemmas: personalization and autonomy

In education, AI enables personalized learning and efficient assessment. Adaptive systems can tailor tasks, feedback, and pacing to each student (Seufert & Meier, 2023). Yet such

personalization risks eroding students' autonomy if it replaces self-regulation with algorithmic steering. The dilemma here is between short-term learning gains and long-term independence.

When teachers rely excessively on AI feedback or grading, the epistemic authority shifts from human educators to technical systems. The ordonomic approach suggests designing rules that preserve the normative primacy of human judgment, for instance through transparency obligations, mixed-assessment formats, and digital literacy education.

- Actors: Students, teachers, educational institutions, and providers of AI-based learning and assessment systems.
- Incentives: AI systems offer incentives for efficiency, personalization, and performance optimization by adapting content, feedback, and assessment to individual learners.
- Collective outcome: When reliance on algorithmic guidance becomes pervasive, educational practices risk prioritizing measurable performance over critical reflection, thereby weakening learners' autonomy, epistemic agency, and responsibility for their own learning processes.
- Ordonomic insight: Preserving educational autonomy requires rules that embed AI as a supportive instrument rather than a substitutive authority; institutional designs such as transparency requirements, mixed assessment formats, and AI literacy education can realign efficiency gains with the public good of independent judgment.

3.3 Cross-sectional synthesis

These four domains reveal recurring structural patterns. Across them, AI generates coordination failures by altering information asymmetries, incentive structures, and moral expectations simultaneously. Actors optimize locally within their constraints, while collectively undermining the institutional trust that sustains cooperation.

From an ordonomic perspective, these tensions can be summarized as three generic conflicts:

- Information asymmetry – AI systems concentrate information and decision power asymmetrically between producers and users.
- Incentive misalignment – rules reward short-term efficiency rather than long-term stability or fairness.
- Moral overburdening – responsibility is displaced onto individuals who lack the power to alter the underlying incentives.

The consequence is a proliferation of what Pies (2017) calls undesirable dilemmas—those that trap actors in collectively inferior equilibria. Recognizing these structures is the first step toward diagnostic clarity. The next analytical move is reflection: to assign differentiated responsibilities across actors, institutions, and the market-discourse system, preparing the ground for cooperative redesign.

3.4 Interim conclusion

AI thus functions as a mirror for the social order. It exposes latent contradictions between efficiency, justice, and legitimacy that modern societies must address not by moralizing technology but by improving their rules of coordination. The ordonomic approach provides a systematic method to analyse these contradictions: it treats AI neither as an

autonomous moral agent nor as a neutral tool, but as an amplifier of human and institutional incentives. The following section applies this ordonomic triad—Diagnosis, Reflection, and Design—to develop concrete pathways for institutional adaptation in the domains of work, education, and governance.

4. Applying the Ordonomic Triad (Diagnosis–Reflection–Design)

Building on the theoretical foundations and the typology of dilemmas, this section applies the ordonomic triad to the domain of artificial intelligence. The triad serves as a heuristic that connects empirical observation with institutional design. It enables a movement from diagnosis—the clarification of conflict structures—to reflection—the assignment of differentiated responsibilities—and finally to design, the development of cooperative solutions that transform dysfunctional equilibria into socially beneficial ones.

The application demonstrates that AI governance, when viewed through an ordonomic lens, is not primarily about restricting innovation, but about shaping the rules that align innovation with collective welfare.

4.1 Diagnosis: AI and employment as an illustrative dilemma

The employment dilemma illustrates the ordonomic logic of social conflicts in a paradigmatic way. On the actor level, companies rationally pursue automation and algorithmic management to increase productivity, flexibility, and competitiveness. The resulting technological substitution effects—especially in repetitive and information-processing tasks—improve firm efficiency and shareholder returns. However, at the systemic level, the same optimization process can produce side effects such as job polarization, regional inequality, and declining middle-class stability (Acemoglu & Restrepo, 2020).

This tension between micro-rationality and macro-irrationality constitutes a classic social dilemma. Each firm acts rationally under competition, but collectively these actions may erode the very demand and trust conditions that sustain markets. In the short term, automation seems welfare-enhancing; in the long term, it risks social fragmentation. Ordonomically, this reveals a coordination failure: rules at the institutional level insufficiently internalize the social costs of technological displacement.

Whether this dilemma is desirable or undesirable depends on the institutional framework. Desirable dilemmas drive progress by rewarding innovation while maintaining adaptive institutions. Undesirable dilemmas, by contrast, persist when institutions fail to redistribute the gains or provide adjustment mechanisms. The ordonomic diagnosis, therefore, distinguishes between functional competition and dysfunctional erosion, identifying the rule conditions that determine the balance.

To exemplify this, consider the differing trajectories of AI diffusion across economies. Countries with strong vocational education systems, portable social insurance, and proactive retraining policies—such as Denmark or the Netherlands—are better equipped to absorb technological shocks without amplifying inequality. In contrast, where institutions rely primarily on labor market deregulation and limited social buffering, the same innovation dynamic produces exclusion rather than empowerment. The dilemma thus manifests not in the technology itself, but in the rules that govern adaptation.

4.2 Reflection: Assigning responsibilities across levels of coordination

Following the ordonomic logic, effective resolution of social dilemmas requires a clear differentiation of responsibilities across actors, institutions, and the market-discourse system.

At the actor level, firms and individuals are expected to act within given rules, but not to unilaterally solve systemic conflicts. The expectation that corporations alone should "act ethically" disregards the competitive pressure that constrains them. However, actors can contribute by integrating AI ethics principles into internal governance, promoting transparency in algorithmic systems, and supporting retraining initiatives for affected employees. These actions mitigate, but do not eliminate, the structural causes of the dilemma.

At the institutional level, governments, regulators, and intermediary organizations must establish the formal and informal rules that correct incentive asymmetries. This includes labor market institutions that facilitate reskilling and job mobility, fiscal systems that support social innovation, and regulatory frameworks that ensure accountability for AI deployment in employment contexts. Educational institutions have a parallel role: integrating AI literacy into curricula and ensuring that learning systems preserve critical reasoning rather than promote dependency.

At the market and discourse level, the competitive and communicative arenas in which norms evolve, public debate plays a constitutive role. The recent open letters on AI governance—whether calling for a moratorium on large-scale experiments (Future of Life Institute, 2023) or for a pause in overregulation (AI Champions, 2025)—illustrate how the discourse level functions as a meta-arena of rule-finding. In this space, societies negotiate the legitimate boundaries between innovation and precaution. Discursive coordination, however, requires procedural rationality: actors must learn to "outbid" purely moralistic arguments by demonstrating that cooperation is instrumentally superior to unilateral action.

This reflective differentiation avoids two extremes: it neither moralizes market behavior nor reduces ethics to compliance. Instead, it conceptualizes social responsibility as a distributed task, embedded in complementary roles that together enable cooperative rule change. The moral paradox of modernity is thus countered: individuals are no longer overburdened, and citizens are not underchallenged, because each level is normatively engaged within its functional competence.

4.3 Design: From moral appeal to rule-based cooperation

The final step of the triad, Design, focuses on developing institutions and mechanisms that realign incentives. The ordonomic ambition is not to impose external constraints on rational actors but to make cooperation rational by transforming the rule environment. In the AI context, this entails the creation of regulatory, organizational, and educational designs that embed normative goals within competitive structures.

Regulatory design addresses the systemic level of rule-setting. The emerging EU Artificial Intelligence Act exemplifies a move toward risk-based governance, where AI systems are classified according to their societal impact. Requirements for transparency, auditability, and human oversight translate moral concerns into operational criteria. Rather than restricting innovation, such regulation creates a predictable playing field in which ethical performance becomes a competitive advantage. By coupling compliance incentives with public accountability, these rules internalize externalities without centralizing control.

Organizational design operates at the meso-level of companies and institutions. Mechanisms such as algorithmic impact assessments, model cards (Mitchell et al., 2019), and data documentation protocols (Gebru et al., 2018) institutionalize accountability. They make the ethical quality of AI systems observable and auditable, turning trust into a measurable asset. Firms that adopt such measures not only reduce reputational risk but also signal reliability to clients and regulators. Over time, these practices generate a market for responsibility, where transparency and fairness are rewarded.

Educational design complements these mechanisms by addressing the formation of competences and moral agency. Integrating AI literacy into education helps future workers, managers, and citizens understand the logics and limits of algorithmic systems. Rules for AI use in schools and universities should protect intellectual autonomy, for example by combining automated feedback with oral defense and reflective documentation. In doing so, education becomes a microcosm of ordonomic reasoning: it aligns individual learning with collective epistemic responsibility.

Across these design levels, the principle of outbidding arguments serves as the connective logic. Actors are motivated to cooperate when the institutional framework allows them to achieve more together than apart. Cooperation becomes rationally superior to defection because well-designed rules transform moral desirability into private advantage. In this sense, ordonomic design is not an abstract ideal but a strategy of institutional evolution—a process of iterative rule improvement through public reasoning and competitive experimentation.

4.4 Synthesis: Toward a cooperative order of AI governance

Applying the ordonomic triad to AI governance reveals that the key to reconciling innovation and responsibility lies in rule-based cooperation. The dilemmas of employment, transparency, and autonomy cannot be solved by technical optimization alone. They require institutional arrangements that transform the structure of incentives and the distribution of knowledge.

The ordonomic contribution is to offer a methodological grammar for such institutional design. Diagnosis identifies the coordination failures; reflection clarifies the normative division of labor; and design formulates concrete mechanisms that realign private and public interests. When applied iteratively, this process enables what Homann (2002) described as the "ethical learning of systems"—a collective capacity to adapt rules in light of new technological realities.

In the domain of artificial intelligence, this means that responsibility is not exhausted in compliance checklists or moral declarations. It resides in the continuous improvement of the institutional order itself—the dynamic adjustment of rules that make innovation sustainable. Through this perspective, ordonomics reframes AI ethics as a project of institutional creativity, in which societies learn to design the conditions under which human and artificial intelligence can coexist productively.

5. Case Studies: Education, Human Resources, and Governance

The ordonomic framework becomes tangible when applied to concrete domains in which artificial intelligence reshapes established coordination patterns. Each domain presents specific constellations of actors, incentives, and rule systems, yet they all reflect the same

underlying logic: individually rational optimization that unintentionally produces collective dysfunction. Analysing these cases through the triad Diagnosis–Reflection–Design demonstrates how cooperation can be reconstructed by institutional means.

5.1 Education: Learning under algorithmic conditions

Diagnosis: In education, AI tools such as adaptive tutoring systems, plagiarism detectors, and generative language models have transformed both teaching and assessment. Students use AI to obtain feedback and accelerate learning, while teachers experiment with automation to manage workloads and personalize instruction. On the actor level, these behaviors appear rational: they promise efficiency, accessibility, and inclusion. Yet at the systemic level, they risk undermining the cultivation of critical thinking, authorship, and intellectual autonomy.

Reflection: The resulting dilemma lies in the tension between personalization and autonomy. Algorithmic systems optimize for engagement and performance metrics; human learning, however, depends on friction, error, and reflection. When algorithms overfit education to predicted outcomes, they silently redefine what counts as competence. This shift threatens the public good of education as a space for independent reasoning (Williamson & Piattoeva, 2022). From an ordonomic standpoint, the task is to design rules that preserve the normative primacy of human judgment within digitally mediated learning environments. Reflection therefore differentiates responsibilities: teachers remain accountable for assessment standards; institutions set transparent usage policies for AI tools; students are obliged to disclose AI assistance; and ed-tech providers ensure traceability and explainability of their systems.

Design: One can translate these normative expectations into practice. Universities may require written process documentation or oral defense to accompany AI-supported assignments; curricula may include modules on digital epistemology and algorithmic bias; accreditation agencies may demand audit trails for automated grading. Through such institutional embedding, AI becomes not an agent of substitution but an instrument of cognitive cooperation—enhancing learning while maintaining the rule of reflective autonomy (Minnameier, 2025).

5.2 Human resources: Algorithmic management and the search for fairness

Diagnosis: The domain of human resource management illustrates another structural dilemma. Organizations increasingly rely on AI-based tools for recruitment, screening, and performance evaluation. These systems promise objectivity, speed, and cost efficiency, allowing firms to handle thousands of applications or to monitor productivity in real time. For each company, adopting such technologies appears strategically rational; no single actor can afford to ignore efficiency pressures. However, the aggregate effect of widespread automation may be the erosion of fairness and trust in labor markets (Raghavan et al., 2020).

The incentive asymmetry is evident because vendors profit from proprietary algorithms, employers from efficiency, while job seekers bear the opacity and potential bias of automated judgments. The collective outcome is a deficit of legitimacy—an institutional externality rather than a moral lapse.

Reflection: The ordonomic analysis reframes this as a rule-design problem, in other words: how can fairness become an economically rational attribute of AI systems? The answer lies in making fairness auditable and commercially valuable. Certification schemes, bias-testing requirements, and transparency clauses in procurement contracts turn moral expectations into enforceable standards. Reflection again distributes tasks means developers document models and datasets; employers provide audit access and impact assessments; regulators define thresholds and procedures for algorithmic accountability.

Design: One could connect these layers through shared metrics—algorithmic impact assessments, model cards, and datasheets for datasets—that create comparability across systems (Gebru et al., 2021; Mitchell et al., 2019). When such reporting becomes a market norm, reputational and legal incentives converge, producing what might be called a market for responsibility. In this order, ethical quality is no longer an externality but a competitive resource.

5.3 Governance: Balancing innovation and precaution

Beyond specific sectors, AI challenges the foundations of democratic governance itself.

Diagnosis: Public authorities experiment with predictive policing, welfare automation, and administrative decision systems. These applications raise legitimacy questions that cannot be answered solely by technical accuracy. The governance dilemma concerns the balance between innovation and precaution. Excessive regulation risks stifling beneficial innovation; too little regulation exposes citizens to opaque and potentially discriminatory systems. Diagnosing this dilemma reveals an institutional misalignment: the pace of technological change outstrips the adaptability of legal frameworks. Governments, motivated by efficiency and cost reduction, adopt algorithmic tools faster than oversight mechanisms evolve. Citizens, meanwhile, lack the information and rights necessary to contest automated decisions. The result is a gap between de facto technological power and de jure democratic control.

Reflection: It requires clarifying who bears responsibility for maintaining this balance. Legislators set general principles; agencies implement and monitor; civil society and academia supply critical feedback. Discourse at the public level functions as a corrective meta-game in which legitimacy is continually renegotiated. The rule of law thus depends on procedural transparency and access to contestation.

Design: One could translate these insights into institutional architecture. Risk-based regulation, such as the EU Artificial Intelligence Act, classifies AI systems by societal impact and imposes corresponding duties of transparency, data governance, and human oversight. Complementary instruments—impact assessments, mandatory documentation, and independent audit bodies—transform moral demands for accountability into enforceable rights. Through these mechanisms, governance evolves from reactive control to anticipatory coordination: a system capable of learning from its own dilemmas.

5.4 Comparative synthesis

Across the domains of education, human resources, and governance, the same structural pattern reappears. AI systems intensify coordination failures by shifting information asymmetries and by multiplying the speed of interaction, while existing institutions lag behind. Yet these failures are not inevitable. When rules are designed to realign incentives—so that

transparency, fairness, and learning become rationally rewarded—the dilemmas turn from destructive to productive.

Ordonomically, this transformation can be understood as a process of institutional learning. Each domain develops specific mechanisms—usage policies in education, audit standards in HR, risk classification in governance—but all follow the same meta-logic: making cooperation more rewarding than unilateral optimization. In this sense, AI becomes a test case for modern societies' capacity to update their moral infrastructures. Rather than appealing to ethical heroism, ordonomics invites the redesign of the game itself, ensuring that innovation and integrity reinforce one another.

6. Discussion: Integrating Ordonomics into the AI Ethics and Governance Discourse

The preceding analysis demonstrates that artificial intelligence confronts societies with classical coordination problems in a new technological guise. The ordonomic framework offers a language and logic for understanding these problems not as isolated ethical controversies but as manifestations of incentive misalignment. This section situates ordonomics within the broader landscape of AI ethics and governance and discusses its theoretical, methodological, and normative implications.

6.1 Bridging AI ethics and institutional economics

Much of the current AI ethics discourse is principle-driven. Frameworks such as the OECD AI Principles (2019) or Floridi and Cowls' (2019) five-principle model—beneficence, non-maleficence, autonomy, justice, and explicability—formulate shared values but often remain detached from the mechanisms that could realize them. Ordonomics contributes an institutional bridge by asking under which rule conditions these values can become self-enforcing.

Whereas traditional ethics seeks moral compliance by individuals, ordonomics focuses on the game architecture that structures behavior. It conceptualizes cooperation as a product of rule-design rather than virtue. This perspective resonates with the notion of responsible innovation (Stilgoe et al., 2013) and value-sensitive design (Friedman & Hendry, 2019) but adds an explicit economic dimension: actors must have incentives to internalize ethical expectations. By converting moral aims into institutional payoffs—through regulation, market signaling, and discourse—ordonomics operationalizes what otherwise remains aspirational.

6.2 Complementarity with socio-technical and accountability approaches

Socio-technical systems theory interprets technology as embedded in human, organizational, and cultural contexts (Bijker, 1997; Suchman, 2007). Algorithmic accountability research similarly seeks procedural safeguards such as audits, documentation, and contestability (Raji et al., 2020; Selbst et al., 2019). Ordonomics complements these approaches by providing a normative-economic grammar for why such procedures matter.

Audits and transparency reports, for instance, can be read as institutionalized outbidding arguments: they make trust commercially valuable. Documentation and contestability mechanisms serve to re-balance asymmetric information. From an ordonomic viewpoint, these measures are not external ethical add-ons but rule-based corrections that restore cooperation in competitive settings. The approach thus links descriptive analyses of socio-

technical complexity with a theory of institutional learning—how societies adjust rules to turn dilemmas into opportunities for collective gain.

Recent work on implicit decision-making and governance further supports this perspective. Hedfeld (2025) shows how implicit voting mechanisms and language models can be conceptualized as normatively legitimate and institutionally implementable rules, thereby illustrating how discursive coordination can be translated into incentive-compatible governance structures in complex socio-technical systems (Hedfeld, 2025a; Hedfeld 2025b).

6.3 The ordonomic contribution to governance theory

In governance theory, ordonomics clarifies how discursive and regulatory arenas interact. The discourse level—public debate, scientific reflection, and stakeholder dialogue—acts as a meta-arena for rule-finding. Here, actors test and refine competing arguments, selecting those that enable broader cooperation. The market and institutional levels then implement these insights through laws, standards, and organizational routines.

This iterative feedback between discourse and order constitutes a dynamic model of ethical governance. It aligns with reflexive governance theories (Voß et al., 2006) but grounds them in incentive logic. Rather than ideal deliberation, ordonomics assumes bounded rationality and competitive pluralism. Ethical progress occurs when rule systems evolve to make cooperative solutions Pareto-superior to conflictual ones. In AI governance, this principle underlies risk-based regulation, co-regulation, and standardization efforts that allow innovation to continue under clear accountability constraints.

6.4 Relating ordonomics to current policy frameworks

The emerging EU Artificial Intelligence Act operationalizes many ordonomic ideas in practice. By linking risk categories to mandatory obligations—such as transparency, human oversight, and post-market monitoring—it creates a graduated incentive structure that rewards responsible innovation. Similar logics guide the OECD AI Principles and the UNESCO Recommendation on the Ethics of AI. These instruments function as meta-rules: they do not prescribe outcomes but specify procedures for ethical alignment.

Ordonomics interprets these developments as part of a broader rule evolution: societies learn to transform moral expectations into enforceable coordination mechanisms. Regulation thus becomes an instrument of ethical learning rather than a barrier to progress. The ultimate objective is not to moralize technology but to embed its operation in a framework of reciprocal advantages—a cooperative order in which compliance with ethical standards coincides with strategic rationality.

6.5 Theoretical synthesis and implications

Integrating ordonomics into the AI governance discourse yields three conceptual insights:

Reconceptualizing responsibility. Responsibility shifts from individual virtue to the design of institutions that make virtuous action rational. This redefinition dissolves the moral paradox of modernity: actors are no longer expected to sacrifice their interests but to pursue them within fair rules.

Ethics as institutional learning. Ethical norms evolve through feedback between action, order, and discourse. AI governance exemplifies this evolutionary process: public controversies trigger regulatory adaptation, which in turn reshapes incentives and behavior.

Cooperation as a design problem. The central question of AI governance is not what values to hold, but how to design rules so that adherence to those values becomes payoff-consistent. Ordonomics provides a methodological grammar for this transformation, translating moral reasoning into institutional architecture.

Through these implications, ordonomics positions itself as both a complement and a corrective to mainstream AI ethics. It does not replace value frameworks but grounds them economically, ensuring that normative aspirations survive contact with strategic reality. In this sense, ordonomics represents a form of design-oriented ethics: an ethics of systems, not of saints.

7. Conclusion and Outlook

Artificial intelligence is both a catalyst and a mirror of contemporary social order. Its diffusion amplifies long-standing coordination problems between efficiency, fairness, and legitimacy. By applying ordonomic reasoning, this paper has shown that these tensions cannot be resolved through moral appeals alone but require institutional learning.

Ordonomics, understood as normative institutional economics, interprets social dilemmas as structural misalignments between individual incentives and collective outcomes. Through the triad Diagnosis–Reflection–Design, the approach provides a systematic grammar for transforming moral conflicts into cooperation problems that can be solved by redesigning rules.

Diagnostically, AI reveals where incentive structures generate collectively inferior outcomes: in the displacement of labor, the opacity of algorithmic decision-making, or the erosion of educational autonomy. Reflection then assigns differentiated responsibilities across actors, institutions, and the market-discourse system, preventing both moral overburdening and political passivity. Finally, design translates normative expectations into incentive-compatible rules—through regulation, organizational standards, and education—so that responsibility becomes a structural feature of the system itself.

The broader implication is that AI governance represents an ongoing experiment in institutionalized ethics. Risk-based regulation, algorithmic audits, and transparency requirements exemplify how societies transform values into enforceable coordination mechanisms. Rather than opposing innovation and morality, ordonomics invites their integration: it demonstrates how cooperative rule design can make responsibility a competitive advantage. This article itself can be read as a discursive contribution at the rule-finding (meta-meta-game) level as theory-driven conceptual paper.

Future research should empirically investigate how such ordonomic mechanisms operate in practice—how firms, regulators, and educational institutions internalize ethical expectations through incentive structures. Comparative studies across sectors and jurisdictions could reveal patterns of institutional learning that help societies anticipate, rather than merely react to, technological change.

Ultimately, the ordonomic approach reframes AI ethics as a project of collective rule intelligence. It asks not only how machines can become intelligent, but how humans can design the social rules that ensure intelligence—human or artificial—serves the cooperative advancement of society.

References

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of political economy*, *128*(6), 2188-2244.

AI Champions. (2025). Stop the Clock: Open Letter Calling for an EU AI Act Pause. Available online at: https://aichampions.eu (Call 16.01.2026)

Beckmann, M., & Pies, I. (2016). The constitution of responsibility: Toward an ordonomic framework for interpreting (corporate social) responsibility in different social settings. In *Order ethics: An ethical framework for the social market economy* (pp. 221-250). Cham: Springer International Publishing.

Bijker, W. E. (1997). *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*. MIT press.

Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (FAT), 149–159.

Buchanan, J. M. (2000). *Reason of Rules—Constitutional Political Economy*. Liberty Fund Incorporated, us.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society*, *3*(1), 2053951715622512.

Eucken, W. (1952/1990). Grundsätze der Wirtschaftspolitik. Tübingen: J.C.B. Mohr (Paul Siebeck).

Floridi, L. & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Philosophy & Technology, 32(4), 685–703.

Friedman, B. & Hendry, D. (2019). Value Sensitive Design: Shaping Technology with Moral Imagination. MIT Press.

Future of Life Institute. (2023). Pause Giant AI Experiments: An Open Letter. Available online at: (https://futureoflife.org/open-letter/pause-giant-ai-experiments/) (Call 16.01.2026)

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86-92.

Hedfeld, P. (2025a). Implicit decision voting made by humans as normative and implementable rules with the help of language models. In R. Buchkremer, O. Koch & A. Lischka (Hrsg.), ifid Schriftenreihe: Beiträge zu IT-Management & Digitalisierung (Bd. 3). FOM-Hochschule für Oekonomie & Management. ISBN 978-3-89275-395-7.

Hedfeld, P. (2025b). Essay: Mit der Langfristigkeit im Herzen–Nachhaltigkeit und Generationengerechtigkeit, eine interdisziplinäre Perspektive zwischen Sozialpädagogik und Wirtschaftsethik. *Zeitschrift für Sozialpädagogik*, (1).

Homann, K. (2002). *Vorteile und Anreize: Zur Grundlegung einer Ethik der Zukunft*. Mohr Siebeck.

Homann, K., & Pies, I. (2000). Wirtschaftsethik und Ordnungspolitik–Die Rolle wissenschaftlicher Aufklärung. *Ordnungstheorie und Ordnungspolitik–Konzeptionen und Entwicklungsperspektiven, Stuttgart*, 329-346.

Jobin, A., Ienca, M. & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence, 1(9), 389–399.

Minnameier, G. (2016). Rationalität und Moralität: Zum systematischen Ort der Moral im Kontext von Präferenzen und Restriktionen. *Zeitschrift für Wirtschafts-und Unternehmensethik*, *17*(2), 259.

Minnameier, G. (2025). Ordonomik und Bildung: Verantwortung für die moderne Gesellschaft (p. 372). wbv Media.

Mitchell, M. (2019). Artificial Intelligence: A Guide for Thinking Humans. Farrar, Straus and Giroux.

Mitchell, M., Wu, S., Zaldivar, A., et al. (2019). Model Cards for Model Reporting. Proceedings of FAT 2019, 220–229.

OECD. (2019). OECD Principles on Artificial Intelligence. Paris: OECD Publishing. https://www.oecd.org/en/topics/ai-principles.html (Call 16.01.2026)

Pies, I. (2000). Ordnungspolitik in der Demokratie: Ein ökonomischer Ansatz diskursiver Politikberatung. Tübingen: Mohr Siebeck. ISBN 3-16-147507-0.

Pies, I. (2017a). Ordonomik als Methode zur Generierung von Überbietungsargumenten: Eine Illustration anhand der Flüchtlings (politik) debatte (No. 2017-03). Diskussionspapier. https://doi.org/10.5771/1439-880X-2017-2-171

Pies, I. (2017b). The ordonomic approach to business ethics. *Available at SSRN 2973614*.

Pies, I. (2022). Kapitalismus und das Moralparadoxon der Moderne. Berlin: wvb Wissenschaftlicher Verlag Berlin. ISBN 978-3-96138-310-8

Pies, I. (2025). The interplay of incentives and ideas: An intellectual journey from order economics through order ethics to ordonomics (No. 2025-08). Diskussionspapier. https://www.econstor.eu/bitstream/10419/325828/1/1936155664.pdf (Call 16.01.2025)

Raghavan, M., Barocas, S., Kleinberg, J. & Levy, K. (2020). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAccT), 469–481.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020, January). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 33-44).

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 59-68).

Seufert, S., & Meier, C. (2023). Hybrid Intelligence: Collaboration with AI Systems for Knowledge Work. *HMD Praxis der Wirtschaftsinformatik*, *60*(6), 1194-1209.

Stilgoe, J., Owen, R. & Macnaghten, P. (2013). Developing a Framework for Responsible Innovation. Research Policy, 42(9), 1568–1580.

Suchman, L. (2007). Human-Machine Reconfigurations: Plans and Situated Actions. Cambridge University Press.

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence (Call 16.01.2026)

Voss, J. P., Bauknecht, D., & Kemp, R. (Eds.). (2006). *Reflexive governance for sustainable development*. Edward Elgar Publishing.

Williamson, B., & Piattoeva, N. (2022). Education governance and datafication. *Education and Information Technologies*, *27*, 3515-3531.

Patrick Hedfeld, Dr., works for the Center of Business Ethics at the Goethe University and as lecturer for the FOM University of Applied Sciences in Frankfurt am Main.

Address: Johann Wolfgang Goethe-Universität Frankfurt Theodor-W.-Adorno-Platz 1 60323 Frankfurt am Main and FOM University of Applied Sciences Franklinstraße 52, 60486 Frankfurt am Main, Germany, E-Mail: hedfeld@econ.uni-frankfurt.dehttps://orcid.org/0000-0002-0385-2829