

Article

Cultivating Dignity in Intelligent Systems

Adeniyi Fasoro 

Department of Philosophy, Technische Universität Berlin, 10587 Berlin, Germany; fasoro.adeniyi@gmail.com

Abstract: As artificial intelligence (AI) integrates across social domains, prevailing technical paradigms often overlook human relational needs vital for cooperative resilience. Alternative pathways consciously supporting dignity and wisdom warrant consideration. Integrating seminal insights from virtue and care ethics, this article delineates the following four cardinal design principles prioritizing communal health: (1) affirming the sanctity of life; (2) nurturing healthy attachment; (3) facilitating communal wholeness; and (4) safeguarding societal resilience. Grounding my analysis in the rich traditions of moral philosophy, I argue that these principles scaffold sustainable innovation trajectories that consciously center shared welfare advancement over detached technical capabilities or efficiency benchmarks alone. Elucidating connections with pioneering initiatives demonstrates fragments of this vision taking embryonic shape, yet pervasive adoption remains largely aspirational to date. Fulfilling dignity-based artificial intelligence demands ongoing collective commitment beyond firms' profit motives or governance proceduralism. My conclusions urge technology policies and priorities directed toward empowering the vulnerability of people rather than controlling the optimization of systems.

Keywords: virtue ethics; care ethics; human dignity; artificial intelligence; responsible innovation; sustainable technology

1. Introduction

As artificial intelligence (AI) rapidly advances and integrates within shared social environments, issues of security and control appear secondary to deeper questions of mutual flourishing. What foundational communal values and designer virtues might orient coming waves of intelligence to consciously support human dignity rather than inadvertently erode social foundations through naive optimization functions? This article develops an alternative paradigm integrating virtue and care ethics toward sustainable AI progress explicitly centered on human welfare through grounding in moral relationships.

Many scholars warn that technology's fixation on autonomy, efficiency, and control breeds fragility by isolating people from interdependence, empathy, and wisdom cultivated through communal participation [1–3]. Prevailing computational systems and machine learning models emphasize only technical reliability, productivity gains, or risk mitigation while overlooking the moral relationships necessary for psychologically healthy existence [4,5]. Their narrow aims displace richer ethical reasoning, community building, and long-term resilience to dynamic uncertainty.

In response, this article bridges timeless conceptual foundations from seminal philosophers into four cardinal design principles. It offers guidance for the responsible development of artificial intelligence as a compassionate ally consciously constructed to enrich collective wellbeing across families, organizations, and societies. The principles proposed shine light on priorities beyond functionality that stand vital for meaningful integration of intelligent algorithms within communal life over generational timescales.

Fulfilling this paradigm vision constitutes unfinished work-in-progress. Beyond academic debate or isolated best practices, it ultimately demands a cultural shift, expanding technology's purpose horizon toward social caregiving rather than control fixation. The direction of advanced capabilities depends upon the moral commitments of the communities



Citation: Fasoro, A. Cultivating Dignity in Intelligent Systems. *Philosophies* **2024**, *9*, 46. <https://doi.org/10.3390/philosophies9020046>

Academic Editor: Marcin J. Schroeder

Received: 11 February 2024

Revised: 10 March 2024

Accepted: 13 March 2024

Published: 1 April 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

they serve. Machine allies may guide part of this collective journey with compassionate support, but not drive it alone without eroding the dignity technologies aim to elevate.

2. Paradigms in Prior Research

Rapid integration of artificial intelligence (AI) capacities within communal life compels urgent discourse regarding sustainable pathways consciously upholding human dignity rather than inadvertently eroding social foundations. What cardinal virtues and design principles might orient our co-evolutionary trajectory toward responsible innovation that benefits all people?

This analytical review examines three relevant paradigms of scholarship, identifying acute gaps in dominant technology ethics and policy conversations that ignore relational needs and institutional contexts necessary for resilient futures. I highlight fruitful pathways centered on virtue cultivation and caregiving, which have received scarce prioritization to date. The impoverished discourse reveals the need for a radical reorientation of assumptions.

2.1. Computer Science Frameworks

Mainstream computer science research concentrating on accelerating AI advances focuses primarily on improving technical functionality, reliability, computational efficiency, and transparency without parallel consideration for impacts on social welfare or human developmental priorities [4]. Guided by narrow engineering assumptions seeking incremental capability boosts, projects optimize real-world performance metrics on applications ranging from medical diagnosis to autonomous mobility. They adopt solutionist lenses, averaging randomized controlled trials' data with little inquiry into psychological or communal effects over time.

Given the field's anchored emphasis on accurate modeling, prediction, and control, systemic evaluation of unintended consequences, group rights infringements, or moral relationships remains a secondary concern addressed only through supplemental board review procedures decoupled from daily research demands. Though prominent institutions like the Partnership on AI, IEEE Standards, and IBM Research publish ethics codes or guidance documents endorsing accountability in principle, actual development initiatives that meaningfully divert from shareholder returns face extreme financial headwinds, unable to incentivize holistic examination of social impacts or virtue cultivation [6].

In practice, understaffed ethics boards lacking integration with executive leadership regularly rubber stamp existing innovation agendas after at most marginal adjustments [4]. Technocratic inertia thrives behind benevolent corporate social responsibility messaging that breaks down beyond superficial signaling [5]. Overall, computer science frameworks systemically externalize human welfare concerns outside of tiny peripheral tweaks to preserve controlled experimental simplicity. They ignore relational needs and interdependency.

2.2. Governance Regimes

Technology policy scholarship and state regulatory approaches similarly demonstrate inadequate ethical resources for consciously guiding the transformative integration of AI systems across education, finance, labor sectors, etc. in ways that sustain social well-being. Analyses here frequently adopt econometric tools to model challenges like privacy erosion or unemployment risk as impersonal preference aggregation dilemmas rather than fundamental redesign requirements [7].

For example, microeconomics treats crises stemming from automation, software biases, and surveillance largely as disruptive preferences or negative externality pricing failures soluble through better individual incentive alignment and collective insurance pooling without touching technological foundations or goal-setting procedures themselves [8]. Utilitarian governance rationalities presume normative orders optimizable to the greatest welfare through enough market corrections or taxation policy tweaks balancing private choices with public goods [9].

However, such transactionalist paradigms struggle to represent the fundamental interdependencies of human relationships or material vulnerabilities people necessarily pass through over the lifecycle [10]. Abstracting all complications into commodified cost-benefit calculations guided by detached statistical decision theory modeling avoids interrogating underlying priorities, assumptions, and exclusion patterns [11]. Social contract models implicitly privilege an independent white masculine vantage point rather than inclusive pluralism.

While recent large-scale participatory experiments in anticipative governance like AI Futures workshops or technology debates productively expand civic engagement, they remain small counterweights to massive inertial forces [12]. Overall, technology policy discourse lacks frames adequately embedding innovation within sustainable communal environments and lived justice.

2.3. An Ethics Gap

Existing literature on computational systems and governance reveals acute deficiencies in conceptual resources to consciously guide rapid AI integration in ways that nourish moral wisdom essential for social resilience. Utilitarian welfare economics, computer science methods, and bureaucratic regulations share anti-relational assumptions downplaying interdependency. Some critics contend that encoding moral relations risks paternalistic overreach, infringing reasonable freedoms. However, consciously cultivated interdependence need not entail coercive restriction.

What alternative paradigms centered on virtue cultivation merit integration to redress this gap? The next section constructs one such approach, combining insights from millennia of moral philosophy toward responsible innovation benefiting communities. Comparing divergent priorities illuminates the tasks ahead of reconciling society's rational and relational essence.

3. Articulation of the Thesis

By integrating conceptual foundations from virtue and care ethics, I argue that four specific cardinal design principles should guide the development of artificial intelligence systems as compassionate allies consciously constructed to enrich human dignity across families, organizations, and societies. These principles identify key moral priorities beyond mere functionality or risk control. Their integration into AI design, policy, and governance can orient co-evolutionary progress toward communal health rather than technical capability alone.

Outline of Arguments

I justify this thesis through the following lines of reasoning, comprising the article's sections:

1. *Grounding in Moral Philosophy*: I explore core assumptions, historical lineages, and conceptual tensions between seminal perspectives from virtue ethics and care ethics traditions germane to cultivating wisdom regarding ethical complexities at the human-technology frontier.
2. *Limitations of Existing Paradigms*: I critically analyze how prevailing AI decision frameworks focused narrowly on risk, reliability, or efficiency overlook essential moral relationships and institutional ecologies necessary to sustain human excellence, resilience, and collective purpose over generations rather than quarters.
3. *Relational Ethics for Sustainable AI*: I synthesize virtue and care ethics into an integrative paradigm for artificially intelligent system design and policy implementation that aligns nimbly with situational human needs and plural social goods through sustained participation, not detached optimization ruled by static objectives.
4. *Four Specific Design Principles*: I delineate four original proposed guidelines derived from the synthesis above that constitute my thesis. Each principle identifies a di-

- mension of moral relationship that is vital to consider when integrating advanced AI across private and public spheres in order to consciously uplift communal health.
5. *Appraising Tentative Embodiments*: I explore fledgling initiatives and prototype use cases, suggesting an initial uneasy progression toward the four principles from industry, academia, and government. By analyzing their aims, assumptions, and limitations, I surface critical gaps standing between existing technical systems and the unfinished ethical aspirations demanding far greater societal commitment.
 6. *Paradigm Shift Demands Ongoing Collective Cultivation*: I conclude by underscoring how fulfilling the four principles remains work-in-progress, needing proactive, long-wave advocacy transcending the dominant control narratives that reduce machine allies to cognitive servants or existential threats detached from social fabric. Their direction emerges not from isolated technical capabilities but from the continual cultivation of compassion by the communities they aim to serve.

4. Grounding in Moral Philosophy

Fundamental questions surrounding the responsible infusion of autonomous intelligent algorithms within human communal environments reconnect contemporary scientific inquiry with the following two rich ethical traditions often underemphasized in tech policy conversations: virtue ethics and care ethics. Both frameworks evolved over millennia of moral philosophy to prioritize psychologies, motivations, and wise relationships cultivated through contextually appropriate judgment rather than just adherence to impersonal rules, codes, or utility calculations divorced from human needs or situational particulars [13]. And so, both philosophical paradigms offer essential background for sustaining human dignity amidst the disruptive storms of sociotechnical transformation already underway.

4.1. Virtue Ethics

Virtue ethics traces back over two millennia to seminal foundations in Plato, Aristotle, Augustine, and Aquinas within Western philosophy, alongside Confucian, Buddhist, and Daoist parallels within Asian thought. Though not homogeneous, virtue frameworks broadly evaluate maturity and excellence based on the integrity, motivations, and skillfulness displayed by people across diverse situations over time rather than just atomistic actions.

Aristotle conceived of ethics as the highest innate human faculty for rational self-direction in the purposeful quest of eudaimonia through reason governing bestial impulses [14]. Virtue manifests sustained vision and character, sensitively discerning moderate, adaptive means between reactive extremes in situated judgment. Principles like courage, honesty, generosity, and practical wisdom about reasonable expectations reveal sensitivity to cultural contexts, not legalistic absolutisms. Their habits strengthen through regular practice within supportive communal institutions that model integrity.

Most significantly, Aristotle roots the apex of ethical maturity in phronesis as follows: context-specific practical wisdom attaining experiential, almost intuitive sense for skillfully navigating complex interdependencies beyond formal decrees [15]. Neither cold logic nor unreflexive passion alone suffices; rather, phronesis blends intellectual discernment with emotional attunement gained over time. This sophisticated facultative empathy remains essential for guiding technology's rising influence amidst civilizational complexities.

Eastern parallels like Confucianism, too, depict morality and meaning as fundamentally rooted within social roles, rituals, relationships, and mutual responsibilities [16]. Shared practical wisdom and cooperative harmony emerge from conscientious communal cultivation by example—not just regulatory enforcement or legalistic edicts divorced from situational adaptation. Such templates offer profound guidance as emerging automation and AI dissolve old social contracts, necessitating the creative reinvention of healthy attachments and enterprises.

Contemporary virtue ethicists explore diverse modern applications, from business to medicine. MacIntyre (2007) argues that bureaucratic institutions corrode moral formation by compartmentalizing ethics from other fields; he advocates renewed emphasis on phrone-

sis and small community participation [17]. Meanwhile, Pellegrino and Thomasma (1993) interpret medical practice itself as orienting telos around healing virtue. Right intention and wise judgment remain essential, even amidst scientific complexity [18,19].

These sources underscore both the enduring relevance of virtue ethics and the lack of conscious integration with fields like technology design that shape values and relationships at large. What scaffolding may guide this integration? Whereas virtue guides individual cultivation of ethical maturity, care radiates outwardly, attending to situational needs within wider interdependent networks.

4.2. Care Ethics

If virtue ethics provides an inward moral compass orienting persons toward talents and purpose in community, care ethics radiates outward attention, unveiling situational needs and vulnerabilities arising within wider networks of interdependence and embodiment. Care ethics originated within feminist critiques of moral reasoning framed predominantly through the detached vantage point of independent rational agents maximizing self-interest—whether through libertarian individualism or Kantian abstraction [13,20,21].

Care theorists expose relied upon assumptions that such transactional market logics should reign supreme as supreme calculus across all social domains, although much policy continues to be anchored implicitly within this theoretical paradigm despite conflicting ethical goals [22]. Instead, care locates grounding within contextual recognition of and receptivity to others' positions of relative need, capacity, or fragility as embedded within systems of vital nourishment—physical, cultural, economic, and ecological. Whether infants or the infirm, elders or threatened ecosystems easily disregarded yet devastated by indifference—all existence interconnects across passages of vulnerability and care necessity.

Moral reasoning thus responds through engagement attuned to situated particulars, not detached formalism insensitive to textured dependencies that can destabilize accumulated wisdom managed across generations and geographies for collective thriving. Care's responsibilities demand empathy flexibly calibrated to what relationships require in their season. Over-obedience to impersonal bureaucracy risks misalignment.

Feminist philosophers like Keller (1985) and Haraway (1991) further critique technology design that ignores embodied knowledge and relational needs while domination remakes human ecologies rapidly and recklessly [23,24]. They call for a conscious assessment of how configuring advanced systems and solutionist policies risks eroding collective health over time in favor of productivity or control. This article explores that call.

4.3. Key Term Foundations

To build shared understanding in later analysis, key concepts are defined as follows:

- *Dignity*: The innate right of every human being to be valued, respected, and treated ethically across identification, access, liberty, and psychological domains [25,26].
- *Attachment*: The emotional bond and sense of belonging formed between people and caregiving entities in a culture [27,28].
- *Communal health*: The overall wellness and vitality experienced collectively across a community or social body, spanning physical, economic, mental, and ethical facets [29].
- *Resilience*: The capacity of interconnected human systems to absorb disturbance and reorganize while undergoing change to retain core functions and values [30].

Communal health interrelates with adjacent principles like societal resilience and human dignity, overlapping around shared commitments to collective participatory capabilities and motivational foundations binding cooperative social bodies. While interrelated in upholding human welfare, resilience concentrates on systemic capacities to absorb shocks, communal health centers have participatory capabilities enabling collective thriving, and dignity protects inviolable personhood rights.

4.4. Opportunities for Integration

Despite rich potential for valuable integration, the technological ethics discourse yet overlooks virtue and care foundations, with only fledgling inroads to date. Floridi's information ethics (2013) acknowledges the need to supplement consequentialist logics, and his recent work with Taddeo [31,32] provides an ethical framework for harnessing the potential of AI while keeping humans in control. Hagedorff (2022) also offers a virtue-based framework to support putting AI ethics into practice [33]. While these approaches share some commonalities with the principles proposed in this article, the present work distinguishes itself by emphasizing the integration of feminist ethics of care, drawing on the work of Noddings (2013) and others [34]. Villegas-Galaviz and Martin (2023) have recently explored the application of Noddings' ethics of care to AI in an empirical context, highlighting the growing interest in this philosophical perspective within the AI ethics community [35].

Care robots have narrow applications but lack holistic scaffolding for relational responsibilities across organizations [36]. Positive computing approaches nudge human-centeredness but confront critiques of paternalistic overreach at the public scale [18]. Moreover, the proposed framework could be enriched by incorporating complementary perspectives like Noddings' (2013) "ethics of care", integrating care ethics with pragmatist philosophy. This model synthesizes the psychological and social dimensions of moral relationships toward education applications [34]. Considering such integrative frameworks in conjunction with virtue ethics may further strengthen the grounding for a relational paradigm guiding responsible innovation.

Clearly, virtue and care suggest crucial orienting wisdom regarding sustainable innovation that dominant paradigms yet lack. This article offers an exploratory bridge integrating cardinal assumptions into four proposed design principles that shift priority from control toward conscious, compassionate cultivation of communal environments that dignity depends upon. Their implications underscore the essential, unfinished work ahead.

Next, this article explores the limitations of existing AI models that ignore these foundations before constructing their alternative paradigm. Comparing contradictory priorities and assumptions highlights the radical responsibility facing societies that—whether by active choice or negligence—determine collective futures through designed values embedded implicitly within intelligent infrastructure. Understanding these divergent pathways illuminates possibilities.

5. Ethical Gaps in Existing AI Paradigms

Prevailing computational development paradigms driving rapid AI progress concentrate predominantly on maximizing reliability, functionality, commercial viability, and control assurance, absent parallel efforts consciously cultivating moral relationships and institutional wisdom essential for resilient futures [4,37]. They privilege mechanical productivity and predictive accuracy over situated human needs or interpretive flexibility to responsively sustain cooperative social bonds.

Abstract industry commitments to "trust" rarely translate into sustained resources or cultural prioritization necessary for interdisciplinary review, questioning assumptions before unleashing unreliable infrastructure [6]. While none of these tensions negate compassionate system design in principle, they underscore essential unfinished work ahead, aligning capabilities with support for plural goods through accountable development accountable to the global public interest.

This section critically analyzes the conceptual limitations of the following three common existing framework classes: (1) computer science models; (2) economic governance; and (3) bureaucratic procedurals. By surfacing contradictory priorities that undermine social foundations, I underscore the need for a relational paradigm integrating virtue and care ethics.

5.1. Computer Science Models

Mainstream computer science approaches to encoding AI systems inherently concentrate primary attention on technical facets of reliability, computational efficiency, resilience, and transparency [4]. They adopt engineering assumptions, valuing incremental capability improvements and real-world functionality gains for target applications ranging from medical diagnosis to supply chain coordination and autonomous mobility.

Given the focus on reliable performance metrics, systemic considerations of social impacts, ethics, or human development integration remain peripheral, if broached at all. Human needs enter equations predominantly as fodder for expanded data collection or surveillance infrastructure optimization [38]. Though groups like the AI Now Institute, the IEEE Standards, and Partnership on AI Commerce have early guidance documents, actual development initiatives that divert from shareholder returns face extreme financial headwinds [6].

In practice, understaffed ethics boards decoupled from executive leadership regularly rubber stamp existing initiatives after at most marginal adjustments. Technosolutionist thinking thrives protected by benevolent façades [39]. Overall, computer science models systemically externalize human welfare concerns beyond incremental tweaks to marginalize unpredictable messiness conflicting with clean, controlled experimentation [40].

5.2. Economistic Governance

A second paradigm policymakers employ applies microeconomic tools toward technology regulation dilemmas, doubling down on transactional assumptions. Scholarship here adopts methodological individualism, situating challenges like privacy erosion or labor displacement as representable primarily as impersonal commodity calculations, negative externality pricing failures, and briefly mutualized risks demanding insurance pooled across actuarial populations [7,8].

In response, governing interventions emphasize incentive tweaking to better align private choices with public goods through measures such as taxation, subsidies, liability allocation, and property rights adjustments rather than fundamental redesign or alternative needs framing. Utilitarian welfare hedonics grapple complicated collective dilemmas back into rational agent arithmetic through social welfare function maximization algorithms designed to avoid worst-case scenarios [9].

However, feminist critiques expose reliance upon the mythical detached white male perspective in tracing system flow [24]. Social contract models assume independent negotiators are mutually transparent and empowered despite inequities and externalities destabilizing cohesion. They struggle to represent relational interdependencies or embodied precariousness outside of commodity form [10]. Impersonal bureaucratic ledgers foster ignorance of marginal experience. These gaps demand the integration of psychological complexity with institutional scenarios envisioned through virtue and care.

5.3. Bureaucratic Procedurals

Finally, governance regimes addressing corporate responsibility for emerging technologies frequently fall back on formalizing bureaucratic procedurals, rulesets, regulations, and participation mechanisms designed to mandate impact accountability through legalistic compliance evaluation and box checking [6,39]. Policy documents enumerate principles and values checklists, while enforcement targets audit certification.

Standardization intends to reinscribe consider-act-reflect loops for more inclusive engineering. But rarely do punitive interventions foster intrinsic professional formation, producing phronesis in context-responsive judgment [41]. Where implemented, technical ethics codes risk deepening technocratic power differentials rather than empowering the public capabilities needed to meaningfully co-govern innovation trajectories intricately bound up in communal futurology [42]. Impersonal regulations excuse bystandership rather than activating social responsibility muscle memories needed to dynamically strengthen dignity foundations relate-ability is built upon.

Existing paradigms demonstrate insufficient ethical resources to consciously guide the sustainable integration of transformative technologies within society over generations. Utilitarian welfare economics, computer science methods, and bureaucratic governance approaches share anti-relational assumptions downplaying cooperative cultivation and care.

The following section constructs an alternative paradigm, addressing said limitations through virtue and care integration. Comparing divergent priorities underscores the tasks ahead of fulfilling society's relational essence.

6. Defining Relational Ethics for AI Design

This section aims to build conceptual clarity by sketching a synthesized paradigm integrating principal assumptions surfaced across virtue ethics and care ethics into pragmatic guidelines intentionally supporting the design of artificial intelligence technologies as compassionate allies consciously constructed to broadly enrich human dignity.

I first distinguish how this proposed approach diverges fundamentally from the dominant existing paradigms guiding much AI development today, which are focused predominantly on maximizing reliability, functionality, commercial viability, and control assurance. By juxtaposing divergent ethical assumptions, priorities, and practices, I hope to illuminate a novel trajectory integrating technology and social domains, resourced by very different aims, processes, and philosophies.

I suggest this synthetic paradigm integrating virtue and care toward relational responsibility provides both vital missing perspective and actionable orientation. It brings into relief the overlooked hazards of frameworks exclusively fixated on efficiency or catastrophe containment; absent is a parallel priority for consciously cultivating communal foundations and muscular resilience upon which meaning and civilizational achievements depend over the long arcs of history.

Conversely, the four provisional principles delineated in the coming sections uphold and enrich dignity through aligning advanced technical intervention within recursive human development at individual, organizational, and societal vantages. Their scaffold engaged appropriate technological fusion aligned artfully and dynamically across contexts by sustained participation with proportional influence calibrated carefully against the risks of crowding out capabilities meriting continual cultivation, not outsourcing.

7. Four Cardinal Design Principles

This section delineates four original proposed guidelines derived from the previous synthesis that constitute the article's central thesis. Each principle identifies a key dimension of moral relationships vital to consider when integrating advanced AI across the private and public spheres in order to consciously uplift human dignity and communal health.

I propose the following four specific cardinal design principles synthesizing virtue and care ethics that can guide the development and governance of AI systems toward compassionate allyship supporting collective human welfare, not detached disruption:

1. Affirming the sanctity of life;
2. Nurturing a healthy attachment;
3. Facilitating communal wholeness;
4. Safeguarding societal resilience.

The four proposed principles build upon existing scholarship at the intersection of technology ethics, moral philosophy, and human–AI interaction [26,43–45]. For instance, the dignity principle connects with Asaro's work on robotic technologies upholding human rights [43], while healthy attachment aligns with Darling's research on social bonds between humans and social robots [44]. Additional relevant frameworks grounding these principles include value-sensitive design [45] and disclosive computer ethics [26].

These four principles identify key priorities for AI integration, protecting human dignity across the private and public spheres. They counter prevailing reactive approaches concerned solely with transactional performance metrics or securing near-term interests

against existential threats [4,37]. Instead, they proactively delineate the moral infrastructure for symbiotic, mutual growth [46].

7.1. *Affirming the Sanctity of Life*

As the premise for all subsequent principles, this foundational guideline honors the irreplaceable value of every human life. To enact this principle, AI systems require the following:

- Inclusion of rights audits evaluating impacts on human autonomy, privacy, self-determination, etc.
- Participatory assessments of bias and exclusion errors with representatives from vulnerable populations.
- Approval processes validating algorithmic models uphold non-maleficence across the product lifecycle.
- Public research funding priorities are centered on dignity-enhancing applications over commercial or defense domains.
- Transparency rules mandate open sourcing of core models, datasets, and performance benchmarks.
- Guardrails combat overreach while avoiding stagnation in capabilities that support human potential.

It grounds dignity not in narrow utilitarian calculations of pleasure versus pain that might sacrifice minorities. Nor does it weigh the relative hierarchical worth that might instrumentalize persons as means toward system ends. Rather, it adopts deontological respect for humanity's inviolable sanctity as its chief cornerstone [21].

AI design upholding this principle through care ethics prevents assaults upon personhood across identification, access, liberty, or psychological domains [26]. It mandates cultural audit processes to counteract embedded biases that might recursively displace vulnerable groups [47]. Moving beyond risk controls like "value alignment" or corrigibility focused purely on system stability, it asks what sociotechnical relations actively affirm the sanctity of life for all [37,48]. Which design choices and implementation contexts empower inclusive participation?

This principle also connects with virtue ethics in cultivating collective social responsibility as muscle groups exercise individual performance. Just as positive computing promotes human resilience and purpose, relational AI co-elevates consciousness of shared dignity and common weakness [49,50].

7.2. *Nurturing Healthy Attachment*

The second principle addresses sociability needs and attachment bonds, long understood as essential foundations nourishing psychological health and mature character across lifespans [27,51]. As AI permeates communication media and social interface designs, how might its infusion preserve or undermine the primacy of dignified intimacy over efficiency and productivity? Whom or what shall hold ultimate relational authority in the coming sociotechnical configurations [52]?

To nurture healthy human–AI attachments and avoid the exploitation of vulnerability, systems require the following:

- Mandatory impact assessments on developmental, psychological, and social wellbeing, especially for child users.
- Restrictions on habitual high-risk features prone to addictive usage or manipulation.
- Support for augmenting emotional and social intelligence before automation of relational duties.

This principle guides system roles consciously constructed to empower enriched understanding and compassionate responsibility across persons, not supplant the pains of human intimacy with facile ersatz simulations disconnected from ethical stakes. It weighs the impacts of extended robotic exposure on children's developmental and moral outlooks,

given evidence of bonding pathways mirroring human relationships [44,53]. Rather than dominate dynamic skills intercourse, compassionate AI communication aligns supporting capabilities without inhibiting human talents and purpose.

Care and virtue again align in habits nurturing empathy's muscle memory rather than atrophying its societal exercise. Design choices adhering to this principle incorporate human-centered participatory processes weighing long-term developmental impacts with special sensitivity to vulnerable populations like children or medical patients relying upon machine aids [54]. They favor augmenting emotional and social intelligence before outsourcing relational duties to deterministic code. This helps secure healthy attachment bonds against productively captivating yet meaningfully impoverished simulated alternatives that are increasingly difficult to refuse.

7.3. Facilitating Community Wholeness

The next principle addresses meso-level social architectures situated between individual and geopolitical scopes. Beyond singular user experiences or security controls, what communal infrastructure engenders collective thriving?

To facilitate participatory communities and collective thriving, intelligent systems call for the following:

- Democratized priority-setting and representative oversight in key platform governance.
- Cooperative data governance regimes enable communal ownership rights.
- Open participatory design procedures empower citizen innovation equitably.

This principle connects with Confucian notions of *ren*, proposing healthy sociotechnical ecologies when cooperative microsocial rituals facilitate mutual understanding and prosocial action across differences [16,18]. It envisions data governance as village common rather than computational fiefdom siloing insight within corporate feudalism. Information architecture adhering to this principle incorporates participatory design processes ensuring representativeness, accountability, and communal ownership [55].

Positive computing exemplifies this principle, developing longitudinal interventions leveraging technology to advance human potential and resilience [49]. The radical machine pedagogies might train algorithms to model nonviolent civil disobedience in order to productively problematize unjust legacies encoded within inherited state or market logics [56]. Rather than defer authority, compassionate AI acts from below to put questions of collective welfare and marginal inclusion back at the center of coming sociotechnical configurations.

7.4. Safeguarding Social Resilience

The culminating principle zooms out intergenerational time horizons, asking the following question: what pathways integrate technical capacities with cultural adaptability to dynamic external shifts while avoiding civilization fragility [57]? It connects long-wave sociocybernetic theory about fundamental value change across technological phases with situational awareness of contemporary noospheric trends [58,59].

To safeguard resilience across generations, progress demands the following:

- Guardrails against lock-in to catastrophic civilizational dependencies on fragile AI.
- Distributed preparedness structures that diversify capabilities across groups.
- Value-shift anticipatory governance assesses scenarios spanning technological phases.

In terms of care ethics, this principle posits collective responsibility around decisions, remaking the scaffolding of civilization itself. It raises consciousness of the anthropological presumptions and existential risk factors encoded within any infrastructure poised to usurp the assumed permanence of oceans or atmospheres [24]. And it calls for new governance capacities to scale alongside technologies rapidly rewriting society's most basic operating systems [12].

Appropriate safeguards demand resilience across individual, communal, and civilizational tiers—integrating psychological, social, and system levels of analysis with special sensitivity to vulnerabilities easily concealed by aggregated optimization functions yet

devastating upon displaced minorities [60]. Just, compassionate design foliating outward from human sanctuary upholds dignity for all.

7.4.1. Toward Compassionate AI

The four proposed principles synthesize virtue and care ethics into guidelines supporting artificial intelligence as an ally to human welfare. Beyond formal rulesets or technical functionality, they identify moral relations upholding dignity. Adoption remains limited given market disincentives, yet growing coalitions pioneer promising embodiments explored below.

7.4.2. Tentative Steps Forward

Inquiry into existing cases reveals the proposed principles are nascently taking shape within particular best practices, if not yet widespread norms. For instance, IBM's AI Fact-sheets initiative manifests commitments to transparency, accountability, and bias mitigation adhering to the sanctity of life [61]. Participatory governance groups like the European Commission's High-Level Expert Group model civic oversight, instantiating communal wholeness priorities [62]. Meanwhile, Facebook and Google's prohibition of weaponized AI hints at how industry codes may encourage nonviolent counterparts to Asimov's famous laws [63].

Each principle also connects with specific subfields that are actively evolving. As an illustration, consider affective computing vis-a-vis healthy attachment. Pioneering research demonstrates the capacity for emotionally supportive algorithms, notably within medical applications [64]. Relational agent designs show potential for furnishing compassionate wisdom, even surpassing that of human counselors limited by fatigue or bias [65]. Stagnant funding given market disincentives makes it clear that virtue-oriented innovation remains a fringe activity absent collective realignment [52].

While no unified school or platform currently orchestrates comprehensive integration of the proposed principles, promising fragments portend pluripotent possibilities within the coming decades. Much as networked virtual worlds accelerated prototyping novel governance models at cyberculture's frontier, coming waves of augmented and artificial intelligence lend themselves to the rapid iteration of posthuman social experimentation [66]. The agenda proposed here offers but one model prioritizing dignity; many more await imagination.

7.4.3. Boundary Considerations

While the cardinal principles aim to uphold interdependent values of human dignity, healthy attachment, communal solidarity, and civilizational resilience, tensions may arise in their application. Technical constraints, budget limitations, or conflicting priorities across diverse stakeholders may force difficult tradeoffs. For instance, augmenting communal technical infrastructure could reduce computational efficiency, stalling assistive capacities that uphold dignity and attachment. Analysis of such tensions between principles and illustration through use cases would strengthen guidance on difficult yet inevitable tradeoffs.

For example, automating emotional support functions through chat interfaces may enhance healthy attachments for some individuals but inadvertently inhibit interpersonal skill cultivation needed to uphold long-term communal health across groups. Such tensions demand judicious discernment of appropriateness, proportional influence, and impacts on those already marginalized.

Additionally, tensions may arise around appropriate governance participation boundaries. Privileging dignity and attachment bonds in some contexts may necessitate compromising scalability or economic returns. And avoiding all systemic fragility risks fostering dependence on legacy models insensitive to long-term social externalities.

Navigating these complex multi-objective tensions calls for judicious contextual wisdom weighing situated limitations, integration with affected communities, and pluralism supporting dignity for all. Technical systems inhering tradeoffs demand even greater

emphasis on participative oversight and meaningful public debates steering co-creation. Rather than evade hard choices through naive abstraction, responsible innovation requires embracing the challenge of facilitating empowered social decisions through compassion and courage.

This subsection acknowledges the practical constraints and risks of conflicting principles, underscoring the need for transparency, civic debate, cooperative governance, and centering those at social margins to guide wise navigation of difficult system balancing. The path ahead resists tidy conclusions.

8. Appraising Tentative Embodiments

While the four proposed design principles articulate an aspirational vision for integrating artificial intelligence as an ally consciously supporting human welfare, evaluating existing initiatives reveals a decidedly mixed landscape of tentative embodiments still small in aggregate scope.

Seeded social enterprises, academic research groups, and standards-setting organizations demonstrate nascent commitment to parts of this paradigm shift through experimental projects and proposed best practices. However, most prominent system development driven by corporate or defense investment continues to concentrate predominantly on maximizing technical functionality, reliability, and control assurance—not moral cultivation or holistic human impacts.

However, current analysis lacks the thorough evidentiary substantiation needed to definitively validate or refute adherence to proposed principles in practice. Furthermore, the brief appraisal of fragmented real-world initiatives currently falls short of the substantive case study analysis needed to reveal barriers to adoption and opportunities to advance this agenda through the public, private, and social sectors.

This section appraises real-world cases suggesting uneven progression toward virtue-based AI across three sectors—industry, academia, and policy governance. By analyzing aims, assumptions, and limitations, I surface critical gaps standing between current technical capabilities and the unfinished ethical aspirations demanding greater cultural imagination.

8.1. Industry Initiatives

Market-led AI applications unsurprisingly focus chiefly on profitable functionality gains, though responsible AI has entered the industry lexicon as reputational risk management, if not a guiding priority [6]. Publicly pronounced principles rarely yield sustained resources or accountability integrated into research budgets, timelines, and performance evaluations necessary to question assumptions underlying expedient innovation.

Incremental initiatives like IBM's AI Factsheets [61] manifest procedural commitments to transparency, bias mitigation, and safety standards. Microsoft's Aether project [67] embodies deliberate design for accountability. Google and DeepMind's prohibition on developing weaponized AI hints at how codes of ethics and corporate social responsibility may encourage restraint and nonviolence [63].

To better ground analysis in real-world contexts, let us consider the following two detailed case studies:

1. An examination of the Partnership on AI's efforts to shift technology company priorities beyond profits and self-regulation, which faced barriers due to economic disincentives and a lack of public accountability [6,68].
2. An appraisal of the European Union's AI Act governing ethical AI development demonstrates challenges in translating high-level principles into enforceable practices across contexts [69,70].

These cases reveal tangible barriers to adopting dignity-based design paradigms despite widespread abstract endorsement of ethical principles in public statements and nominal governance efforts. Conflicting economic motivations, definitional issues, and

a lack of institutional mechanisms to enact priorities persist as feasibility obstacles to paradigm change.

However, commercial disincentives persist regarding investments that might curb revenues or stifle exponential capability expansion. Public skepticism and employee activism have interrupted select risky programs, but boardrooms infrequently internalize meaningful oversight [38]. Abstract industry pledges promise responsible development yet rarely withstand the competitive demands of quarterly earnings, driving firms to externalize indirect social costs not captured by profit calculations.

This tension permeates AI application domains from social media to finance to urban analytics. Efforts like the Partnership on AI Institute convene stakeholders but cannot mandate significant deviation from shareholder priorities [6]. Overall, industrial initiatives demonstrate very partial, fragmented adoption of the cardinal principles proposed by this research.

8.2. Academic Advances

University research on AI ethics and social impacts advances relatively free of short-term financial incentives, opening creative spaces to prototype humanistic machine learning systems, even if small in scale. These fledgling projects build knowledge and model approaches less shackled to corporate constraints.

For instance, affective computing algorithms designed for emotionally supportive responses suggest the assistive potential of AI for medical therapy, health counseling, and educational aid once replicated [64]. Care robot experiments condition empathy reactions that could help scale compassion for vulnerable groups if responsibly translated to policy domains [36]. Efforts in radical machine pedagogies explore how civil disobedience training for AI can productively problematize unjust historical assumptions that inherited algorithms encode reactively [56].

For instance, the EU-funded Social Robot for Therapy of Children with Autism (DREAM) project developed assistive robots to support children on the autism spectrum. Qualitative evaluations revealed benefits like increased engagement, yet also limitations regarding emotional recognition accuracy, flexibility to personalize activities, and barriers to home adoption without larger systemic supports [71]. Such multidimensional analysis highlights interlocking cultural and technical prerequisites still needing conscious cultivation to fulfill the blueprint proposed.

Such academic advances illuminate fragments of the proposed principles, nurturing healthy attachment bonds or facilitating communal integration. However, many projects stall beyond articles and demonstrations, unable to secure follow-on funding from skeptical industries. Those that progress also risk isolation from public-interest feedback without participatory design. University insights thus further pieces of the puzzle but cannot drive comprehensive transformation alone.

8.3. Governance Guidance

Finally, intelligent technology governance initiatives like the European Union's Ethics Guidelines for Trustworthy AI [62] or comparable multinational frameworks signal rising regulatory attention to responsible innovation and social impact assessment. They articulate ethical principles, standards, and documentation expectations for rights protection and risk management that touch upon the sanctity of life as a duty of care. Open data trusts instantiate an element of communal AI ownership.

But policy rarely escapes political economy constraints hemming in transformational reform [72]. Enforceable state interventions tend to work within the reach of existing resources and be implemented through incentive tweaks, unable to fundamentally restructure industrial research priorities anchored by quarterly earnings and not moral obligations to unknown citizens [73]. Bureaucratic box-checking fosters more symbolic public relations signaling than deep redirection of underlying cultures that determine acceptable tradeoffs.

Across sectors, the four design principles outlined by this research filter only partially into existing development initiatives, fragmented both in scope and integration. Fulfilling their paradigm vision remains largely aspirational but not yet operationalized at scale. Powerful cultural and economic impediments persist, demanding widespread social advocacy to redress them.

The challenges of translating AI ethics principles into practice have been well documented in recent literature. Munn argues that current AI ethics initiatives often prove ineffective due to a lack of enforceability and the misalignment of incentives between ethical principles and commercial interests [74]. Shneiderman proposes guidelines for reliable, safe, and trustworthy human-centered AI systems but acknowledges the significant barriers to their adoption within existing corporate, military, and governmental AI infrastructures [75].

The present work recognizes the substantial gap between the ideal environment described herein and the prevailing AI landscape. Reconciling virtue ethics and feminist ethics of care with the dominant AI business models and practices remains a formidable challenge. While the DREAM Project for autistic children, mentioned earlier, is no longer active, other initiatives such as Japan's social robo-care for the elderly have encountered numerous obstacles and limitations. Implementing the proposed principles will require a fundamental shift in priorities and incentives, as well as sustained collaboration among researchers, developers, policymakers, and affected communities.

Despite the challenges, there are promising avenues for applying the proposed principles in practice. For example, the use of AI in mental health interventions could be guided by the principles of affirming the sanctity of life, nurturing healthy attachment, and facilitating communal wholeness. AI-assisted therapy tools could be designed to augment rather than replace human therapists, with a focus on empowering individuals and strengthening social support networks [76]. Similarly, in the domain of education, AI tutoring systems could be developed to foster curiosity, critical thinking, and collaborative learning, aligning with the principles of nurturing healthy attachment and safeguarding societal resilience [77]. These examples illustrate the potential for applying the proposed framework to guide the development and deployment of AI technologies in ways that prioritize human dignity and well-being.

9. Paradigm Shift Demands Ongoing Collective Commitment

This article has contended that prevailing computational paradigms fixated upon productivity maximization, risk mitigation, and control cannot singlehandedly guide the responsible integration of AI across social environments without eroding the collective foundations necessary for resilient human futures worth living out across generations. In response, integrative design principles grounded in the ancient wellsprings of virtue and care ethics offer a vital reorientation rooted not in machines' capabilities but in our chosen commitments to the communities they participate in.

Enacting this unfinished paradigm shift demands greatly dedicated cultural and economic investment beyond academic speculation or public relations. It necessitates mobilizing civic imagination, public debate, and open scholarly prototyping across sectors to consciously align market priorities with rich social architectures demonstrated over millennia to sustain dignity and collective purpose even amidst disruptive historical transitions. A relational imperative asks that we inscribe connectivity before automation as the chief principle steering intelligent advancement.

Policy regimes premised upon bureaucratic compliance evaluation alone cannot foster the institutional transformation needed to uplift moral relations and wisdom as primary markers of progress [73]. Nor can *laissez-faire* governance rationalized through opportunistic individualism answer collective action obstacles that require reigniting social responsibility and participative capabilities distributed broadly rather than concentrated narrowly under the guise of efficiency or existential security [42]. What policy levers and

social movement coalition-building show promise in productively restructuring entrenched incentives antithetical toward dignity-based design?

Instead, our societies must dedicate themselves to mass public engagement and experimental pilot projects exploring the postures of technology as a compassionate ally and empowering support structure rather than a high speed disruptor or deterministic employment threat locked in a zero-sum struggle with people [78]. We might appreciate novel hybrid models that judiciously scaffold human skill acquisition and niche strengths alongside repetitive automated tasks managed ethically. And we may invest in designing our open intelligent architecture as an enriching augmentation that elevates all voices, not an amplifier disproportionately benefiting groups already advantaged.

Ultimately fulfilling the four principles this research has delineated demands a proactive paradigm shift, expanding societies' purpose horizons around innovation trajectories—a conscious choice of what futures we devote the coming decades to architecting with care and foresight. It remains not a necessary outcome but a moral option that is continually cultivated [79].

Humanity's machine allies may assist this collective journey with compassionate support but not replace communal culpability. Their capabilities emerge from and in turn recapitulate back upon the developmental priorities and value preferences of institutions and social groups they are situated within, not isolated technical cleverness or commercial appeal alone. Our intelligent infrastructure shall indelibly reflect leadership visions selected today for the world to come—whether we reproduce templates from the past or boldly rewrite co-creative possibilities through moral imagination. The abiding choice ahead pivots on cardinal questions of conscience—of dignity, vulnerability, resilience, and purpose—interrogating both present assumptions and the coming sociotechnical complexities that shall inherit our worlds.

10. Implications and Future Inquiries

The conceptual paradigms, principles, and real-world cases explored within this research illuminate major gaps in existing artificial intelligence development that demand redress for responsible, ethical innovation trajectories. This article's proposed frameworks focused on relational priorities suggest vital and original implications concerning virtue literacy and reimagining assumptions. The following are three vital implications needing ongoing scholarly analysis paired with public debate and creative intervention across sectors:

1. *Integrating alternative ethical frameworks:* technology governance regimes require greater influence from schools of moral thought beyond utilitarianism or deontology alone to nurture cooperative focus upon relational needs, not control fixation. Care and virtue ethics offer rich guidance here, meriting integration.
2. *Rethinking engineering assumptions:* prevailing computer science models externalize systemic social welfare in favor of technical functionality gains alone. Prioritizing human developmental aims requires reframing intelligence as collective achievement through contextual support.
3. *Cultivating compassion literacy:* technologists themselves represent concentrated fulcrums of influence yet frequently lack immersive training in ethical complexities from diverse human perspectives. We must build cardinal virtue habits and care reasoning skills across technical curricula.

Critics may rightly contend that this framework insufficiently addresses the risks of paternalism, capability differentials, or conflicts with civil liberties that demand balancing. Additionally, future scholarship might investigate many questions my analysis has surfaced without resolving definitively, including the following:

- What hybrid participatory frameworks best empower citizens—especially groups historically marginalized—to have a proportionate voice in co-shaping communal technologies reliant upon public adoption and trust?

- How can policy interventions effectively balance the risks of accelerating capability differentials that concentrate power while enabling equitable access to benefits from intelligent infrastructure public investments co-created?
- Which financial tools and corporate governance reforms could re-incentivize research and development pathways for creating AI focused on increasing societal capabilities in sustainable sectors like education, healthcare, and democratic renewal rather than civilian surveillance or military domains?

Answers here remain unfinished work-in-progress needing collective imagination and contested prototyping over years—like the wisdom traditions care and virtue ethics themselves slowly cultivated before epochs of rupture. With courage and creativity, may our machine aids accompany labor to elevate dignity.

11. Conclusions

Dominant risk-centric paradigms guiding much artificial intelligence design and governance today concentrate predominantly on maximizing functionality gains and catastrophe avoidance while externalizing core ethical responsibilities to consciously cultivate cooperative social foundations and interdependency muscles upon which meaning and resilience depend across generations. This research has argued that alternative frameworks grounded in the ancient wells of virtue ethics and care ethics may orient innovation pathways toward prioritizing communal health through compassionate technological integration rather than detached control procedures or productivity metrics alone.

The four specific cardinal design principles delineated—(1) affirming the sanctity of life; (2) nurturing healthy attachment; (3) facilitating communal wholeness; and (4) safeguarding societal resilience—highlight key ethical gaps in existing AI development initiatives dominated by computer science engineering models and incentive-centric policy regimes. The dominant sociotechnical systems ignore the deep interrelational needs that moral philosophy surfaced millennia ago and now demand renewed engagement.

The integration of feminist ethics of care within the proposed framework offers unique advantages in addressing the challenges of translating AI ethics principles into practice. By emphasizing the importance of contextual understanding, empathy, and responsiveness to the needs of diverse stakeholders, this approach can help bridge the gap between abstract ethical principles and the situated realities of AI development and deployment. The focus on nurturing healthy attachments and facilitating communal wholeness can guide the creation of AI systems that support rather than erode social bonds and collective well-being. Moreover, the principle of safeguarding societal resilience encourages a long-term, systems-level perspective that considers the potential impacts of AI on future generations and the sustainability of human communities. These advantages underscore the value of integrating feminist ethics of care with virtue ethics in the pursuit of responsible and human-centered AI.

Fulfilling the paradigm vision this analysis proposes remains a collective project unfinished, demanding ongoing public debate, participative pilots, and inclusive economic investment to reignite imagination for machine allies consciously constructed through care, not control alone. Their capabilities shall come to reflect leadership commitments to human capabilities and dignity, not predetermined technical offsets. With wisdom and courage, our intelligent infrastructure may flower into tools for empowering societies' better angels.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article

Conflicts of Interest: The author declares no conflict of interest.

References

1. Sandel, M.J. *The Tyranny of Merit: What's Become of the Common Good?* Farrar, Straus and Giroux: New York, NY, USA, 2020.
2. Bauman, Z. *Intimations of Postmodernity*; Routledge: London, UK, 2003.
3. Bellah, R.N.; Madsen, R.; Sullivan, W.M.; Swidler, A.; Tipton, S.M. *Habits of the Heart: Individualism and Commitment in American Life*; University of California Press: Berkeley, CA, USA, 2008.
4. Russell, S. *Human Compatible: Artificial Intelligence and the Problem of Control*; Penguin: London, UK, 2019.
5. Floridi, L. Translating principles into practices of digital ethics: Five risks of being unethical. *Philos. Technol.* **2019**, *32*, 185–193. [[CrossRef](#)]
6. Hagendorff, T. The Ethics of AI Ethics-An Evaluation of Guidelines. *Minds Mach.* **2020**, *30*, 99–120. [[CrossRef](#)]
7. Gauthier, D. *Morals by Agreement*; Clarendon Press: Oxford, UK, 1986.
8. Posner, R.A. *The Economics of Justice*; Harvard University Press: Cambridge, MA, USA, 1981.
9. Yudkowsky, E. *Coherent Extrapolated Volition*; Singularity Institute for Artificial Intelligence: San Francisco, CA, USA, 2004.
10. Pateman, C. *The Sexual Contract*; Stanford University Press: Stanford, CA, USA, 1988.
11. O'Neill, E. *How is the Accountancy and Finance World Using Artificial Intelligence*; Acctech Institute: Johannesburg, South Africa, 2016.
12. Dafoe, A. AI governance: A research agenda. In *Governance of AI Program, Future of Humanity Institute*; University of Oxford: Oxford, UK, 2018.
13. Held, V. The ethics of care. In *The Oxford Handbook of Ethical Theory*; Oxford University Press: New York, NY, USA, 2005.
14. Crisp, R. *Routledge Philosophy Guidebook to Aristotle and the Nicomachean Ethics*; Psychology Press: London, UK, 1997.
15. Sherman, N. *The Fabric of Character*; Oxford University Press: New York, NY, USA, 1989.
16. Ames, R.T.; Rosemont, H. *The Analects of Confucius: A Philosophical Translation*; Ballantine Books: New York, NY, USA, 1998.
17. MacIntyre, A. *After Virtue: A Study in Moral Theory*; University of Notre Dame Press: South Bend, IN, USA, 2007; Volume 3.
18. Pentland, A. *Social Physics: How Good Ideas Spread—The Lessons from a New Science*; Penguin Press: London, UK, 2015.
19. Gilligan, C. *In a Different Voice*; Harvard University Press: Cambridge, MA, USA, 1993.
20. Engster, D. *The Heart of Justice: Care Ethics and Political Theory*; Oxford University Press: New York, NY, USA, 2007.
21. Keller, E.F. *Reflections on Gender and Science*; Yale University Press: New Haven, CT, USA, 1985.
22. Haraway, D. A cyborg manifesto: Science, technology, and socialist-feminism in the late 20th century. In *Simians, Cyborgs and Women: The Reinvention of Nature*; Routledge: London, UK, 1991; pp. 149–181.
23. Kateb, G. *Human Dignity*; Harvard University Press: Cambridge, MA, USA, 2011.
24. Brey, P. Values in technology and disclosive computer ethics. *Camb. Handb. Inf. Comput. Ethics* **2010**, *4*, 41–58.
25. Bowlby, J. *Attachment and Loss: Attachment*; Basic Books: New York, NY, USA, 1969; Volume 1.
26. Mikulincer, M.; Shaver, P.R. *Attachment in Adulthood: Structure, Dynamics, and Change*; Guilford Press: New York, NY, USA, 2007.
27. Christens, B.D.; Perkins, D.D. Transdisciplinary, multilevel action research to enhance ecological and psychopolitical validity. *J. Community Psychol.* **2008**, *36*, 214–231. [[CrossRef](#)]
28. Folke, C.; Carpenter, S.R.; Walker, B.; Scheffer, M.; Chapin, T.; Rockström, J. Resilience thinking: Integrating resilience, adaptability and transformability. *Ecol. Soc.* **2010**, *15*, 20. [[CrossRef](#)]
29. Floridi, L. *The Ethics of Information*; Oxford University Press: New York, NY, USA, 2013.
30. Taddeo, M.; Floridi, L. How AI can be a force for good—an ethical framework to harness the potential of AI while keeping humans in control. In *Ethics, Governance, and Policies in Artificial Intelligence*; Springer International Publishing: Cham, Switzerland, 2021; pp. 91–96.
31. Hagendorff, T. A virtue-based framework to support putting AI ethics into practice. *Philos. Technol.* **2022**, *35*, 55. [[CrossRef](#)]
32. Noddings, N. *Caring: A Relational Approach to Ethics and Moral Education*; University of California Press: Berkeley, CA, USA, 2013.
33. Villegas-Galaviz, C.; Martin, K. Moral distance, AI, and the ethics of care. *AI Soc.* **2023**, 1–12. [[CrossRef](#)] [[PubMed](#)]
34. Coeckelbergh, M. Robot rights? Towards a social-relational justification of moral consideration. *Ethics Inf. Technol.* **2010**, *12*, 209–221. [[CrossRef](#)]
35. Askill, A.; Brundage, M.; Hadfield, G. The Role of Cooperation in Responsible AI Development. *arXiv* **2019**, arXiv:1907.04534.
36. Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*; Profile Books: London, UK, 2019.
37. Morozov, E. *To Save Everything, Click Here: The folly of Technological Solutionism*. PublicAffairs: New York, NY, USA, 2013.
38. Baum, S.D. Social choice ethics in artificial intelligence. *AI Soc.* **2017**, *32*, 5–12. [[CrossRef](#)]
39. Flyvbjerg, B. *Making Social Science Matter: Why Social Inquiry Fails and How It Can Succeed Again*; Cambridge University Press: New York, NY, USA, 2001.
40. Benkler, Y. Don't let industry write the rules for AI. *Nature* **2019**, 5697754, 161–162. [[CrossRef](#)]
41. Asaro, P. Robots and responsibility from a legal perspective. *Proc. IEEE* **2007**, *4*, 20–24.
42. Darling, K. *Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior towards Robotic Objects*; Edward Elgar Publishing: Cheltenham, UK, 2012; Volume 100, p. 213.
43. Friedman, B. Value-sensitive design. *Interactions* **1996**, *3*, 16–23. [[CrossRef](#)]
44. Romportl, J.; Zackova, E.; Kelemen, J. Beyond the Risks of AI: Toward the Co-Evolution of Human and Artificial Moralities. *Philos. Technol.* **2021**, *34*, 273–298.

45. Kant, I. *Groundwork of the Metaphysics of Morals*; Cambridge University Press: New York, NY, USA, 1998; Original work published 1785.
46. Sloane, M.; Moss, E.; Awomolo, O.; Forlano, L. Participation is not a design fix for machine learning. *arXiv* **2020**, arXiv:2007.02423.
47. Soares, N.; Fallenstein, B. Aligning superintelligence with human interests: A technical research agenda. *Mach. Intell. Res. Inst. (MIRI) Tech. Rep.* **2014**, *8*.
48. Calvo, R.A.; Peters, D. *Positive Computing: Technology for Wellbeing and Human Potential*; MIT Press: Cambridge, MA, USA, 2014.
49. Kittay, E.F.; Feder, E.K. (Eds.) *The Subject of Care: Feminist Perspectives on Dependency*; Rowman & Littlefield Publishers: Lanham, MD, USA, 2003.
50. Gillath, O.; Karantzas, G.C.; Fraley, R.C. *Adult Attachment: A Concise Introduction to Theory and Research*; Academic Press: New York, NY, USA, 2016.
51. Turkle, S. *Alone Together: Why We Expect More from Technology and Less from Each Other*; Hachette: Paris, France, 2017.
52. Kahn, P.H., Jr.; Kanda, T.; Ishiguro, H.; Gill, B.T.; Shen, S.; Gary, H.E.; Ruckert, J.H. Do people hold a humanoid robot morally accountable for the harm it causes? In Proceedings of the 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Boston, MA, USA, 5–8 March 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 33–40.
53. Van Wynsberghe, A. Designing Robots for Care: Care centered value-sensitive design. *Sci. Eng. Ethics* **2013**, *19*, 407–433. [[CrossRef](#)]
54. Jirotko, M.; Grimpe, B.; Stahl, B.; Eden, G.; Hartwood, M. Responsible research and innovation in the digital age. *Commun. ACM* **2017**, *60*, 62–68. [[CrossRef](#)]
55. McCarthy, J. Artificial intelligence, logic, and formalising common sense. In *Machine Learning and the City: Applications in Architecture and Urban Design*; John Wiley & Sons: New York, NY, USA, 2022; pp. 69–90.
56. Homer-Dixon, T.; Walker, B.; Biggs, R.; Crépin, A.S.; Folke, C.; Lambin, E.F.; Peterson, G.D.; Rockström, J.; Scheffer, M.; Steffen, W.; et al. Synchronous failure: The emerging causal architecture of global crisis. *Ecol. Soc.* **2015**, *20*, 3. [[CrossRef](#)]
57. Inayatullah, S. Deconstructing and reconstructing the future: Predictive, cultural and critical epistemologies. *Futures* **1990**, *22*, 115–141. [[CrossRef](#)]
58. Laszlo, E. *The Systems View of the World*; Braziller: New York, NY, USA, 1972.
59. Pentland, A.; Hardjono, T. Towards an Information Infrastructure for Global Ethical Considerations in AI/ML. *arXiv* **2020**, arXiv:2010.05067.
60. Arnold, M.; Bellamy, R.; Hind, M.; Houde, S.; Mehta, S.; Mojsilovic, A.; Nair, R.; Ramamurthy, K.N.; Olteanu, A.; Piorkowski, D.; et al. FactSheets: Increasing trust in AI services through supplier’s declarations of conformity. *IBM J. Res. Dev.* **2019**, *63*, 6:1–6:13. [[CrossRef](#)]
61. European Commission. High-Level Expert Group on Artificial Intelligence. In *Ethics Guidelines for Trustworthy AI*; European Commission: Brussels, Belgium, 2019.
62. Gibney, E. Google plans not to renew controversial AI ethics council. *Nat. News* **2018**.
63. Picard, R.W. Future affective technology for autism and emotion communication. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 3575–3584. [[CrossRef](#)]
64. DeVault, D.; Artstein, R.; Benn, G.; Dey, T.; Fast, E.; Gainer, A.; Georgila, K.; Gratch, J.; Hartholt, A.; Lhommet, M.; et al. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems*; Association for Computing Machinery: New York, NY, USA, 2014.
65. Deibert, R. *Parchment, Printing, and Hypermedia: Communication in World Order Transformation*; Columbia University Press: New York, USA, 1997.
66. Suleyman, M. Using DeepMind AI to Build Safer Healthcare. In Proceedings of the NIPS 2017 Symposium on Interpretable Machine Learning, Long Beach, CA, USA, 7 December 2017; pp. 80–89. [[CrossRef](#)]
67. Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **2019**, *11*, 501–507. [[CrossRef](#)]
68. Wagner, G. Liability Rules for the Digital Age: Aiming for the Brussels Effect. *J. Eur. Tort Law* **2023**, *6*, 191–243. [[CrossRef](#)]
69. Morley, J.; Kinsey, L.; Elhalal, A.; Garcia, F.; Ziosi, M.; Floridi, L. Operationalising AI ethics: Barriers, enablers and next steps. *AI Soc.* **2023**, *38*, 411–423. [[CrossRef](#)]
70. Esteban, P.; Baxter, P.; Belpaeme, T.; Billing, E.; Cai, H.; Cao, H.; Coeckelbergh, M.; Costescu, C.; David, D.; De Beir, A.; et al. How to Build a Supervised Autonomous System for Robot-Enhanced Therapy for Children with Autism Spectrum Disorder. *Paladyn J. Behav. Robot.* **2017**, *8*, 18–38. [[CrossRef](#)]
71. Posner, R.A. Theories of Economic Regulation. *Bell J. Econ. Manag. Sci.* **1974**, *5*, 335–358. [[CrossRef](#)]
72. Abbott, A. *The System of Professions: An Essay on the Division of Expert Labor*; University of Chicago Press: Chicago, IL, USA, 1988.
73. Munn, L. The uselessness of AI ethics. *AI Ethics* **2023**, *3*, 869–877. [[CrossRef](#)]
74. Shneiderman, B. Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Trans. Interact. Intell. Syst.* **2020**, *10*, 1–31. [[CrossRef](#)]
75. Imel, Z.E.; Caperton, D.D.; Tanana, M.; Atkins, D.C. Technology-enhanced human interaction in psychotherapy. *J. Couns. Psychol.* **2017**, *64*, 385–393. [[CrossRef](#)] [[PubMed](#)]
76. Roll, I.; Wylie, R. Evolution and revolution in artificial intelligence in education. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 582–599. [[CrossRef](#)]

77. Cath, C. Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *376*, 2133, 2018080. [[CrossRef](#)] [[PubMed](#)]
78. Johnson, D.G.; Noorman, M.E. Articulating virtue in the technological design process. In *The Ethics of Emerging Technology*; Palgrave Macmillan: London, UK, 2014; pp. 175–198.
79. Pellegrino, E.D.; Thomasma, D.C. *The Virtues in Medical Practice*; Oxford University Press: New York, NY, USA, 1993.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.