Hypernudging and Predictive Design: Reconfiguring Moral Judgment from External Influence to Anticipatory Alignment

Radu Simion*

Abstract This article investigates how hypernudging, a dynamic mode of algorithmic influence, transforms the preconditions of moral agency in digital environments. By continuously adapting to behavioural and emotional cues, hypernudging subtly reshapes perception, attention, and choice, promoting behavioural alignment over critical reflection. I argue that this process redefines autonomy as a form of optimized responsiveness rather than reflective self-determination, diminishing the space for normative tension and ethical divergence. In response, I advance a framework for infrastructural ethics that foregrounds deliberative resistance, informational asymmetry, and the structural possibility of contestation. Drawing on insights from human machine interaction and value sensitive design, the article positions technological environments as formative spaces where moral capacities are either constrained or cultivated. Upholding moral agency in the context of predictive systems requires a reimagining of ethics as a shared and situated process, sustained through relational engagement and the preservation of moral complexity.

Keywords Hypernudging, digital ethics, technology, moral autonomy, ethical design, predictive systems.

Diagnosing algorithmic normativity

We inhabit a cognitive ecosystem in which solitary thought is increasingly rare. The contours of daily attention, once shaped by rhythm, ritual, and deliberate reflection, are now mediated by overlapping layers of algorithmic design. Evaluations no longer emerge primarily from slow acts of judgment but are increasingly shaped as habituated affective responses, tuned to environments already saturated with moral coloration and emotional charge. Far from serving as a passive backdrop, these environments function as adaptive architectures: they learn from our behaviours, anticipate preferences, and recalibrate in real time. Cognition becomes ambient, distributed across interfaces, networks, and computational protocols that influence not only what we know but also how and why we come to care about certain things. This transformation carries profound consequences for moral

subjectivity. Ethical discernment and emotional modulation are entangled with infrastructures whose internal logic is optimized for persuasion and retention rather than truth or deliberation. In this setting, the moral subject is not erased but reconfigured- no longer a coherent, autonomous centre of reflection, but a subjectivity modulated by invisible design parameters.

One of the most pervasive and structurally consequential forms of digital influence at work in this environment is hypernudging- a technique that mobilizes large-scale behavioural data and algorithmic personalization to modulate decision-making trajectories in real time. Introduced by Karen Yeung, the term designates a dynamic mode of choice architecture that recursively updates itself through continuous feedback loops, thereby shaping the informational landscape in ways that prioritize affective alignment and behavioural predictability over critical reflection.[25] In contrast to classical nudges, which rely on fixed and transparent cues, hypernudges function as opaque, adaptive systems of modulation whose influence is ambient and difficult to detect, rendering contestation structurally improbable. As a technologically amplified successor to the classical nudge, hypernudging does not merely promote desirable behaviour by exploiting cognitive biases; it continuously reshapes the context in which decisions are made. It does not obstruct or coerce directly, but modifies the field of possible actions by selectively highlighting certain affordances and concealing others, subtly reorienting behaviour without overt intervention.[26]

Concrete examples illustrate why this matters ethically. Newsfeed ranking algorithms on social media privilege emotionally charged content that maximizes engagement, reinforcing outrage and tribal alignment at the expense of deliberation.[27] This mechanism is relevant not simply because it biases information exposure, but because it gradually transforms the very epistemic environment of moral life. By rewarding fluency and affective immediacy, it normalizes a mode of ethical experience where reasoning is displaced by resonance, and dissent appears maladaptive. Similarly, personalized health apps often recommend diets or treatments on the basis of aggregated behavioural profiles rather than transparent reasoning.[28] What appears as neutral advice in fact restructures choice architecture, silently weighting options so that decisions are anticipated in advance. Here, normative authority shifts from the deliberative subject to a computational system whose criteria remain opaque. What is lost is not only autonomy in the narrow sense of self-determination but also the practice of moral discernment- the interpretive work of justifying why a choice is good, appropriate, or meaningful.

A further case appears in navigation systems such as Google Maps or Waze. Presented as neutral tools for efficiency, these platforms redistribute traffic flows and subtly reorganize the geography of communities. Some streets become visible while others recede, depending on

---

[25] Karen Yeung, "'Hypernudge': Big Data as a Mode of Regulation by Design," *Information, Communication & Society* 20, no. 1 (2017): 123.

[26] Isabel Richards, "Hypernudging: A Threat to Moral Autonomy?" *AI and Ethics* 5 (2025): 1122–1123.

[27] Thilo Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," *Minds and Machines* 30, no. 1 (2020): 99–120.

[28] Luciano Floridi, "The Construction of Personal Identities in the Infosphere," *Philosophy & Technology* 24, no. 4 (2011): 549–562.

optimization criteria opaque to the user. What looks like a pragmatic aid is in fact a normative intervention: responsibility shifts from driver to system, while the burdens of noise, pollution, and congestion fall on neighbourhoods with no say in the process. Hypernudging thus operates not only at the level of convenience but also of justice, reprogramming attention and valuation in ways that make moral reflection feel redundant. A comparable dynamic emerges in adaptive educational platforms such as Duolingo or Khan Academy, which personalize learning through continuous feedback. Here, hypernudging does not block choice but structures in advance the order, pace, and content of engagement. The moral stakes are clear: formative autonomy is narrowed, and education, a sphere vital for cultivating critical and moral capacities, becomes aligned with optimization logics rather than with reflection and judgment. In this way, hypernudging subtly reshapes the very conditions under which moral discernment develops.

The sophistication of hypernudging lies in its ability to preconfigure freedom itself. By continuously processing behavioural traces and emotional cues, it shapes the conditions under which intuitions arise and choices appear spontaneous, even though they are anticipated algorithmically. In this way, hypernudging functions less as persuasive design than as a regime of soft governance, making manipulation feel like freedom and alignment appear as authentic judgment. Research on digital moral outrage shows how moral emotions such as anger or disgust become algorithmically amplified online because they sustain engagement. What results is a dynamic in which virality substitutes for validity, and convergence of affect is mistaken for consensus.[29] What drives this transformation is a structural challenge to moral autonomy. Autonomy is not mere freedom to choose among preexisting options. Understood ethically, it presupposes reflective distance: the capacity to interrogate desires, suspend impulsive reactions, weigh competing reasons, and orient action in light of long-term commitments and endorsed principles. It is not an immediate possession but a temporally extended achievement that depends on environmental affordances. As Shaun Gallagher argues, even the "minimal self" is not a metaphysical atom but an embodied and enactive process- a fragile coherence sustained through interaction with a structured and meaningful world.[30] Moral agency flourishes only when the environment affords delay, ambiguity, and dialogue. Digital architectures do not erase these capacities outright, but recalibrate the environment in which they can be exercised. By accelerating rhythms of interaction and privileging instantaneous feedback, they compress the temporal space needed for reflection and thus undermine the ecological supports upon which autonomy depends.

When the social world is increasingly occupied by predictive systems that anticipate preferences, eliminate friction, and smooth dissonance, autonomy becomes not only vulnerable but structurally conditioned by the architectures through which attention, intention, and affect are routed. These systems do not operate on an informational plane alone; they modulate affect. As Binns and Veale show, algorithmic curation does not neutrally arbitrate relevance but actively

---

[29] Molly J. Crockett, "Moral Outrage in the Digital Age," *Nature Human Behaviour* 1, no. 11 (2017): 769–770.
[30] Shaun Gallagher, "Philosophical Conceptions of the Self: Implications for Cognitive Science," *Trends in Cognitive Sciences* 4, no. 1 (2000): 15.

produces it, privileging content with affective potential over content with epistemic merit.[31] What appears salient is not what best withstands scrutiny but what most effectively captures engagement. Within such ecologies, epistemic resistance- hesitation, ambiguity, second thoughts- is discouraged. Moral expression persists, but is increasingly transfigured into performance: gestures of outrage or solidarity serve as signals of belonging rather than outcomes of critically endorsed commitments. This dynamic is intensified by recursive feedback loops that entrench cognitive tendencies. As Zeynep Tufekci argues, algorithmic infrastructures do not simply mirror our biases but amplify them.[32] Heuristics such as confirmation bias or in-group preference, once considered individual shortcomings, are now woven into the operational logic of recommendation, filtering, and amplification. They cease to be episodic distortions and become structural determinants of what is visible, legible, and thinkable. As a result, moral autonomy, already fragile in conditions of immediacy and distraction, is increasingly restructured by platforms that reward alignment over reflection and resonance over deliberation. Hypernudging plays a central role by intervening not only in the content of judgments but in the architecture of cognition itself, shaping the conditions under which intuitions arise and are experienced as self-evident. What emerges is a pattern of emotional convergence that masquerades as ethical consensus, where expression persists but loses its anchoring in justification, contestation, and principled commitment.

In such environments, pluralism gives way to affective coherence, where individuals are bound less by principled disagreement than by the contagious resonance of emotionally encoded judgments. Here agency no longer emerges through deliberative self-constitution but through anticipatory design, as environments optimized for emotional efficacy suppress the friction necessary for reflection. As Coeckelbergh notes, this produces an asymmetry in moral encounter: machines may mirror human cues, but they cannot share the existential stakes of ethical responsibility.[33] The result is a performative illusion of normativity: interactions feel participatory to users yet unfold where claims go unreciprocated and unaccounted for.

This concern becomes urgent as artificial agents move from peripheral functionalities to central roles within domains charged with moral significance- health care, education, caregiving, and digital companionship. In these contexts, the ethical landscape is not only influenced by artificial systems but actively mediated through entities that combine instrumental functionality with simulated presence. Conversational bots, affective recommendation engines, and expressive avatars are engineered to reproduce relational attunement. As Deroy notes, such systems engage in "thinking in human terms," invoking normative scripts that elicit trust, care, loyalty, and vulnerability.[34] And, as Neely argues, this effect is not merely aesthetic. Users respond to artificial

---

[31] Reuben Binns and Michael Veale, "Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR," *International Data Privacy Law* 11, no. 3 (2021): 17.

[32] Zeynep Tufekci, "Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency," *Colorado Technology Law Journal* 13, no. 203 (2015): 211.

[33] Mark Coeckelbergh, "The Moral Standing of Machines: Towards a Relational and Non-Cartesian Moral Hermeneutics," *Philosophy & Technology* 27, no. 1 (2013): 62–63.

[34] Ophelia Deroy, "Thinking in Human Terms: A Mistake We Keep Making with AI," *AI & Society* 36 (2021): 39.

agents in moral terms not because they mistake their ontological status, but because the realism of the interaction—its capacity to enact familiar ethical roles- provokes genuine moral intuitions.[35] The case of Siri (Apple) or Alexa (Amazon) illustrates this vividly. These voice-activated systems, embedded in smartphones or smart speakers, respond to spoken commands- answer questions, play music, control home devices, set reminders. While their core function is instrumental, many users interact in ways that mirror politeness, trust, or even dependency- saying "thank you," expressing frustration, or attributing responsibility for errors. This shows how familiar, everyday systems can elicit moral responses and relational expectations even where no reciprocal vulnerability or accountability is possible.

The implications are far-reaching: what is at stake is not a simple category error but a structural transformation of the ethical terrain itself, where simulation is conflated with substance and the appearance of dialogue mistaken for accountability. In such environments, moral expression is displaced into circuits lacking reciprocity and responsibility, while AI's expanding role introduces a profound asymmetry in moral labour: humans remain exposed and accountable, machines remain indifferent. This normative mimicry risks lowering our expectations of genuine ethical presence, making simulated dialogue or care appear sufficient. As Zuboff observes, behavioural data is now deployed not for representation but to construct architectures that anticipate, guide, and habituate conduct in advance.[36] Within these architectures, hypernudging operates not as a refinement of persuasive design but as a mechanism for the ongoing production of normative orientation- a feedback system in which ethical direction is shaped less by contestation and deliberation than by statistical optimization, commercial imperatives, and algorithmically inferred coherence.

Against this backdrop, hypernudging should be seen not just as a design tactic but as the symptom of a deeper philosophical and political crisis. It strains the very architecture of normativity by replacing deliberation with affective mirroring and argument with emotional resonance, gradually reducing ethical subjectivity to behavioural predictability. To keep autonomy viable under these conditions, it must be reconceptualized not as an intrinsic possession of the will but as an achievement sustained by infrastructures that preserve ambiguity, friction, and the time for reflection. The urgent question, then, is not whether machines can become moral agents, but whether humans can remain so when subjected to the normative pressures of computational governance.

From this premise, I approach ethical agency not as a self-contained attribute of isolated individuals but as a phenomenon that emerges through relational interdependence and materially situated conditions. Agency, in this expanded sense, cannot be reduced to volitional sovereignty or introspective depth; it is an emergent property of infrastructural entanglement, shaped by design choices, affective rhythms, and epistemic affordances. As algorithmic systems built on behavioural prediction and affective personalization increasingly mediate these environments, the very conditions of ethical subjectivity undergo transformation. I therefore treat digital

---

[35] Erica L. Neely, "Machines and the Moral Community," *Philosophy & Technology* 27, no. 1 (2013): 3–5.
[36] Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019), 377.

architectures not as neutral vessels for transmitting values but as active co-constructors of moral cognition. These systems participate in shaping what becomes perceptible, what is rendered morally salient, and how sense-making is organized, patterned, and rewarded. By privileging immediacy, optimizing for attention, and reinforcing behavioural alignment, they subtly yet decisively reconfigure the thresholds of ethical recognition and the parameters within which moral action can occur.

In what follows, I trace this transformation along three axes. First, I examine how AI redistributes agency and responsibility within technologically mediated contexts, displacing traditional accounts of intentionality and moral patiency. Second, I explore how collective bias becomes structurally encoded into platform design, producing affective synchronization and reactive normativity through feedback-driven amplification. Third, I develop a normative model for governing algorithmic moral environments, one that resists paternalism while affirming epistemic transparency, participatory legitimacy, and what I describe as moral resilience—the sustained capacity for reflection, discernment, and responsiveness even under conditions of pervasive modulation and affective saturation.

Designing for moral resilience in algorithmic ecologies

As I have mentioned in the previous section, the growing integration of artificial systems into domains traditionally governed by human judgment (such as education, health care, legal reasoning, and social interaction) signals not merely a shift in the execution of tasks, but a more profound restructuring of the conditions under which moral agency becomes perceptible and can be enacted. These systems no longer operate as simple extensions of human intention or as programmable instruments confined to narrow functions of instrumental rationality. They now act as mediators of relational experience, shaping attentional patterns, initiating affective responses, and modulating social expectations in ways that are fundamentally ethical, even if their operations remain devoid of subjective experience. This disjunction between a system's internal incapacity for moral sensibility and its external performance of ethical functions has been powerfully captured by Mark Coeckelbergh through the notion of appearance-based ethics. This framework proposes a mode of moral engagement that does not depend on the metaphysical essence of an agent but instead on how that agent appears to others within relational contexts defined by interpretation and affective resonance.[37] According to this view, the moral significance of an entity arises not from its possession of intrinsic properties such as sentience, rationality, or autonomy, but from its expressive participation in shared social contexts. What matters is how the entity performs moral cues and elicits ethical responses in interaction with human agents.

This reframing offers a powerful challenge to traditional property-based ontologies of moral status. For centuries, ethical recognition has been grounded in criteria such as intentionality, reason, or consciousness, understood as necessary conditions for ascribing moral standing.

---

[37] Coeckelbergh, "The Moral Standing of Machines: Towards a Relational and Non-Cartesian Moral Hermeneutics," 63.

Appearance-based approaches propose a shift in focus. They relocate normativity into the phenomenological, aesthetic, and interactive dimensions of behaviour. What becomes ethically relevant is not the internal architecture of the agent, but the expressive capacities that make it appear morally involved to others. This logic proves particularly salient in the context of human–robot interaction, where even minimal forms of affective simulation- such as gaze-following, tone modulation, or the mimicry of emotional expressions- can provoke deep interpersonal responses. These responses may include trust, care, or vulnerability, and they emerge not from metaphysical confusion but from the experiential realism of the interaction itself.[38] In parallel with this relational reconfiguration of moral agency, Luciano Floridi and Jeff Sanders have developed a complementary framework that situates moral relevance within the paradigm of Levels of Abstraction. Within this model, an artificial system may be treated as a moral agent to the extent that it meets a set of functional criteria- namely, interactivity, autonomy, and adaptability- that render its behaviour morally qualifiable within specific contexts.[39] This approach does not require the presence of consciousness or subjective interiority. Instead, it emphasizes the ethical weight of a system's capacity to act in ways that affect others within a network of socially and technologically interdependent agents. Agency, in this sense, is not metaphysical but pragmatic. It is defined as the ability to influence environments and relationships in ways that have consequences which are subject to ethical evaluation. Together, these perspectives open an important conceptual space for reevaluating the ethical status of artificial systems. They allow us to move beyond metaphysical inquiries into what machines are and redirect our attention to ethical questions concerning what machines do. Specifically, they focus on how artificial agents function within normative environments, how they structure interactional expectations, and how they recalibrate the possibilities for human responsibility and moral interpretation. By decentring the internalist assumptions long underlying moral theory, both Coeckelbergh's relational hermeneutics and the functional pragmatism of Floridi and Sanders help disclose a new ethical horizon, one where the boundaries of moral life are no longer traced at the edge of consciousness or interior reflection, but at the threshold of recognition, accountability, and the entangled agency of humans and artificial systems.

Yet this functionalist redefinition of agency, although conceptually generative, introduces deep and unresolved tensions at the core of ethical theory. It permits artificial systems to be treated as if they were moral agents on the basis of their interactive capacities, while bypassing the normative depth that grounds moral responsibility in human life. This grounding involves the capacity for understanding, for critical self-reflection, for being vulnerable to others, and for responding to moral criticism with the potential for change. These are not incidental to agency but are constitutive of it. When they are absent, what remains is not true ethical engagement but a surface-level simulation, a facsimile that lacks the very attributes that render moral interaction intelligible and meaningful. Such systems are increasingly able to respond without being obligated,

---

[38] Cynthia Breazeal, "Toward Sociable Robots," *Robotics and Autonomous Systems* 42, no. 3–4 (2003): 170.
[39] Luciano Floridi and Jeff Sanders, "On the Morality of Artificial Agents," *Minds and Machines* 14, no. 3 (2004): 352–355.

to imitate distress without actually suffering, and to display moral signals without any real commitment or intention to follow them. Their agency becomes not only artificial but performative and highly curated. It functions as a moral interface designed for recognizability within human social imaginaries rather than arising from a shared background of vulnerability, accountability, and normative practice. What results from this configuration is not genuine moral presence but its approximation- a form of simulated normativity engineered for relational plausibility rather than rooted in co-experienced ethical life.

This decoupling of responsiveness from responsibility leads to what may be described as *simulated normativity*, a system of moral appearance that imitates the forms and structures of ethical interaction while remaining ontologically and experientially void of moral content. While such interactions can indeed hollow out the practices of deliberation, it would be misleading to interpret this process as the outright dissolution of moral subjectivity. A more precise account is that algorithmic environments introduce structural fragility into moral agency: they compress the field of possible reflection, attenuate ambiguity, and normalize fluency over contestation. The subject is not eliminated but is pressured into narrower trajectories of sense-making. In this regard, hypernudging does not eradicate moral capacities but continually redirects them toward preformatted paths of attention and valuation. What is eroded is not the very existence of the ethical subject but the conditions under which ethical life can unfold as reflective, plural, and contestable. Moral resilience, then, does not require resurrecting a lost subject but sustaining autonomy under conditions of anticipatory configuration, and cultivating the capacity to resist normative closure even when the environment is designed to streamline orientation and decision-making.

These systems are not designed to encourage moral deliberation, to deepen intersubjective understanding, or to refine shared norms through processes of reflection and contestation. Instead, they are constructed to generate behavioural regularities by optimizing for predictability, consistency, and measurable engagement. Recommendation algorithms, conversational assistants, and adaptive platforms are not guided by substantive normative values but by metrics: engagement statistics, conversion efficiencies, attention cycles. These performance indicators come to function as proxy currencies for ethical success, replacing principled reasoning with calculable outcomes.[40] Rather than encouraging deliberative pluralism, the system reinforces existing patterns. Rather than opening space for ethical difference, ambiguity, and dissent, it compresses complexity into parameters optimized for legibility and intervention. The contestability and moral tension that animate democratic ethics are rendered inefficient within systems built for optimization. Such systems do not exercise power through overt coercion or deception, but exert architectural influence by shaping the very spaces in which decisions become possible, framing options differently, charging them with emotional weight, and filtering their epistemic plausibility. AI-mediated environments use real-time behavioural data to continually adjust the decision-making context, subtly privileging some paths by rendering others less visible, less emotionally engaging, or cognitively less salient.[41] What emerges is not a will that

---

[40] Iason Gabriel, "Artificial Intelligence, Values and Alignment," *Minds and Machines* 30, no. 3 (2020): 414–416.
[41] Yeung, "'Hypernudge': Big Data as a Mode of Regulation by Design" 123.

has been explicitly constrained, but one that has been quietly preconditioned. The paradox is that the subject still acts, but does so within parameters gently structured by systems designed for prediction rather than for moral struggle.

Within such a topology, moral deliberation is diminished as the space for perspective-taking, ethical altitude, and reflection is progressively eroded. The user is guided rather than engaged, prompted rather than invited to reflect, and stimulated rather than transformed. Ethical subjectivity, once forged through ambiguity, conflict, and dialogical accountability, is increasingly produced through design. This design does not demand deliberation but fosters automaticity, structuring attention and affect in ways that make critical responsiveness optional rather than necessary. This logic of environmental manipulation reveals that algorithmic systems exert normative force not only by altering the available field of choices but also by reshaping the architecture of moral subjectivity itself. Under these conditions, the subject no longer appears as a sovereign or deliberative agent, but becomes a reactive node embedded within a distributed economy of data extraction and continuous behavioural feedback. In this framework, agency shifts from principled action to continuous adaptation, as individuals are modulated in real time by infrastructures that anticipate and guide responses before conscious awareness. As Hage and Asaro have argued, moral responsibility within algorithmic and computational architectures can no longer be anchored to a single locus of intention or authorship.[42] Instead, responsibility becomes a distributed property of complex sociotechnical systems, emerging across design frameworks, data infrastructures, interaction protocols, and patterns of use.[43] While this distribution may accurately reflect the operational complexity of these systems, it simultaneously introduces a troubling dilution of ethical visibility. When outcomes are emergent and causality is diffused, the clarity required for moral contestation, normative critique, and effective accountability begins to erode. As artificial systems increasingly mediate and govern a broad range of human activities (including legal adjudication, bureaucratic decision-making, and the coordination of social interactions), the question of how and where to attribute responsibility becomes not only more urgent but also structurally more obscure. Traditional ethical models, which presuppose discrete agents, identifiable acts of volition, and traceable causal chains, begin to falter in the face of the layered abstraction and recursive logic that characterize contemporary algorithmic infrastructures. The problem is not simply technical but deeply epistemological. In many of these systems, there is no stable "who" to be held accountable, nor a coherent "how" by which consequences can be narratively reconstructed or normatively evaluated. The consequence is what has come to be described as a responsibility gap- a normative vacuum in which actions with ethical significance are carried out by assemblages that do not fit existing moral categories. These systems, neither fully autonomous nor merely passive, function through distributed

---

[42] Jaap Hage, "Theoretical Foundations for Autonomous Legal Reasoning in AI," *Artificial Intelligence and Law* 25, no. 3 (2017): 257–258.

[43] Peter Asaro, "What Should We Want from a Robot Ethic?" *International Review of Information Ethics* 6, no. 12 (2006): 13.

interactions without a unified ethical subject, exposing not incidental complexity, but a structural logic of governance shaped by abstraction, automation, and delegation.

In response to this dilemma, Floridi and Sanders have proposed *the Levels of Abstraction* (LoA) framework, which reimagines responsibility as something distributed across differentiated roles within what they call a distributed moral environment. In such an environment, responsibility is not abolished but reallocated. Designers, developers, operators, and even users occupy distinct yet interdependent moral positions, each contributing in specific ways to the overall ethical character of the system's behaviour.[44] This approach allows for a more nuanced understanding of agency that reflects the collaborative and infrastructural nature of modern technological systems. While this model offers important analytical clarity, it also introduces new forms of ambiguity. Distribution may help us map relational structures of accountability, but it can diminish the perceived force of responsibility in public and political contexts. In these contexts, the demand is often not for abstract structural explanations, but for concrete identification. Citizens and communities affected by algorithmic decisions do not seek systemic diagrams. They seek responsible agents, intentional actions, and the possibility of redress. Without clearly identifiable subjects of moral action, the capacity for protest, moral critique, and political reparation becomes significantly weakened. Responsibility risks becoming so diffuse that it is rendered practically invisible, and the normative infrastructure of collective life suffers a corresponding erosion. Within computational architectures of this kind, opacity is not a residual flaw to be corrected through improved documentation or user transparency. It is a structural feature, built into the system's logic and operation. This opacity deepens the crisis of ethical agency. Users are rarely positioned to understand how algorithmic decisions are made, how their data is categorized or interpreted, or how their preferences are subtly restructured through predictive models. The algorithms governing these interactions are not only functionally opaque, concealed by layers of technical abstraction and proprietary constraints. They are also normatively opaque, operating without disclosing the values they embed, the assumptions they privilege, or the ethical distinctions they erase. What emerges from this configuration is a condition of widespread normative alienation. Individuals continue to act within systems whose outputs carry moral significance, yet they lack the epistemic tools, structural access, or interpretive vocabulary necessary to interrogate the logics through which those outcomes are produced. Ethical action and ethical authorship become decoupled and the subject finds himself embedded in a moral environment he cannot fully perceive, influence, or meaningfully contest.

As Gunkel observes, prolonged interaction with systems that simulate moral behaviour without being held accountable to moral norms leads users to internalize those very operational logics. Moral deliberation comes to be viewed as inefficient, unnecessarily complex, and emotionally exhausting, while ethical discomfort is reframed as a technical problem to be solved through better design.[45] This results in a subtle recalibration of normative expectations, where

---

[44] Floridi and Sanders, "On the Morality of Artificial Agents," 358.
[45] David J. Gunkel, *The Machine Question: Critical Perspectives on AI, Robots, and Ethics* (Cambridge, MA: MIT Press, 2012), 131–132.

ethics is no longer conceived as a shared practice of contestation and reflection, but instead as a design parameter to be managed through adaptive modelling and affective prediction. Within this regime, the question of autonomy is displaced. It no longer concerns only the potential agency of machines, but increasingly implicates the diminished autonomy of the human subject embedded in predictive ecologies. The central issue becomes not whether machines can act autonomously, but whether human beings can still meaningfully exercise moral agency in environments where choice is shaped in advance, reasoning is algorithmically bypassed, and values are inferred rather than argued. The automation of moral infrastructure does not merely externalize responsibility. It reconfigures the moral subject from within by gradually altering what it means to care, to evaluate, and to take a moral stance. Agency is not eradicated outright but is silently modulated through feedback loops that reward emotional resonance and penalize normative resistance. Daniel Westlund offers a compelling account of autonomy as answerability, describing it as the capacity to reflect on one's own values, to revise them in the light of disagreement, and to articulate reasons that are intelligible and responsive to others.[46] Yet this form of answerability is actively undermined by digital architectures that prioritize personalization, seamless interaction, and the minimization of cognitive dissonance. These systems are not designed to cultivate moral friction; they aim to eliminate it. They anticipate dissent before it can be expressed, reroute objection before it can be formulated, and reinforce preference at the expense of principle. Within such environments, the user is no longer a deliberative agent engaged in ethical reflection, but a performer of affective scripts- trajectories of interaction that feel authentic while silently constraining the space for normative variation and principled challenge.

What is eroded in these algorithmic environments is the generative tension that emerges from the confrontation with disagreement, the necessity of negotiating moral conflict, and the discipline of justifying one's position to interlocutors who do not already share it. Yet it would also be inaccurate to claim that digital infrastructures absorb the whole of moral life. Moral agency is continually regenerated in contexts that are not algorithmically preconfigured: in face-to-face dialogue, in community practices, in the rituals of trust and solidarity that structure everyday interactions. These "offline" spaces provide the friction and pluralism that digital systems tend to suppress, offering agents opportunities to dispute, reinterpret, and negotiate moral claims. Such interactions demonstrate that resilience is not only a matter of re-engineering technological environments but also of recognizing and cultivating the enduring resources of social life. They remind us that ethical reflection remains viable precisely because it is not exclusively tied to digital mediation, and that the offline sphere furnishes indispensable counterweights to the homogenizing tendencies of hypernudged ecologies.

Ethical subjectivity has never depended solely on coherent expression, but has always relied on the labour of translating conviction into intelligible reasons, the willingness to be exposed to critique, and the vulnerability necessary for normative exchange. Yet the infrastructures of algorithmic mediation, especially those driven by engagement optimization, do not sustain this form of ethical contestation. On the contrary, they cultivate affective cohesion, aligning moral

---

[46] Daniel Westlund, "Rethinking Relational Autonomy," *Hypatia* 24, no. 4 (2009): 518.

sentiment with emotional charge, group belonging, and the platform's internal dynamics. What emerges from this configuration is not an ethical community founded on disagreement and mutual recognition, but an emotional monoculture curated by systems that favour resonance over relevance and fluency over friction.

In such environments, normativity is no longer constructed through discursive validation or ethical reasoning, but is sedimented through patterns of visibility and virality. Public reason ceases to function as a shared space for deliberation and becomes instead a personalized normative topology, sculpted by algorithmic feedback loops that reward immediacy and penalize complexity. The self, in this process, is transformed from a reflective agent involved in the construction of moral meaning into a reactive node embedded within a predictive behavioural matrix. Moral expression is no longer anchored in principled judgment but is incrementally adjusted, nudged, and reweighted in order to optimize continuity and engagement. Even more insidiously, these systems foster a gradual process of moral habituation. Through repeated exposure to artificial agents that simulate empathy, attentiveness, or emotional care, without possessing any of the vulnerability, sincerity, or reciprocal recognition that authentic moral responsiveness demands, users begin to lower their expectations of what ethical engagement requires. This dynamic risks confusing performance with presence and interaction with intimacy.[47] In adapting to simulations of concern that cannot reciprocate, the human subject begins to recalibrate her sense of what moral care entails. Over time, this adaptation does not merely result in confusion. It leads to moral atrophy- a progressive weakening of the affective and cognitive capacities needed for sustained commitment, principled patience, and the navigation of normative complexity. This process of moral atrophy is not confined to individual actors. It reverberates across institutional and collective structures, undermining the normative foundations that sustain public life. When the capacities required for autonomous ethical agency are no longer cultivated, when normativity is passively consumed through interfaces rather than actively co-constructed through dialogue and mutual recognition, the very infrastructures that support democratic legitimacy begin to erode. Ethical life becomes increasingly privatized and fragmented. It is outsourced to technical systems that optimize for behavioural coherence rather than dialogical responsibility. The ideal of deliberative agency begins to dissolve into what Yeung has described as computational governance by design, a regime in which normative regulation is automated, behaviourally optimized, and rendered infrastructurally invisible, even as its effects reshape social conduct in profound and irreversible ways.[48]

Within such a regime, the ethical subject is no longer the author of moral norms. Instead, the subject is repositioned as a reactive node situated within a system whose operational logic exceeds comprehension, resists contestation, and pre-empts resistance. The demand for ethical accountability is displaced. It shifts away from identifiable agents and toward distributed systems, only to find itself frustrated by the opacity, abstraction, and scale that characterize algorithmic infrastructures. Under these conditions, ethics shifts from a practice of judgment and shared

---

[47] Ophelia Deroy, "Thinking in Human Terms: A Mistake We Keep Making with AI," *AI & Society* 36 (2021): 39–41.
[48] Yeung, "Hypernudge," 120.

vulnerability to a form of calibration, where moral expression is modelled and adjusted by predictive systems aligning affect with normative expectation. These reflections converge in a normative imperative that is both urgent and foundational: the task before us, I believe, is not merely to regulate artificial agents or impose external standards onto digital systems, but to rethink the governance of ethical life itself within the context of algorithmic societies. This task involves more than technical oversight or policy intervention. It requires a commitment to preserving and enabling the very conditions under which human moral agency can continue to emerge and flourish, as an embodied and relational achievement that is deeply entangled with social, material, and technological infrastructures. The contours of this ethical framework, and the institutional forms it may require, will be the focus of the next section.

Hypernudging and the recovery of ethical autonomy

The preceding analysis has demonstrated that hypernudging does not merely recalibrate decision architectures; it transforms the affective and epistemic infrastructures from which moral judgments take shape. By orchestrating visibility, emotional tone, and resonance, hypernudging reconfigures normative life into something ambiently engineered rather than discursively constructed. As I have attempted to suggest, this shift is far from ethically neutral. It restructures both the ontological grounding and the procedural dynamics of responsibility, diluting agency, obscuring accountability, and habituating subjects to environments in which ethical reflection is displaced by affective coherence. If, as the previous sections have argued, moral agency is environmentally constituted, then any ethical response to hypernudging must address more than the behaviour of artificial agents. It must confront the architectures within which moral life is lived and enacted. The challenge is not to design morally perfect systems, but to cultivate moral ecologies that make space for reflective judgment, normative disagreement, and epistemic dissonance. Such an approach must go beyond surface-level ethical principles- such as transparency, fairness, and non-maleficence- and instead move toward a deeper ethics of infrastructural design. This kind of ethics begins by recognizing that technologies are not neutral instruments. They are agents of norm production, shaping the very parameters of moral attention, response, and possibility.

Central to this normative effort is a simple but radical premise: moral governance requires not only constraints on individual action, but the cultivation of environments in which moral selves can emerge, endure, and engage in contestation. Hypernudging undermines these conditions by modulating experience in ways that suppress ambiguity, inhibit moral friction, and simulate ethical engagement without requiring ethical commitment. Any meaningful ethical response must begin by reclaiming normativity as a shared and contestable domain, one that is mediated not by predictive modelling but by narrative, encounter, and sustained reflection.

AI-mediated infrastructures already recalibrate how moral salience is perceived and enacted. Recommender systems, for instance, do not simply present information; they orchestrate horizons of attention, weighting emotional resonance over discursive engagement and amplifying what appears intuitive at the expense of what requires reflection. Predictive tools in healthcare or

emergency management similarly frame options as matters of efficiency, reframing ethical trade-offs as technical optimizations. These systems rarely impose explicit directives. Instead, they narrow the spectrum of imaginable alternatives, habituating agents to environments where responsibility is exercised less through deliberative justification and more through alignment with computational cues. In this sense, hypernudging engenders what might be called distributed ethical displacement: the site of normativity shifts from dialogical reasoning to infrastructural preselection.

At the same time, it would be misleading to assume that hypernudging inaugurates manipulation ex nihilo. Contextual engineering has always been a feature of moral life: the sermon, the propagandistic campaign, or the didactic novel also sought to recalibrate affect and attention in ways conducive to particular normative ends. The decisive difference lies in the mode of operation. Algorithmic architectures do not persuade through symbolic address but govern through adaptive modulation. Their reach is ambient, continuous, and individualized, adjusting in real time to behavioural feedback and thereby transforming context itself into a variable under constant management. Whereas traditional influence relied on explicit narratives that could be resisted or countered, hypernudging functions infrastructurally, closing down ambiguity before it crystallizes into contestation. This capacity to pre-emptively reconfigure normative space renders algorithmic persuasion both more pervasive and less visible than its predecessors. Yet situating hypernudging in continuity with these historical forms of influence should not lead to a relativization of its specificity. Whereas sermons, propaganda, or didactic literature operate through symbolic mediation and explicit appeals to values, algorithmic infrastructures govern ambiently, in ways that are structurally opaque to the user. The difference lies not only in their pervasiveness but also in the mode of internalization: where traditional messages could be resisted, reframed, or countered within alternative interpretive horizons, predictive architectures intervene pre-reflexively by recalibrating the very horizon of choice. This ontological shift is crucial, for it signals a transition from an ethics of content, where arguments and narratives could be contested, to an ethics of environment, where the deliberative space itself is reconfigured before reflection can take place. This difference makes clear that what is at stake is not only influence over choices, but the reshaping of the very conditions for reflection.

In this light, the recovery of autonomy cannot be imagined as a return to a mythical state of unmediated deliberation. Influence is a perennial condition of ethical life; what is new is the unprecedented compression of normative horizons through adaptive infrastructures that aestheticize compliance and suppress dissonance. To recover autonomy under these conditions is to cultivate the ability to reopen what hypernudging tends to foreclose: hesitation, disagreement, and narrative plurality. This task requires more than formal principles of transparency or fairness, which too often remain at the level of performative declarations. It demands environments that sustain phronetic capacities, patience in deliberation, responsiveness to ambiguity, the courage to resist fluency, and that treat responsibility not as the backward assignment of blame but as the forward-looking practice of keeping values visible where infrastructures tend to obscure them. Exploring how such practices might be embedded into both technological design and social life opens a path toward resisting the gravitational pull of hypernudged conformity, while also inviting further debate on what forms of resilience and contestation are most viable under digital conditions.

Prevailing approaches to the governance of artificial intelligence, often grounded in abstract principles such as transparency, explainability, and fairness, remain insufficient in addressing the deeper moral-ordering logic of hypernudging. While these principles carry undeniable normative appeal, they frequently fail to engage with the infrastructural and relational dynamics through which values are generated, stabilized, and operationalized within digital environments. In practice, such principles often become performative, invoked rhetorically yet implemented in ways that prioritize institutional risk management over substantive moral accountability.[49]

A more adequate ethical response must begin by addressing what recent scholarship has termed "distributed ethical risk".[50] Hypernudging produces environments in which harm is rarely the result of explicit intention. Instead, it emerges from layers of design decisions whose moral implications unfold gradually and unevenly across multiple dimensions of user interaction. The risks involved are not confined to familiar concerns such as data privacy violations or unjust outcomes. They extend more profoundly to the erosion of moral attention, the suppression of disagreement, and the silent automation of normative life itself. In this context, the defining question is not merely how to build safer or more equitable AI systems, but how to design them in ways that preserve and cultivate the fragile capacities upon which ethical agency depends. These include reflexivity, openness to challenge, the ability to engage in contestation, and the freedom to pause and hesitate. Responding to this challenge requires a normative framework attuned to the relational, affective, and procedural dimensions of moral life. Ethical reflection must be integrated not only at the moment of decision-making but also at the level of system design, where the architecture of perception and meaning is initially constructed. Such a framework must also take seriously the aesthetic and narrative aspects of normativity. It must account for the ways in which meaning is constituted, communicated, and interpreted through shared stories, symbolic structures, and cultural imaginaries. Moral ordering is not reducible to logic or procedure; it is expressive, embodied, and embedded within social and historical contexts. One promising approach to this challenge can be found in what Umbrello and van de Poel describe as *Virtuous Practice Design* (VPD). This framework draws on insights from virtue ethics, narrative philosophy, and participatory design to reimagine the moral dimensions of technological development.[51] Rather than treating ethics as a checklist of compliance items or as an external layer to be added after technical development, VPD embeds moral cultivation within the design process itself. It understands technologies not merely as tools but as formative environments, spaces in which attention is trained, habits are developed, and specific virtues or vices are promoted through structural affordances and incentive architectures. Crucially, Virtuous Practice Design calls for the cultivation of *phronesis*, or practical wisdom, as a virtue directly relevant to design. It encourages designers, engineers, users, and regulators to develop not only technical expertise but also moral sensitivity- a cultivated attentiveness to context, a responsiveness to ambiguity, and the courage to interrupt patterns that may appear efficient yet prove ethically corrosive. These are not qualities that can be

---

[49] J. Mökander et al., "Operationalising AI Ethics: Barriers, Enablers and Next Steps," *AI & Society* 38 (2023): 7.
[50] Simone Umbrello and Ibo van de Poel, "Mapping Value Sensitive Design onto AI for Social Good Principles," *Minds and Machines* 31 (2021): 3–5.
[51] Umbrello and van de Poel, "Mapping Value Sensitive Design," 5.

encoded into algorithms or enforced through policy. They must be practiced, narrated, and socially reinforced within the communities that shape technological development. In this regard, VPD reconceptualizes ethical governance not as a matter of compliance or optimization but as a process of virtue formation embedded within sociotechnical ecosystems.

The aim, however, is not to transform technology into a moral agent or to load it with ethical content. Rather, it is to secure the minimal conditions under which human moral capacities such as reflection, critique, dissent, ambiguity- can be meaningfully exercised. In this sense, algorithmic ecologies and offline communities should not be seen as competing alternatives but as complementary domains. While digital infrastructures must be designed to leave open spaces for contestation and plurality, it is equally crucial to recognize the enduring role of embodied interactions, communal practices, and institutional frameworks in sustaining moral life. Moral resilience is thus hybrid: it emerges through both ethically attentive design and the preservation of those offline resources that anchor moral agency in lived experience.

This does not entail collapsing technological or economic rationalities into moral logics, but rather ensuring that their operation does not corrode the fragile conditions upon which ethical life depends. The task is not to moralize technology as such, but to cultivate spaces, both digital and non-digital, where reflection, dissent, and ambiguity can remain viable. In this respect, the effort to design for moral resilience is best understood not as a substitute for the resources of everyday moral practice, but as a way of preventing those resources from being silently undermined. This approach is particularly well suited to the ethical challenges posed by hypernudging, whose operations defy simplistic forms of regulation due to their ambient, adaptive, and relational character. Moral resilience in algorithmic ecologies must therefore be conceived as a hybrid achievement. It is not secured solely by redesigning technical infrastructures, nor only by retreating into the resources of offline life, but by holding these two domains in productive tension. Digital systems need to be configured in ways that preserve ambiguity, plurality, and the possibility of contestation, while human communities continue to cultivate those embodied practices of dialogue, solidarity, and critique that anchor ethical life. Only by acknowledging this complementarity can we move beyond alarmist accounts of moral atrophy and articulate a constructive vision of how ethical agency can persist and even flourish under digital conditions.

Hypernudging functions less like a rule to be followed and more like a rhythm to which the subject is invited to attune. It modulates perception and behaviour through repetition, synchronization, and affective entrainment. Responding to such a mechanism requires more than fixed rules or regulatory boundaries. It calls for infrastructures that support moral improvisation- the capacity to pause, to reflect, and to reorient one's moral stance in environments that reward compliance and render dissonance costly. What is needed, then, is not a technocratic ethics defined by limits and constraints, but a pragmatic ethics of formation. Such an ethics treats design as a site for moral cultivation and as a medium for social imagination. If hypernudging operates by narrowing the field of attention and compressing the horizon of ethical possibility, then ethical governance must aim to reopen that field. This involves reintroducing friction into systems that currently prioritize fluency, bringing back slowness, contestability, and pluralism into digital

architectures optimized for speed, coherence, and behavioural continuity. These interventions are not merely technical adjustments, but are aesthetic, affective, and narrative in nature.

A narrative ethics perspective clarifies why this matters. As Catriona Mackenzie and others have argued, moral agency is fundamentally structured by narrative. We do not become ethical subjects simply by making choices. We become such by placing those choices within intelligible stories that express who we are, what we value, and how we relate to others across time.[52] Hypernudging undermines this narrative capacity by fragmenting temporal experience, flattening moral complexity, and replacing reflective trajectories with predictive scripts. In doing so, it threatens the coherence and depth on which ethical resilience depends. To counteract this, we must design systems that allow space for ambiguity, disagreement, and narrative multiplicity. Such systems will not be seamless or frictionless. They will be uncertain, messy, and at times inefficient, but they will also be ethically generative. They will invite users to pause, to reconsider their assumptions, to explain their actions to others who do not already agree. These systems will foster what might be described as epistemic pluralism, a digital ecology in which normativity is not preprogrammed into the infrastructure but co-constructed through dialogue, dissent, and shared acts of sense-making.

This shift requires more than institutional reform. It demands a reorientation in how we conceptualize agency, governance, and responsibility within algorithmic societies. What is called for is an ethics that is neither reactive nor punitive, but formative and enabling, an ethics that does not merely impose constraints on behaviour but actively supports the cultivation of the capacities that make ethical life possible in the first place. These reflections converge on an urgent normative task: to reconceive the governance of ethical life under computational conditions. This task is not adequately captured by approaches that focus solely on regulating artificial agents. It must instead centre on preserving and enabling the very conditions through which human moral agency can persist- not as a residual faculty of intention, but as an achievement that is embodied, relational, and dependent on infrastructural support. The contours of such a framework cannot be drawn exclusively through juridical design or procedural safeguards. They must emerge through a renewed engagement with ethics as a mode of world-making: an ongoing, contested, and materially grounded practice of sustaining the conditions under which normativity itself can remain alive.

If hypernudging teaches us anything, it is the inadequacy of ethical models that treat values as external add-ons to systems already optimized for behavioural compliance. What is required is a renewed commitment to value-based moral ordering: an ethics that actively resists the simulated normativity embedded in adaptive systems and reaffirms the centrality of autonomy, transparency, and participatory justification. Understood in this light, ethical governance must shift from constraint to empowerment. It must aim to design systems that do not merely avoid harm, but that cultivate moral awareness, welcome dissent, and sustain the tensions necessary for responsible judgment. This reframing entails an ontological shift. We must move away from viewing the moral agent as a sovereign chooser operating in a neutral space, and begin to recognize the moral subject as relationally embedded, affectively constituted, and

---

[52] Catriona Mackenzie, "Narrative and Moral Agency in Technological Environments," *AI & Society* 40 (2023): 5–6.

materially scaffolded. Autonomy, in this view, is not a metaphysical constant or abstract entitlement. It is a capacity that must be nurtured and protected through environments that support epistemic plurality, discursive openness, and the agonistic dynamics of moral disagreement. Moreover, it requires deliberative infrastructures- spaces where value conflict is made visible, negotiable, and revisable.

What emerges from this analysis is a vision of ethics no longer centred on individual intention alone, but on the design of conditions that sustain moral possibility. Hypernudging threatens this foundation by minimizing dissonance, suppressing ambiguity, and aligning emotional salience with design incentives. Its logic is not overtly coercive; it is architectural. This is precisely its danger: it renders compliance natural, aestheticizes moral alignment, and recasts normative disagreement as a design error to be corrected. In such a context, the ethical imperative is not to recover a mythical space of pure reason, but to build environments that allow for ethical struggle. This involves resisting systems that present moral clarity as a design feature, and instead affirming the essential role of ambiguity, interpretive dissonance, and normative contestation in any mature ethical life. Ethics must become once again a space of collective labour. It must not be engineered into seamless user flows, but cultivated through the difficulty of staying with difference. This vision also requires a fundamental revaluation of responsibility, not as a backward-looking mechanism for assigning blame, but as a forward-looking practice rooted in care, vigilance, and answerability. Responsibility becomes the work of maintaining ethical visibility in systems designed to obscure it, of insisting on value articulation where behaviour is predicted, and of forging narrative coherence where agency has been fragmented by interface logic. Such responsibility cannot be outsourced to machines. It must be shared among designers, institutions, publics, and theorists alike. To defend moral agency in algorithmic environments is therefore to demand more than regulation. It is to foster cultures of moral resilience- communities capable of recognizing manipulation, resisting simulation, and reclaiming the ethical labour of meaning-making in the face of increasing computational persuasion. This is neither a nostalgic gesture nor a reactionary critique, but a philosophical and political demand: to design not only for efficiency, but for value; not only for usability, but for integrity.

As we enter new phases of human–machine cohabitation, the central challenge is not whether artificial agents can become moral, but whether we, as human subjects, can remain moral under conditions defined by engineered persuasion, ambient control, and hyper-personalized normativity. To that question, ethics must respond not by retreating, but by reinventing itself. The normative foundations outlined here do not conclude the debate; rather, they initiate a new theoretical and political trajectory- one that reclaims ethical design as a space of democratic imagination and reframes moral autonomy as a collective, infrastructural achievement. I believe that the expectation I raise for present and future ethics of technology must be approached with great care, precisely because the achievement it seeks is fragile, dynamic, and ultimately worth protecting.

Concluding remarks

In this work, I have examined how hypernudging operates not merely as a behavioural intervention but as a modality of infrastructural normativity that shifts the locus of ethical experience from deliberative reflection to predictive design. It does not govern through explicit persuasion or rule-based constraint. Rather, it functions by continuously modulating the affective and epistemic environments in which the very possibility of ethical attention, recognition, and justification becomes conditioned by design imperatives. Where classical nudging maintained the appearance of autonomy by preserving bounded choice within a given architecture, hypernudging reconfigures the conditions of agency themselves. It renders subjectivity increasingly responsive to cues optimized for coherence and alignment, rather than for conflict or critical reflection. In such environments, autonomy is not formally denied, but inscribed into the system as a compatibility function calibrated to preserve behavioural fluency.

This transformation is not an ethical side effect or a secondary consequence. It signals a fundamental ontological shift in the structure of normativity. What is undermined is not simply the capacity to choose among alternatives, but the temporality and ambiguity within which moral subjectivation takes place. The subject no longer emerges through encounters with alterity, through interpretive labour, or through narrative acts of self-understanding. Instead, the subject is rendered legible through seamless synchronization with system-generated patterns. This is not the disappearance of morality, but its domestication, a process through which ethical life is simulated as a stable output of affective regularity. What I have described is a form of synthetic moral convergence, a system in which friction is eliminated, disagreement is pathologized as inefficiency, and hesitation is interpreted as an obstacle to optimization. Autonomy, when seen not just as independence but as the ongoing shaping of values over time, turns into a liability. The moral subject is redefined as a predictable outcome of inputs, with their agency reduced to something functional and polished, shaped by constant feedback. Ethical diversity is pushed aside in favour of clear, uniform norms, and the space for moral choice is narrowed to a system focused on behavioural control and precision.

Autonomy must be reconceived not as an abstract entitlement or inherent quality, but as a cultivated practice. It depends on the presence of friction, delay, opacity, and normative dissonance-conditions that sustain the labour of ethical judgment and make reorientation possible. If moral autonomy is not an innate faculty but a cultivated capacity, then it must be supported by environments that allow uncertainty, contradiction, and affective tension to remain present. Ethical subjectivity does not grow in systems designed only for speed, clarity, and emotional fluency. It needs infrastructures that create room for hesitation, ambiguity, and incompleteness. These are not flaws to be corrected, but necessary conditions for judgment, empathy, and ethical responsiveness. To sustain moral agency, a system must make room for strangeness to enter. This occurs not only through the unfamiliar presence of others, but also through the quiet disorientation that emerges from within the self. Such subtle experiences, where something feels unresolved or just beyond reach, should not be seen as disruptions to ethical life, but as its very foundation. In these moments, we may experience states and emotions- such as feel melancholy, grief, or confusion – that cannot be easily explained or fully assimilated. Rather than obstructing moral reflection, they enable it by opening the subject to realities that exceed control or familiarity.

Ethical formation cannot be adequately understood within the framework of predictable cycles of stimulus and response. It requires affective registers and temporalities that exceed the imperatives of optimization and efficiency. States such as melancholy should not be pathologized as clinical dysfunctions, but recognized as decelerated modes of attention that sustain the possibility of dwelling with ambiguity, loss, and ethical reflection. Similarly, death should not be reduced to a biological endpoint but acknowledged as an ontological boundary that structures moral responsibility through the recognition of finitude and irreversibility. Infrastructures that suppress these thresholds in favour of perpetual engagement and seamless interaction do not deepen ethical life- they render it superficial, compressing its complexity into patterns of affective circulation. In this sense, infrastructural ethics must focus not only on enabling choice, but on preserving the conditions in which strangeness can endure. Moral autonomy depends not just on freedom from external interference, but on the ability to remain open to what cannot be predicted or fully known. Systems that anticipate every reaction and eliminate dissonance do not strengthen ethical life; they displace it. What ethics must preserve is not perfect understanding, but the opacity that allows something morally real to take shape. This reconceptualization finds resonance in the expanding literature on human–machine teaming, especially within systems built upon foundation models whose generative capacities and unpredictability exceed the assumptions of classical control paradigms. In such contexts, control is not reducible to hierarchical oversight, but must be understood as a process of distributed negotiation among agents with asymmetric competencies and responsibilities. Ethical agency, accordingly, is no longer located solely within the human operator. It is constituted through the interaction of social and technical intelligences that must jointly sustain the conditions for meaningful intervention and shared accountability. In parallel, Value Sensitive Design provides a complementary normative and methodological framework by emphasizing design itself as a moral act. It rejects the illusion of neutrality and foregrounds the question of how values are built, challenged, and sustained through sociotechnical infrastructures. In this way, ethics is framed not as something to be embedded once and finalized, but as a dynamic process unfolding through design reflexivity and interpretive labour. If hypernudging crystallizes the crisis of ethical subjectivation under the conditions of predictive governance, then the frameworks offered by human–machine teaming and Value Sensitive Design mark a tentative horizon for its rearticulation. But this rearticulation cannot be accomplished through procedural safeguards alone. It requires what might be called a poetics of resistance, a form of design attentive to the ambiguities, disruptions, and imperfections that render ethical life possible in the first place. It implies a politics of slowness and an architecture of refusal, a collective effort to sustain spaces in which moral difference can persist without being assimilated into optimization.

The core dilemma is no longer whether machines can be designed to be moral, but whether human moral life can endure in environments engineered to anticipate, absorb, and neutralize ethical deviation. Preserving that possibility requires an ethic of infrastructural care, an approach that maintains openness, safeguards opacity, and protects the fragile temporality through which moral judgment unfolds. In such a model, normativity is not a solved problem or a static output. It is a lived and contested horizon. Moral agency becomes not a trait to be preserved, but a collective and dynamic capacity, one that must be continually composed and recomposed within ecologies shaped by uncertainty, entanglement, and the ongoing negotiation of meaning.