

Concept Paper

Anthropological Singularity and Algorithmic Stratification: A Medical-Sociological Reframing of the Bioethical Principles Governing the Use of Artificial Intelligence in Healthcare

Antonio Sandu ^{1,*}  and Petronela Polixenia Nistor ² 

¹ Faculty of Law and Administration, “Ștefan cel Mare” University of Suceava, 720229 Suceava, Romania

² Faculty of Orthodox Theology, “Alexandru Ioan Cuza” University of Iași, 700506 Iași, Romania; polixenianistor@yahoo.com

* Correspondence: stefan.sandu@fdsa.usv.ro

Abstract

The integration of artificial intelligence (AI) into medical practice is a paradigmatic transformation that compels a reconsideration of the very conceptual structure of bioethics. We argue that this transformation is the local symptom of a wider mutation we designate the anthropological singularity: the threshold at which humanity withdraws from natural evolution and enters an artificial, self-directed one, along three converging vectors—the virtualisation of social space, the cognitive and relational reconfiguration of the human agent through interaction with AI, and the modification of the biological substratum through synthetic biology and gene editing. This approach must be seen as a conceptual study, and not as a systematic literature review, since the study does not aggregate empirical findings but constructs them by defending a theoretical framework and drawing on medical sociology, medical anthropology and biolaw, in order to examine the recurrent tensions between transparency vs. performance, **justice** vs. **efficacy**, precaution vs. innovation. We argue that algorithmic bias is part of socio-technical ecosystems, being a systemic property of these. We also argue that responsibility must migrate towards distributed frameworks between human and non-human agents, and that the digital revolution could produce an algorithmic stratification of populations, now prolonged through gene editing into the very biology of future generations. An *augmented bioethics* is required—not mere delegation of moral judgement to algorithms, but a reflexive framework that revises the ontological presuppositions of principlism, sustained by multi-stakeholder governance and deliberate public policies.

Keywords: anthropological singularity; artificial intelligence in medicine; bioethics; algorithmic bias; distributed responsibility; distributive justice; medical sociology; algorithmic stratification; augmented bioethics; medical anthropology



Academic Editor: Joaquim J. F. Soares

Received: 20 June 2026

Revised: 4 September 2026

Accepted: 7 September 2026

Published: 9 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Ethical deliberation has been, since the dawn of political and juridical thought, an essential component of medical culture, and the obligation of the physician to exercise every diligence in the patient’s favour—within the limits of professional competence and of available resources [1]. The contemporary integration of artificial intelligence (AI) into medical practice reactivates this ancient obligation under conditions that the classical tradition could not have anticipated: the medical act is no longer mediated solely by the clinical judgement of a human agent, but increasingly by socio-technical systems

whose internal operations remain, at least in part, opaque even to those who design them. It is the working hypothesis of the present study that this displacement is not an incremental technological innovation but a paradigmatic transformation, one that compels a fundamental reconsideration of the conceptual frame of bioethics itself. We want to emphasise here the birth of a new paradigm: the appearance of AI entities that are non-human constitutes an ontological mutation: the space of social interactions is radically reconfigured—no longer confined to exchanges between strictly human individuals, but extended to encompass interactions between human individuals and non-human entities endowed with artificial intelligence—while the very medium of interaction is no longer determined by immediate spatial proximity, but by a form of *digital vicinity*. Foundational categories such as space, time, and causality undergo ontological mutation in the process of their recovery and reappropriation within the horizon of AI. This, we argue, marks a moment of **anthropological singularity**—one that manifests distinctly across the various domains of human experience—in the case of the present study, within the sphere of bioethics [2].

The four principles of biomedical ethics formulated by Tom Beauchamp and James Childress—respect for autonomy, beneficence, non-maleficence and the principle of justice [3]—continue to furnish the normative apparatus within which the moral acceptability of medical action is assessed. These principles, however, do not operate in isolation in clinical practice; they form an interdependent system of values that must be balanced according to the concrete decision at hand. Artificial intelligence may extend access to systems of care while, at the same time, introducing new and systematic risks for non-maleficence, so that what is gained along one axis of the principlist scheme is liable to be lost along another. We consider, accordingly, that a unified ethical framework for algorithmic medicine cannot be constructed through the mechanical application of traditional principles, but requires a conceptual reconfiguration capable of recognising the disparities inherent in the technology and proposing mechanisms for arbitration between competing ethical values and principles.

The perspective adopted in this paper is that of medical sociology rather than that of computer science. This choice is deliberate: the harms and the benefits of medical AI are not distributed by the architecture of the algorithms alone, but by the social, economic, and cultural structures within which those algorithms are trained, deployed and trusted. A medical-sociological reading therefore treats algorithmic bias, the diffusion of responsibility and the stratification of access not as isolated technical defects but as properties of the wider ecosystem—properties that reproduce, and frequently amplify, pre-existing patterns of advantage and disadvantage. The argument therefore situates itself within a wider literature on the digital transformation of expert systems, on professional autonomy under conditions of automation, and on the institutionalised inequalities that accompany the diffusion of new technologies, and contributes to that literature a specifically medical example.

The central thesis of this article is that these phenomena are the local symptoms—in public health—of a wider mutation, which we propose to designate the **anthropological singularity**. We construct the concept by deliberate analogy with the technological singularity, the latter understood, after Bostrom [4] and in the recent reformulation of Kurzweil [5], as the point at which machine intelligence outstrips its human counterpart and either exits human control or merges with it. The technological singularity concerns the intelligence; the anthropological singularity concerns the human itself. By this we mean the threshold at which humanity withdraws from the line of natural, Darwinian evolution and enters an artificial, self-directed evolution—a process that O’Lemmon, in an explicitly anthropological reading, describes not as a punctual event but as a continuous transition [6], and that others have characterised as the gradual escape of intelligence from the constraints

of biology [7]. The decisive feature of the anthropological singularity is therefore not that the machine surpasses us, but that the laws of social-historical development—and, with them, the very principles that structure our bioethics—require profound revision, on pain of becoming inoperative as foundations of technological governance.

Three converging vectors, in our analysis, drive this withdrawal from natural evolution. The first is the virtualisation of social space: the transfer of human interaction into a virtual context that is no longer governed by effective spatial distances but by symbolic and communicational ones, so that proximity ceases to be a matter of metres and becomes a matter of meaning [8,9]. The second is the cognitive and relational reconfiguration of the human agent through habitual interaction with AI—a reshaping of clinical reasoning, expertise and the therapeutic relation that no longer leaves the agent of the four principles intact. The third, and the most literal, is the direct modification of the biological substratum through synthetic biology and gene editing, which converts the human species from the product of natural selection into the object of its own design [10,11]. Our claim is that the bioethical, ethical and philosophy-of-medicine arguments developed in the sections that follow are partial arguments for this single thesis: each structural tension we analyse—between transparency and performance, between justice and efficiency, between precaution and innovation—is a regional manifestation of the same ontological mutation in what it means to be human. These three vectors organise the Discussion (Section 4); the partial arguments, by contrast, structure the demonstration as a whole, and the two schemes intersect rather than coincide—a single vector may surface in more than one partial argument, and several partial arguments bear upon no vector but upon the principlist scheme itself.

A word is required on the genre of the present study, since the choice is not incidental to its argument. This is a concept paper: no new empirical data are presented, and existing concepts and knowledge are mobilised in order to propose a different way of construing them. Its contribution is architectural rather than evidentiary, and it is to be assessed accordingly—by the coherence of the distinctions it draws, by their power to discriminate between phenomena that were previously undifferentiated, and by the research questions they open—rather than by the reproducibility of a retrieval procedure. In the typology of conceptual designs proposed by Jaakkola [12], the study combines theory adaptation, in that it reconstructs the principlist framework through a lens external to it, with typology, in that its central claims are advanced through the systematic contrast of Tables 1 and 2.

Table 1. Technological singularity and anthropological singularity: a conceptual contrast.

Criterion of Comparison	Technological Singularity	Anthropological Singularity
Locus of transformation	The machine: the rise of a superhuman artificial intelligence.	The human: the ontology of the human person and of the species.
Threshold	Machine intelligence outstrips and escapes human control (Bostrom; Kurzweil).	The laws of social-historical development and the bioethical principles become inoperative as given.
Driving process	Recursive self-improvement of artificial intelligence.	Virtualisation of social space; cognitive-relational reconfiguration through AI; gene editing and synthetic biology.
Effect on human control	Loss of control over an external intelligence.	Loss of the stable referent—‘the human’—that control was meant to protect.
Effect on bioethics	Raises the question of the moral status of machines.	Renders revisable the ontology of the unitary moral agent presupposed by principlism.
Temporal form	Often imagined as a punctual, future event.	A continuous transition already under way (O’Lemmon).

Table 2. Documented instances and the arguments they anchor.

Instance	Vector	Principle at Stake	Partial Argument It Anchors	Section
Epic Sepsis Model, external validation	Cognitive	Non-maleficence; precaution	Dissolution of the unitary responsible agent	Section 3.1
Pulse oximetry, differential accuracy by skin tone	Instrumental	Justice; non-maleficence	Stratification inscribed upstream of the algorithm	Section 3.2
Race coefficient in eGFR, withdrawn	Classificatory	Justice; autonomy	Structural inscription is revisable	Section 3.2
Dermatological datasets and model sensitivity	Cognitive	Justice	Algorithmic stratification of diagnostic access	Section 3.2
Commercial risk algorithm, expenditure as proxy	Cognitive	Justice	Historical bias converted into protocol	Section 3.2
Epic Sepsis Model v2, institutional variability	Cognitive	Justice; beneficence	Ethical adequacy is locally constituted	Section 3.3
Endoscopic de-skilling, multicentre	Cognitive	Beneficence	Directed evolution of clinical cognition	Section 4.1
Germline editing, 2018	Biological	Non-maleficence; precaution	Governance constructed retrospectively	Section 4.4
Psychotic phenomena associated with conversational agents	Virtualisation	Non-maleficence	Pathologies of the next clinical epoch	Section 4.5

This approach, rather than the systematic review, answers the present state of the field as a claim we make deliberately and are prepared to defend, and not merely as a synthesis of hypotheses verified by other authors. Systematic aggregation presupposes a stabilised object, a settled question and a body of findings rendered comparable by shared descriptors; it consolidates a domain whose terms have already been agreed. None of these conditions obtains here. The broad diffusion of large language models is a matter of some three years, the incorporation of such systems into clinical workflows is more recent still, and the awareness of the general public that medical judgement may be algorithmically mediated is more recent again—a public awareness which is itself, as we argue in Sections 4.2 and 4.5, a constituent of the phenomenon and not merely a reaction to it. To systematise prematurely, under these conditions, is not a neutral exercise: it confers upon a domain in formation the appearance of a consolidation it has not attained, and it fixes as descriptors the very categories whose adequacy is precisely what is in question.

What the field requires at this juncture is therefore of another order: the clarification of what artificial intelligence alters, not in medicine alone, but in the public's relation to medicine and in the anthropology of health that underwrites it. This is the office of what Blumer called the *sensitising concept*—a concept which, in default of fixed benchmarks in the observed material, supplies clues and directions along which to look, and from which one moves outwards towards the distinctiveness of the instance rather than subsuming the instance beneath an established class [13]. The anthropological singularity, the algorithmic stratification and the augmented bioethics proposed here are offered in exactly this capacity. Conceptual and empirical work stand in a relation of sequence, not of rank: the systematic reviews that this domain will properly require are not thereby deferred but made possible, since they cannot be conducted until it is settled what is to be counted, under what description, and as an instance of what. Section 4.7 accordingly sets out the research

questions that follow from the framework, in the conviction that the value of a conceptual proposal is measured by the empirical work it renders formulable.

Table 1 sets out the conceptual contrast on which the argument rests. The remainder of the article is organised as follows. After a statement of the conceptual framework and of the analytical strategy (Section 2), we set out the three structural tensions on which the analysis turns (Section 3): the reconfiguration of precaution and distributed responsibility, the emergence of algorithmic stratification and systemic bias, and the structural conflicts between the bioethical principles. Section 4 discusses these in the light of the transformation of the therapeutic relation, of gene editing as directed evolution, and of the emergent pathologies of algorithmic life, before setting out the controversies that remain open and the research agenda that follows (Section 4.7); Section 5 draws the conclusions and sketches the contours of what we term an **augmented bioethics**.

2. Conceptual Framework and Analytical Strategy

The present work is theory-driven and reports no empirical experiments, human or animal data. It proceeds by conceptual distinction and classification—a procedure congenial to bioethical reasoning—articulated around the principlist framework of Beauchamp and Childress [3] and complemented by the philosophical literature on autonomy, deontology and consequentialism that underwrites the contemporary doctrine of informed consent [1,14]. The literature is engaged here as conceptual material to be reconstructed, not as a corpus to be aggregated: sources supply the distinctions the argument works with, the empirical anchorages against which it is tested, and the contrary positions against which it must hold. What the study identifies are the recurrent tensions that arise when each principle is confronted with the realities of algorithmic decision-making.

2.1. The Medical-Sociological, Anthropological and Biolaw Framework

The perspective announced in the Introduction requires, if it is to operate as an analytical instrument and not merely as a declaration of disciplinary allegiance, that the traditions upon which it draws be named and that the office of each within the demonstration be specified. Four strands are mobilised here, and we indicate for each the partial argument it is called upon to sustain. We do not claim to survey these literatures; we claim only to have borrowed from them the determinate concepts on which the argument turns.

From **medical sociology** we retain six. The thesis of professional dominance, in the sense given to it by Freidson [15], furnishes the measure against which the displacement of clinical authority towards the algorithm is assessed in Section 4.1: what is at stake in de-skilling is not the competence of the individual practitioner but the jurisdictional autonomy of a profession. The concept of medicalisation, as elaborated by Conrad [16], and its technoscientific successor, the biomedicalisation described by Clarke and colleagues [17], supply the frame within which the passage from the medicalisation of the individual life to that of the species is read in Section 4.3, while Armstrong's account of surveillance medicine [18] names the mechanism by which risk itself becomes the object of clinical attention. The sociology of standardisation developed by Timmermans and Berg [19] establishes that a protocol is never a neutral instrument of coordination but a redistribution of discretion, and it is this insight that governs our treatment of the explicit hierarchisation of values in Section 3.3. Nettleton's analysis of the informationalisation of medical knowledge [20] and Lupton's critical account of digital health [21] describe the terrain on which that redistribution now takes place. Finally, the theory of fundamental causes advanced by Link and Phelan [22] supplies the mechanism our second partial argument requires: where a resource confers advantage across whatever intervening risk factors happen to prevail, the arrival of a new technology reproduces the existing gradient rather

than dissolving it. This is precisely the claim we defend under the name of algorithmic stratification, and it is what distinguishes that claim from the weaker proposition that access is at present unevenly distributed and will even out in time. The corresponding remedy, so far as medical formation is concerned, is what Metzl and Hansen have termed structural competency [23]—the capacity to recognise, and to be trained to recognise, the structural determination of what presents itself as a clinical fact.

From **science and technology studies** we retain four. The actor-network account of agency, in the formulation of Latour [24], supplies the ontological warrant for the notion of distributed responsibility defended in Section 3.1: agency, once dispersed across human and non-human actants, cannot afterwards be reassembled into a single author for the purposes of attribution. The analysis of classification and of its infrastructures by Bowker and Star [25] establishes that every operative taxonomy encodes a politics which becomes invisible in the exact measure of its success—the condition of algorithmic categorisation examined in Sections 3.2 and 3.3. The thesis, formulated by Winner [26] and extended to algorithmic systems by Benjamin [27], that artefacts embody political arrangements, furnishes the secular counterpart of the claim we advance in Section 3.3, that no technical instrument is in fact morally neutral; and the mechanism by which informatic interventions in health disproportionately benefit the already advantaged has been given a name and an evidentiary basis by Veinot, Mitchell and Ancker under the heading of intervention-generated inequality [28], with Eubanks documenting its operation across the administration of public and social services more widely [29]. The asymmetric extraction of health data examined in Section 3.2 is read, correspondingly, through the concept of data colonialism proposed by Couldry and Mejias [30], which names as an appropriation of life itself what the political economy of platforms had described only as an accumulation of information.

From **medical anthropology**, we retain four, and this strand is not ancillary: it is what licences the adjective in the concept we propose. The distinction between disease and illness, and the theory of explanatory models formulated by Kleinman [31], governs our treatment of synthetic empathy in Section 4.2, since what an algorithm can approximate is a nosological category and not the meaning a patient assigns to their own suffering. The notion of local biologies, in the sense given to it by Lock and Nguyen [32], bears directly upon Section 4.4: if biological life is itself constituted differently across material and social environments, then neither the algorithm trained upon one population nor the germline intervention designed for it can be transposed to another as though the substratum were universal. The concept of structural violence, as developed by Farmer [33], names the mechanism by which institutional arrangements inflict harm without an identifiable agent—the very structure of the responsibility gap examined in Section 3.1—while Rose's account of a molecular biopolitics [34] describes the correlative formation of identities organised around susceptibility rather than around diagnosis. To these we add the argument, recently restated by Costa and Serra [35], that healthcare technologies and clinical decisions are to be interpreted not through ethical and juridical principles alone but through the cultural contexts, lived experiences and social meanings within which illness is constituted; an argument whose pertinence to our thesis is that it establishes, on independent grounds, why an **augmented bioethics** cannot be exhausted by technical robustness or normative regulation.

From **biolaw** we retain, finally, three orientations. The analysis of technoscience proposed by Hottois [36] furnishes the vocabulary in which the human is understood as having become an object of technical operation rather than solely a subject of contemplation—the philosophical premise of the anthropological singularity itself. The notion of technological management, in the sense given to it by Brownsword [37], identifies the displacement that our third partial argument presupposes: where conduct is rendered impossible by

design rather than prohibited by norm, regulation ceases to address a moral agent capable of compliance or breach, and the juridical subject is dissolved along with the moral one. And the doctrine of human dignity as a constraint upon biomedical intervention, articulated by Andorno [38] in his reading of the international instruments, together with the prohibition of germline modification laid down in Article 13 of the Oviedo Convention [39], supplies the positive-law horizon against which the discussion of gene editing in Section 4.4 is conducted.

These four strands converge upon a single methodological commitment, which we state explicitly because the whole demonstration rests upon it: that the harms and the benefits of a medical technology are not properties of its architecture but of the socio-technical ecosystem within which it is trained, deployed, and trusted. It follows that the reconfiguration of bioethics we propose cannot be derived from an analysis of algorithms, and must be derived instead from an analysis of the social, cultural, and juridical forms that algorithms take up and transform.

2.2. *Criteria of Pertinence and Documentation of the Evidence Base*

Conceptual work is exposed to the charge of selective citation, and we therefore make explicit the criteria of pertinence that have governed the engagement of sources. First, thesis-relevance: a source is drawn upon where it bears either on one of the three vectors of the anthropological singularity—the virtualisation of social space, the cognitive-relational reconfiguration of the agent, or the modification of the biological substratum—or on one of the determinate empirical claims (algorithmic bias in dermatology and in cardiology; clinical de-skilling; synthetic empathy) on which the conceptual argument relies. Second, authority and verifiability: peer-reviewed publications, monographs issued by established academic presses, and primary normative instruments are preferred, in that order; sources outside these categories are admitted only to attest a recent phenomenon and are marked as such in the text. Third, weighted recency: for claims concerning the performance of contemporary AI systems, we rely on work published between 2022 and 2026, as the field is evolving on a monthly cadence; for conceptual foundations—singularity, autonomy, personhood, directed evolution—the pertinent sources are classical and are engaged irrespective of date. Fourth, dialectical balance: for each strong thesis advanced, a source articulating the contrary or a qualifying position is sought and addressed, so that the argument is tested rather than merely illustrated; Sections 3.3 and 4.4 record the concessions this has required.

The empirical anchorages of the argument were located through targeted searches of Web of Science, Scopus and PubMed, supplemented by Google Scholar for recent and pre-print material, using as seed terms algorithmic bias and healthcare, dermatology AI and skin tone, gender bias and clinical AI, de-skilling and clinician, artificial empathy and chatbot, and germline editing and enhancement, in each case with backward and forward citation tracking from the works retained. We report no counts of records screened or excluded, and no flow diagram, because none would be meaningful: the searches were not designed to be exhaustive, and the material they returned does not function evidentially in the sense that would make its completeness relevant. Each empirical anchorage carries the weight of an instance, not of a summary estimate, and the conclusions it supports are correspondingly conditional—the finding constrains the conceptual claim without being aggregated with others to establish it. Where a single study bears an argumentative load, its singularity is stated in the text, and the caution proper to generalising from it is recorded, as in the treatment of Obermeyer and colleagues [40] in Section 3.2.

Three analytic axes structure the argument. The first concerns the equilibrium between beneficence and non-maleficence, reframed through an adapted principle of precaution and

through models of responsibility for algorithmic harm. The second concerns distributive justice and the equitable access to AI-based technologies, examined in relation to algorithmic bias as a systemic phenomenon. The third concerns the structural conflicts between the principles themselves—transparency against performance, justice against efficiency, precaution against innovation—and the consequent need for an explicit hierarchisation of ethical values within automated decision systems. Where regulatory instruments are invoked, reference is made to the European Union’s Artificial Intelligence Act, to the adaptive, safety-oriented approach of the United States Food and Drug Administration (FDA) towards software as a medical device, and to the universal yet contextually flexible ethical guidance promoted by the World Health Organisation (WHO) [39,41–43].

The scope of the method is thus determinate. The study does not measure, and its conclusions are arguments rather than estimates; the analyses set out in Section 3 are theses defended, not data obtained. The criteria stated above are intended to discipline, though they cannot eliminate, the interpretive latitude that conceptual work allows. The conceptual content and its argumentative structure are those of the authors; a declaration concerning the use of generative AI for language translation from the original language of the authors and editing and structural reformulation is provided in the Acknowledgements. Because the study is entirely theoretical, no new data were created or analysed, and questions of clinical-trial registration, human-subjects approval, and dataset deposition do not arise, as recorded in the back matter.

3. Conceptual Analysis: Three Structural Tensions

3.1. *Precaution, Distributed Responsibility, and the Dissolution of the Unitary Agent*

The equilibrium between beneficence and non-maleficence in the use of artificial intelligence requires the adoption of a principle of precaution adapted to the present technological context. Such a principle does not entail the renunciation of innovation; it entails, rather, the implementation of mechanisms of evaluation, monitoring and intervention that permit the benefits of AI to be realised while the risks of its use are minimised—among these risks, the abdication of the physician’s own professional judgement in favour of a presumed technological superiority, the dehumanisation of the medical act under conditions of hyper-technologisation, and the imbalance in access between health systems of unequal financial strength. Responsibility in the development of medical AI systems therefore demands rigorous validation across diverse populations before large-scale clinical implementation, together with the inclusion of the relevant stakeholders—physicians, patients and ethicists alike—across every phase of the technology’s life cycle, and a transparency that takes explicit account of the limitations and of the uncertainty attaching to algorithmic predictions.

A regulatory framework is required that imposes clear, explicit, and efficacious standards of safety and that secures both efficacy and equity of access to AI-based medical devices. The clear attribution of responsibility must be evident between the responsibility borne by developers, medical institutions, and clinicians; and this apportionment of responsibility becomes essential to the human factor and to its accountability when interacting with expert systems endowed with artificial intelligence. At the European level the question of liability for harm caused by automated systems is being debated—whether such harm arises from defective functioning, from a simple hallucination, or from incorrect predictions of the automated system—and it is widely accepted that responsibility can be exclusive neither to the producer nor to the implementer, even though both parties ought to assume responsibility in part, and even juridical liability for those errors that ought to have been avoided [41].

Algorithmic errors, when they appear, generate challenges that are without precedent for the traditional models of medical responsibility, and this imposes a transition from individualistic conceptions towards frameworks of distributed responsibility—not only among the three actors of producer, user and physician, but also at the level of a wider society or community, with respect to communal expectations, lifestyle, the level of information and the degree of acceptance of these technologies. The notion of shared accountability organisations recognises, in effect, the collaborative nature of the AI-endowed medical ecosystem, within which developers, health institutions and clinicians contribute complementarily to the decisional process. This approach facilitates the proportional attribution of responsibility and thereby bridges the responsibility gaps that arise in those situations in which no single actor fully satisfies the traditional conditions of moral responsibility or of juridical liability. AI systems cannot themselves be held responsible, since from a moral point of view they possess no status as autonomous moral agents; yet in the process of automated decision moral agency is at times simulated, with the consequence that the producer, the end user, and likewise the intermediaries of access to AI assume responsibility for the automated decision—even though, in purely formal terms, this presents itself as a liability for the acts of third parties.

The charge of excessive abstraction is best met by concrete cases. The Epic Sepsis Model, a proprietary system deployed across hundreds of hospitals in the United States, was for years relied upon in clinical practice before an independent external validation established that its discrimination and calibration were substantially poorer than its adoption presupposed [42]. The instructive feature, for our purposes, is not the technical failure but the distribution of answerability that surrounded it: the vendor had disclosed no adequate evaluation, the adopting institutions had conducted none, the clinicians who acted upon its alerts were not positioned to assess it, and the patients affected had no standing from which to interrogate it. No party satisfies the traditional conditions of moral responsibility, and yet harm is systematically possible. This is the responsibility gap in its concrete form.

It is here that the first partial argument for our paper declares itself. The migration from individual to distributed responsibility is not merely a juridical convenience; it is the symptom of a deeper dislocation. The principlist scheme of Beauchamp and Childress presupposes, as its silent ontological premise, a unitary moral agent—a human author who decides, who knows, and who may therefore be held to answer. When the decisional process is genuinely shared between human and non-human contributors, that unitary agent ceases to be the basic ontological unit of the medical act, and dissolves into a hybrid in which authorship is irreducibly diffuse. The responsibility gap is, on this reading, the first fissure through which the anthropological singularity becomes visible within bioethics: the framework continues to address a subject whom the technology has already begun to disaggregate.

A tradition external to the one within which principlism was formulated permits this conclusion to be sharpened rather than merely restated. Eastern Christian thought has long distinguished the person, constituted in and by relation, from the individual conceived as a self-enclosed substance, and has held the latter to be an abstraction rather than the primary ontological unit [43]. Read from that perspective, the responsibility gap does not disclose the loss of something the medical act formerly possessed; it discloses that the unitary agent presupposed by the four principles was itself a construction, and that its dissolution under algorithmic mediation exposes a deficiency in the framework rather than a novelty in the technology. The convergence with the distributed account of agency invoked in Section 2.1 is not accidental: both deny that authorship is ever as concentrated as attribution requires it to be.

3.2. Distributive Justice and Algorithmic Stratification of Access

Distributive justice postulates that the benefits and the burdens of medical care ought to be equitably distributed across society, without arbitrary discrimination founded on criteria extraneous to medical need. This principle becomes especially pressing in the context of artificial intelligence, for these emergent technologies risk amplifying the inequities already present within health systems, creating new stratifications in access to quality care. Disparities will arise in access to digitalised medicine, since the digital revolution generates a novel form of inequity that stratifies populations according to differential access to technology: advanced systems of diagnosis and treatment are disproportionately concentrated in the medical institutions of high-income countries and in urban centres, leaving rural populations, marginalised communities and resource-limited nations without access to these transformative capabilities.

The prohibitive costs of the necessary infrastructure—high-performance servers, data-storage systems, specialised technical personnel—raise substantial barriers for institutions with constrained budgets. Even where the technology becomes available, the requirement of digital literacy, for medical professionals and for patients alike, introduces a further level of exclusion for vulnerable groups: for the elderly, for persons with limited education, for linguistic minorities. The paradox of this situation is that the populations who would benefit most from the predictive and preventive medicine facilitated by AI—those with a high burden of chronic disease, limited access to specialists, and reduced economic resources—are precisely those systematically excluded from access to the technology. Such an inversion of ethical priorities transforms artificial intelligence from a potential instrument for the democratisation of medical access into a mechanism for the consolidation of existing privilege.

A second order of stratification, less visible than that of access to diagnostic tools, yet more consequential in the longer term, concerns the asymmetric extraction of the very data on which medical AI is trained. We hold that the disparity is not only one of who receives the technology, but also of whose bodies furnish the corpus from which the technology learns and whose bodies remain epistemically invisible to it. As Pope Leo XIV has recently observed, the appropriation of personal lives in the form of exploitable information constitutes a new mode of colonialism: health data, epidemiological profiles, genetic maps and demographic information of entire peoples—frequently collected under the appearance of aid, research or innovation—operate as the rare earths of contemporary power, conferring upon those who aggregate them a structural leverage over the future of the populations from whom they were drawn [44]. The medical-sociological consequence is twofold: the populations from whom data are extracted gain no proportionate share in the benefits of the systems thereby produced; and the systems themselves, trained on corpora skewed towards populations with the means to be measured, may generalise poorly to those who were never represented. The algorithmic stratification examined here is, on this reading, the downstream symptom of an upstream stratification of epistemic inclusion, and the two together compose a single mechanism by which the digital revolution displaces equity rather than enacting it.

The very technical architecture of AI systems perpetuates and amplifies systematic discrimination through what is termed *algorithmic bias* [45]. Machine-learning algorithms can reproduce and encode the prejudices present in society, because these are inherited from the data on which the systems are trained, transforming historical disparities and existing discriminatory practices into automated decisional protocols. A body of studies quoted above documents that algorithms used at scale in health systems—at least in those of the United States—may systematically underestimate the medical needs of African American patients, thereby perpetuating racial disparities in access to care; the canonical demonstra-

tion concerns a single widely deployed commercial algorithm, and the generalisation from it ought to be made with corresponding caution. Such biases are not accidental; they reflect the asymmetries within the training data, which arise from the under-representation of certain populations in medical databases and from problematic design choices, including the use of prior medical expenditure as a proxy for health need—a choice that penalises patients with a limited care history, because disadvantaged groups historically spend less on care not because they are healthier but because they have had less access [45].

In practice, populations with limited access to health systems will be under-represented in the training bases of artificial intelligence; and where particular genetic profiles or mutations are more prevalent within those very populations, they will not be adequately integrated into the algorithms of medical decision, so that it becomes difficult or impossible for an expert system to detect pathologies specific to these vulnerable groups. The empirical record on dermatological imaging is now unambiguous on this point: the principal public datasets are severely skewed towards lighter phototypes—in a systematic review of publicly available skin-cancer image datasets, only a small fraction of the images with recorded skin type represented darker phototypes [46]—and the consequence is a measurable loss of diagnostic sensitivity, with one widely cited system (DeepDerm) falling from a sensitivity of 0.69 on lighter skin to 0.23 on darker skin, and a second (ModelDerm) from 0.41 to 0.12 [47]. A recent preprint reports that the disparity persists even where darker phototypes are proportionally sampled, which suggests that representativity must be understood as a dynamic property of model outputs rather than a static feature of the dataset [48].

Two further cases show that the disparity is inscribed upstream of the algorithm. Pulse oximetry, the ubiquitous instrument on which oxygenation is assessed and from which the training data of countless downstream systems derive, has been shown to overestimate arterial oxygen saturation in patients with darker skin, so that occult hypoxaemia goes undetected at differential rates [49]. Here the bias is not learned from the data; it is manufactured by the sensor that produces them. Conversely, the removal of the race coefficient from estimated glomerular filtration rate, following sustained normative deliberation, demonstrates that such inscriptions are revisable [50]: a correction which had been defended as empirically grounded was withdrawn once the question was reframed from what the data showed to what the classification did. The two cases together mark the boundary of our claim—stratification is structural, but it is not fated.

Gender bias likewise appears, though its mechanism must be stated with precision, for the temptation to explain it by differential care-seeking is both unsupported and misleading. Diagnoses are delayed or mistaken—for women with cardiovascular conditions, for instance—not because men consult more readily and are therefore better quantified, but because women have been historically excluded from clinical trials and under-represented in cardiovascular research, and because the diagnostic criteria themselves were calibrated upon male physiology, so that female symptomatology diverges from the male patterns that the training data treat as the norm [51,52]. The reproduction of this asymmetry by contemporary systems is now measurable: a large language model prompted to generate cardiovascular cases over-represented men by some thirty per cent and under-represented women by a comparable margin, and diagnosed female patients less accurately for conditions as grave as coronary artery disease and abdominal aortic aneurysm [53]. Algorithms for the recognition of dermatological images, similarly, perform sub-optimally for darker skin tones owing to the want of diversity in their training data, creating disparities in the early detection of melanoma—a disparity of life and death, since darker-skinned patients already present later and die more often of the disease [47,48]. Such algorithmic discrimination presents itself invested with technical authority and an apparent objec-

tivity conferred by data science, masking subjective prejudice beneath a semblance of mathematical neutrality; and the systematic, scalable nature of algorithmic error means that these discriminations do not affect isolated cases but may impact entire populations simultaneously. This observation admits, moreover, of generalisation: algorithmic decision is biased, quite generally, by the want of axiological neutrality in the informatic system—a system conceived in conformity with the prevailing current of thought within a given domain, and one that thereby diminishes the prospects of any alternative opinion securing recognition and acceptance. In the medical case, the intuition of an exceptional clinician may outperform an automated diagnosis; yet the presumption of trustworthiness that attaches to the algorithm, set against the opinion of the human physician, may issue in the rejection of the treatment the latter recommends—thereby vitiating the informed consent so obtained. Algorithmic bias could become an axiological bias and an epistemological one.

This is the second partial argument. Algorithmic stratification is the social form of the artificial evolution we have named: it determines, before any biological intervention, who is admitted to the trajectory of AI-augmented medicine and who is left upon the older, slower line. The stratification is not a contingent failure of distribution to be corrected by better sampling; it is the mechanism by which a self-directed evolution selects its beneficiaries. The realisation of distributive justice in the age of artificial intelligence therefore requires deliberate and sustained interventions of public policy, capable of counteracting the natural tendency of the market towards concentration and inequity, lest the membrane between the included and the excluded harden into the first frontier of the post-human condition.

A further stratum of inequity, less visible than the demographic one yet no less consequential, concerns the medical profession itself. The skilled solicitation of artificial-intelligence systems—what is now termed prompt engineering—constitutes, in our view, an emerging third literacy, taking its place alongside the classical alphabet of letters and the digital literacy of the recent past. Physicians who have not acquired this competence will be disadvantaged in their access to the most refined outputs of the technology, and so will their patients by extension; physicians who, conversely, surrender clinical reasoning altogether to the responses generated through their prompts risk being substituted, in the patient's regard, by the virtual practitioner itself. The implication is twofold: medical curricula must integrate this third literacy as a formal competence, and regulatory frameworks must take account of a stratification that no longer falls only between populations, but also within the profession that serves them.

Public investment in digital health infrastructure ought to be directed strategically towards disadvantaged regions and communities; programmes of training and capacity-building for medical professionals in resource-limited settings can secure the competent use of the available instruments; standards of regulation ought to impose the obligatory validation of algorithms across diverse populations, in a manner analogous to the validation of therapeutic schemes; and policies of open data and the responsible sharing of medical data sets can facilitate the development of representative algorithms. Mechanisms of independent audit and of the continuous monitoring of algorithmic equity, together with legal frameworks against algorithmic discrimination that impose clear sanctions upon developers and institutions, create the incentives for the prioritisation of equity. The democratisation of the benefits of AI in medicine is, on this reading, not merely an ethical aspiration but a strategic necessity for the legitimacy and the sustainability of health systems in the digital era.

3.3. *Structural Conflicts Between the Principles and the Opacity of the Human*

The integration of artificial intelligence into medicine generates tensions between the fundamental bioethical principles [54,55], the most salient of which is the conflict between

the transparency required by the principle of autonomy and the algorithmic performance required by beneficence. Deep-learning methods, with their superior performance, are frequently regarded as opaque, functioning as so-called black boxes whose internal processes remain insondable even to their own developers. Models built upon deep-reasoning systems are sometimes said to expose the information on which the response was generated together with the structure of the reasoning that led to it; this appearance, however, must be received with caution, for the reasoning trace a model displays is not a faithful report of its internal computation—it may be reconstructed after the fact, and the premises that genuinely conditioned the output are by then already sedimented in the training and no longer rendered explicit. The opacity is thus displaced rather than dissolved.

Something formally similar arises within specifically human processes of thought, where premises that operate beneath consciousness nonetheless influence rational processes, rendering human deliberation itself identifiable as a black box—its internal workings differing significantly from one individual to another according to premises that are neither conscious nor, in some cases, identifiable as such. We draw from this symmetry not the reassurance that algorithmic opacity is benign, but a conclusion that serves our central thesis: if the human mind and the algorithmic system are alike opaque, then the epistemic frontier that the tradition drew between the transparent human reasoner and the inscrutable machine begins to blur. The privilege of intelligibility on which the autonomous human agent was founded is precisely what the anthropological singularity calls into question.

To prioritise explicability through the use of simpler algorithms may compromise diagnostic accuracy and so impair beneficence; conversely, to opt for maximal performance undermines the patient's capacity to understand and to control the decisional process that affects them. Distributive justice and the efficiency of resources enter into a similar competition: algorithms for the allocation of scarce medical resources—intensive-care capacity, organs for transplantation, clinical time—may optimise aggregate outcomes at the level of the population, maximising the years of life saved or the total quality of life, yet these utilitarian optimisations may discriminate systematically against elderly patients, against those with disabilities or with a reserved prognosis, thereby violating the principle of equity and of the equal dignity of all persons. A conflict arises, finally, between precaution as a dimension of non-maleficence and innovation as potential beneficence: the validation of AI systems prior to implementation may delay patients' access to potentially life-saving technologies, while premature implementation risks systematic harm.

A particular instance of the same tension concerns the legal and ethical instrument by which patient data are made available for the training of future medical AI. Specific consent—in the classical sense of an authorisation given for a determinate purpose, communicated in advance—becomes structurally impossible where the very purposes for which the data will be reused are not yet defined at the moment of collection. The notion of broad consent, by which the patient authorises an open class of future research uses subject to ethical oversight, presents itself as a candidate solution, and it has been defended on the ground that the distinction between kinds of consent is a distinction between the kinds of choice a person makes rather than between degrees of information, so that a broad authorisation may be as fully autonomous as a specific one [56]. Against this it has been argued that the justifications ordinarily advanced—high benefit, respect for autonomy, consistency with existing practice, low risk—do not survive examination, and that broadening consent dilutes the very ethics it is invoked to satisfy [57]. The dispute is not resolved here, and we note only that it turns on the same question as the conflict between explicability and performance: whether an authorisation whose substantive content remains, by design, indeterminate preserves autonomy or merely satisfies it procedurally.

The programming of AI systems thus imposes an explicit hierarchisation of ethical values—a decision that, in human medical practice, ordinarily remains implicit. Algorithms cannot feel ethical tension and cannot exercise contextual judgement; they must be furnished with explicit rules as to how to arbitrate between competing imperatives, and such a codification of priorities forces an uncomfortable clarity upon the values that guide medicine. Where the system instead infers its priorities from the analysis of prior practice—as deep-learning models trained on vast clinical databases will do—the emergence of algorithmic bias makes the system decide who are to be the priority beneficiaries based on past patterns rather than of any axiological deliberation, and the decision can no longer be contextualised to the concrete bioethical or clinical situation, nor to a particular social, cultural, geographic or ethnic setting. When a triage algorithm decides who receives limited resources, the programmed criteria—the maximisation of survival against the equality of chances, efficiency against equity, the saving of the youngest against respect for older people—are presented as neutral, yet they ought to reflect choices of profound axiological charge.

A further difficulty lies in the rigidity of algorithmic hierarchisation. Human physicians can modulate ethical principles according to context, applying different rules in situations of emergency, of routine care, or of patient vulnerability; algorithms tend, by contrast, to apply the same rules uniformly. Such consistency may be ethically correct, eliminating the arbitrary, but it may equally constitute a problem where the context calls for moral flexibility. Since contemporary models can be rendered sensitive to context, such biases may be mitigated by personalising the operation of the automated decision according to the prior practice of the beneficiary institution—a personalisation for which agent-based architectures, updatable locally, are better suited than large language models alone. It must be conceded that personalisation according to prior institutional practice reintroduces, at a local scale, the very reliance on past patterns criticised above; the difference we defend is not that local patterns are bias-free, but that they remain answerable to, and revisable by, the community upon whom they bear—whereas the historical bias encoded in population-level training data answers to no one. The legitimacy of contextual variation rests on this revisability, not on any presumed neutrality.

This is not a speculative proposal. When the sepsis model discussed above was revised and prospectively evaluated across four major health systems and some two hundred thousand encounters, aggregate discrimination proved acceptable, yet performance varied markedly by institution, positive predictive value remained low, and the alert burden was high—leading the investigators to conclude that adopting institutions must conduct their own local validation rather than rely upon the manufacturer's figures [58]. The finding is precisely what our argument predicts. A system's ethical adequacy is not a property it carries with it from the site of its production; it is constituted locally, and it is therefore locally answerable—which is the whole of the case for contextual pluralism advanced above.

The proposal—defensible on grounds of clarity and accountability—that medical AI systems should embed an explicit hierarchy of ethical values within their constitutive algorithms, in the manner of Asimovian laws applied to clinical robotics, carries with it a tension that ought to be acknowledged. A universal codification of priorities, while it disciplines algorithmic decision and forecloses arbitrary inferences from training data, simultaneously erases the cultural and axiological diversity of the communities upon whom the technology is deployed. The risk is that of an axiological imperialism enacted by software: the imposition, under the appearance of mathematical neutrality, of a single moral architecture upon health systems whose underlying value-frameworks differ. Such a risk is not, we hold, an exterior abuse of an otherwise neutral instrument: as Pope Leo XIV

has noted, no technical tool is in fact morally neutral, since each embodies, through what it measures, ignores and optimises, and through the manner in which it classifies persons and situations, an implicit vision of the human and of society [44]. Ethical discernment of medical AI cannot therefore confine itself to asking whether a system is used for good or for ill; it must also interrogate how that system has been designed, and what anthropology has been inscribed in the data and models that drive it. The proposal of an embedded hierarchy of values is, in this sense, not the introduction of an axiology into a value-free engine, but the substitution of an explicit axiology for one that was already operative under the appearance of neutrality. The augmented bioethics we shall advocate must therefore include, alongside its mechanisms of standardisation, a deliberate provision for contextual ethical pluralism—that is, for the legitimate variation in priorities across jurisdictions whose social, religious and political traditions cannot be simply averaged away. This is the third partial argument: the impossibility of arbitrating conflicts of principle from within the principlist scheme reveals that the scheme presupposes an ontology of the human—autonomous, transparent, culturally situated—that algorithmic medicine is in the act of unsettling.

4. Discussion

4.1. From Clinical Intuition to Algorithmic Decision: The Cognitive Vector

The analyses set out above converge upon a single, broader transformation: a reconfiguration of the nature of medical expertise [59]. Where classical medicine valorised clinical intuition—that amalgam of experience and judgement, sedimented across hundreds or thousands of cases, that constituted the summit of competence—algorithms perturb this epistemology, substituting for intuition an explicit probability calculated from massive data. An experienced physician senses when something is amiss with a patient; an algorithm computes a percentage probability for a given pathology on the basis of a number of variables. The physician becomes, in consequence, less a repository of knowledge than an interpreter of algorithmic output, a mediator between probabilistic mathematics and the human reality of suffering, and the clinical role shifts from decision-making towards supervision—a new competence requiring both technical understanding and ethical vigilance towards the limitations and the biases of expert systems.

Paradoxically, dependence upon algorithms risks atrophying the very clinical intuition that ought to supervise and correct algorithmic error. The phenomenon now has an empirical name and an empirical record: clinical de-skilling, the gradual erosion of independent reasoning under habitual reliance on automation. Multicentre evidence from endoscopy is striking—continuous exposure to AI assistance was associated with a measurable fall in unassisted adenoma detection during subsequent procedures, even among experienced operators [60]; and the scoping literature extends the concern from technical to cognitive and even moral competence, warning of a ‘moral de-skilling’ in which the automation of decision distances the clinician from the ethical dimensions of care, and of a ‘never-skilling’ of trainees who delegate before they have learned [61]. The risk is not, however, fated: the same literature is careful to record that de-skilling is not an inevitable consequence of AI but a consequence of its unreflective use, and that suitable institutional arrangements—calibrated confidence signals, hybrid training, the deliberate preservation of unassisted practice—can convert the technology from a substitute for judgement into an augmentation of it [61].

Read through the lens of our paper, de-skilling is the cognitive face of the anthropological singularity: a directed evolution of human capability itself, in which the clinical faculties that took a professional lifetime to sediment are reshaped—and, in part, surrendered—under the pressure of the technical environment. The transformation is not of the tools the

physician uses but of the kind of cognitive being the physician is becoming. A correlative figure emerges from this displacement, which we propose to name digital paternalism. Classical medical paternalism rested on the asymmetry between an informed physician and a patient presumed less competent to weigh the relevant considerations; digital paternalism rests, by contrast, on a triadic asymmetry, in which the human physician becomes the operator of a system whose recommendations are no longer fully intelligible even to those who designed it or managed it. The patient, in such a configuration, no longer enters a therapeutic alliance with the physician; he enters, rather, a relation of compliance with an automated decision that the physician transmits. The risk is intensified where the AI system is approached, by clinicians and patients alike, in the register of what Bostrom has termed a *God-like superintelligence* [4]—that is, as an authority whose pronouncements need not be deliberated but only obeyed. The right of the patient to refuse, central to the entire architecture of consent, is precisely what such a configuration places under pressure.

4.2. Virtualised Therapeutic Relations and the Limits of Synthetic Empathy

The physician–patient relation, with its humanist foundation, undergoes subtle but profound transformations through the introduction of algorithm-based medicine. When an algorithm generates the diagnosis or the therapeutic plan, the question arises of who is the epistemic and moral author of that decision, and whether the patient establishes a relation of trust with the physician, with the algorithm, or with a hybrid physician–AI entity whose responsibility is diffuse. Medical chatbots and virtual assistants create simulacra of therapeutic relations—interactions that mimic empathy and personalisation without genuine subjectivity. This artificial empathy may be therapeutically useful, reducing anxiety and improving adherence to treatment; yet experimental research demonstrates a paradoxical effect, in which expressions of empathy that presuppose experiential capacities the user does not credit a machine with having are perceived as inauthentic, and this perceived inauthenticity suppresses the very gain in trust that empathy was meant to produce—a backfiring effect that the same studies do not find in human-to-human interaction [62]. The distinction between the dimensions of empathy becomes essential here: while the cognitive dimension, the understanding of another’s perspective, may be approximated algorithmically through the processing of language, the affective and motivational dimensions—the sharing of another’s feeling—remain inalienably human, at least for the moment, defining the intrinsic limits of synthetic empathy [63].

Algorithmic telemedicine, where interactions are mediated by digital platforms that gather, analyse and present data, may distance the physician from the patient both physically and emotionally, transforming the clinical encounter from an experience into an exchange of digitalised information. This is the virtualisation of social space brought to bear upon the medical act itself. The sociology of space helps to name what is at issue: space, on Löw’s relational account, is constituted not by metric distance alone but by the placing of social goods and persons and by markings that are primarily symbolic [8]; and phenomenological analyses of the virtualised life-world show that the ‘paramount reality’ of co-presence can no longer be equated with the zone of primary relevance for actors who live and act within mediated realms [9]. The clinical encounter, transposed into this register, is governed no longer by the effective distance between two bodies in a consulting room but by symbolic and communicational distances—by the protocols, interfaces and latencies of the platform. Such virtualisation may increase efficiency while risking the impoverishment of medicine in its human and relational dimension. Communication between physician and patient is held to possess in itself a therapeutic effect, if only through the placebo effect it generates; for the dialogue between human and machine to generate a comparable effect, the patient’s trust in the technology would have to equal, if not exceed, their trust in

their own attending physician. Yet there is registered, justifiably or not, a growing distrust of science—medical science included—and this distrust is liable to extend to advanced technologies, medical technologies among them.

The democratising potential of AI-assisted telemedicine—particularly for populations whose access to specialised care would otherwise be prohibitive—must be weighed against a recent and instructive risk: the temptation of patients to treat the outputs of conversational AI as prescription, and so to substitute self-medication for medical consultation. Legislative responses have moved, within a short interval, from proposal to binding law. In the State of New York a draft initiative would have obliged providers of AI systems offering professional advice—medical or legal—to disclose explicitly the non-human nature of the counsel given; what that initiative proposed for a single jurisdiction, Article 50 of the Union’s Artificial Intelligence Act now imposes across the internal market, its transparency obligations having become applicable in August 2026 and having been expressly excluded from the deferral effected by the amending Regulation of July that year [39,62]. Such instruments, however imperfect, mark an important conceptual shift: from the diffuse moral responsibility of producers towards a juridical liability anchored in a specific and enforceable duty to inform.

What is at stake in this displacement, we hold, exceeds the technical question of whether the synthetic relation is therapeutically effective. The quality of a civilisation, as Pope Leo XIV has recently formulated it, is measured less by the power of its means than by the care it is able to offer, and by its capacity to recognise the other as a face and not merely as a function [44]: a criterion under which the therapeutic relation, however efficiently mediated, is one of the privileged sites in which a society discloses what it takes the human person to be.

The criterion admits of a more exacting formulation. Where an ethics is constituted as conformity to rule, the therapeutic relation can be satisfied by the correct execution of a protocol, and there is then no principled objection to its execution by a system; where ethics is understood instead as an event of relation, in which the moral quality of an act is inseparable from the encounter that carries it, no execution however correct can substitute for the encounter [64]. The distinction is not devotional but analytic, and it converges with what the anthropology of illness establishes on independent grounds [31]: what an algorithm approximates is the correct handling of a category, whereas what a patient brings is a meaning that only another person can receive.

4.3. Emergent Ethical Dilemmas: Prediction, Prevention and the Medicalisation of the Species

Predictive medicine, founded upon artificial intelligence, can identify the risk of a given pathology decades before its symptomatic manifestation, and this capacity creates dilemmas without precedent. What is to be done with a prediction of, say, a sixty-five per cent risk of a given pathology twenty years hence? Is the patient to be informed, at the cost of a chronic anxiety concerning an uncertain future, or is the information to be withheld, in disregard of their autonomy? Preventive interventions founded upon probabilistic predictions subject healthy individuals to treatments bearing adverse effects, in order to prevent diseases that might never have manifested but that are identified, through the analysis of the genome, as a strongly indicated risk. Such an over-medicalisation of health and of social life transforms risk into pathology, raising questions concerning the proportionality of intervention within a society already profoundly medicalised.

These dilemmas are sharpened by the crisis of trust that accompanies the rapid development of biotechnology. Artificial intelligence is increasingly enlisted in the generation of new active substances, by way of genetic engineering and of synthetic and molecular biology; yet the velocity of this development generates distrust among a broad part of the

population, a distrust bearing both upon insufficient testing and upon the massification of the medical act to the neglect of the real needs of the individual patient. The case of the messenger-RNA vaccines is instructive: the introduction of mRNA-based vaccines barely a year after the onset of the COVID-19 pandemic engendered, in many persons, uncertainty as to the correctness—medical and ethical alike—of their testing and as to the sufficiency with which possible long-term adverse effects had been observed, contributing to a decline of trust not only in that technology but in modern medicine and in the goodwill of the medical world more generally. The transition this section marks is decisive for our argument: from the medicalisation of the individual life to the medicalisation of the species, the threshold at which predictive and preventive medicine begins to act not upon the patient before us but upon the biological future of those not yet born.

4.4. Gene Editing as Directed Evolution: The Biological Vector

Within this same horizon, the conjunction of artificial intelligence with the techniques of genetic editing—CRISPR being the paradigmatic instance—opens a question of a different order, and it is here that the anthropological singularity ceases to be metaphor and becomes literal. Where AI-guided analysis permits the identification of congenital predispositions before birth, and where germline editing offers the prospect of eliminating them, medical technology shifts from the protection of health to what is now openly called human enhancement: a step beyond the imperatives of cure towards the deliberate modification of the species' biological condition. This is the precise sense in which humanity withdraws from natural selection. John Harris has given the thesis its sharpest formulation, arguing that we should enter a new stage of evolution in which a directed evolution through genetic enhancement replaces the traditional means of natural selection, so that enhanced evolution succeeds Darwinian evolution as the mechanism of our becoming [10]. The cumulative effect of such interventions, on the more radical readings, is nothing less than the conversion of the human into a successor species—the 'posthuman' of the transhumanist programme, whose capacities so exceed our own as to be no longer unambiguously human by present standards [65].

The threshold has already been crossed once, and the manner of the crossing is itself instructive. In 2018 a researcher announced the birth of children whose germline had been edited, an intervention conducted outside any established framework of oversight and repudiated by the international scientific community; the subsequent criminal conviction and the moratoria proposed in its wake established, in effect, that the governance of germline modification was to be constructed retrospectively, around an act already committed [66]. What the episode demonstrates is not that such interventions are impossible to regulate, but that the regulatory apparatus is at present reactive—a condition that the precautionary reasoning of our Section 3.1 was intended to forestall.

The framework of procreative beneficence, in the sense given to it by Savulescu [11], places upon prospective parents a responsibility not merely for the health of their descendants but for the possible inequity between those whose genome has been edited and those whose has not. A novel form of stratification thereby announces itself, one that prolongs the algorithmic stratification of Section 3.2 into the very flesh of future generations, and that does so on the basis of a consent—that of the unborn—which only the parents can be presumed to give. The medical-sociological reading is unavoidable here: just as differential access to AI sorts the living into the augmented and the unaugmented, differential access to germline editing threatens to sort the unborn into the 'valid' and the 'invalid', reinscribing an old eugenic logic beneath the appearance of reproductive choice [67].

Taking as our point of departure the model of procreative beneficence proposed by J. Savulescu [11], the concept of a *digital beneficence* might be constructed—in the sense of an

obligation incumbent upon care-givers in the domain of health to make responsible use (responsibility by design) of AI in diagnosis and in the proposal of personalised treatments, for so long as such technology is available; otherwise, the duty of diligence that falls to medical personnel would not be observed. Were this obligation to employ AI wherever it is possible and available to be instituted, there would be created, correlatively, the patient's right to *digital autonomy*—that is, the right to be informed as to the manner in which AI is employed, as to the beneficial results and, respectively, the risks attendant upon its use—thereby instituting a specific informed consent for the use of AI, one that would be prior to any such use, yet a *dynamic* consent, sustained throughout the entire course of the employment of these instruments.

Intellectual honesty requires that the contrary voices be given their due, and they are not negligible. A first qualification is technical: the euphoria that greeted CRISPR in transhumanist circles rests, as Ranisch has argued, upon an over-simplified understanding of genetics and upon exaggerated expectations of what germline intervention can in fact achieve, most complex traits being far less tractable than the rhetoric of enhancement supposes [68]. A second qualification is ontological, and it is the more important for our purposes: the transhumanist case presupposes a naturalist conception of the human, the person reduced to modifiable matter, and this conception stands in direct contrast to the classical, Aristotelian-Thomistic and personalist ontologies for which the human person is constituted by culture, freedom and an irreducible dignity rather than by an editable genome [69]. A third qualification, convergent with the personalist one and recently articulated within the contemporary public debate on AI, comes from the social doctrine of the Church, where Pope Leo XIV has set the distinction in unusually sharp terms: it is one thing to integrate technology within a human-centred, relational vision, quite another to be guided by an outlook that devalues human limits and promises a purely technical form of salvation [44]. The same Encyclical recalls that incapacity, illness, old age, suffering and vulnerability—the very conditions which germline enhancement proposes to eliminate before they appear—are not, in a more capacious anthropology, mere defects to be corrected, but realities through which the human person matures and opens to relationship [44]. We hold that this is the same ontological disagreement we have signalled above, registered now in the register of public theology rather than of academic philosophy, and significant precisely because it places the dispute over enhancement within the wider cultural assumptions that accompany the digital revolution. The dispute over enhancement is, thus, at bottom, a dispute over the ontology of the human—which is precisely the terrain on which the anthropological singularity is decided. Our thesis does not require us to adjudicate the dispute; it requires only the recognition that the dispute is now a clinical, and not merely a speculative, matter.

The position is not confined to a single confessional tradition, and its recurrence across traditions that differ in much else is itself evidence that the disagreement is ontological rather than doctrinal. The social ethos adopted by the Ecumenical Patriarchate of Constantinople (Orthodox Church) addresses medicine, healing and the new technologies within a frame in which the human person is understood as constituted relationally and eschatologically, so that a finitude which enhancement proposes to abolish is construed as a condition of personal maturation rather than as a defect awaiting correction [70]. Engelhardt, arguing from within the same tradition but in the register of academic bioethics, has pressed the point further: a bioethics severed from any account of the good it serves can adjudicate procedures but not ends, and the question of enhancement is a question of ends [71]. We do not require the reader to share these commitments. We require only the recognition that the dispute over enhancement cannot be settled by weighing risks against benefits, since what counts as a benefit is precisely what is contested.

4.5. Emergent Pathologies of Algorithmic Life

A further class of dilemmas, presently underestimated in the bioethical literature, concerns the iatrogenic effects of the technology itself—not on the body but on the psyche, and it represents the pathological extremity of the same virtualisation vector examined in Section 4.2 of this paper, now turned upon the psyche rather than upon the therapeutic relation. A clinical literature on these effects has begun to constitute itself, and it must be reported with the caution its present state requires. The possibility was anticipated before the cases appeared [72]; the first documented instance of new-onset psychosis associated with such interaction has since been reported [73]; and a functional typology has been proposed which distinguishes the roles a system may play—as catalyst, precipitating symptoms in a person previously well; as amplifier, worsening symptoms already present; as pseudo-coauthor, participating in the construction of a harmful narrative; or as object, becoming itself the focus of the delusion [74]. A scoping review of the surrounding public discourse records concerns extending to dependence, self-harm and suicide, while noting that the evidence base remains preliminary, that reporting is largely non-systematic, and that questions of prevalence and of causal contribution are as yet unsettled [75]. We register the phenomenon at this level of confidence and no higher. The point that bears upon our argument does not require more: that the same fluency which makes these systems therapeutically promising is what makes them capable of sustaining a distorted account of reality, and that a bioethics constructed for the pathologies of the body has no instrument with which to evaluate this. The investment of artificial intelligence with quasi-pastoral or quasi-divine attributes—a phenomenon that classical bioethics is ill-equipped to analyse—adds a further layer to this picture [76]. Most strikingly, the emerging practice of virtual immortality—the post-mortem emulation of a person’s conversational style, increasingly indistinguishable from the original through deep-fake reconstructions—risks pathologising the anxiety of death itself, and blurring the patient’s capacity to discern the human interlocutor from its synthetic double [77]. These are not marginal curiosities, but the symptoms of a life increasingly lived at symbolic rather than spatial distances: forms of ubiquity and atemporality, mediated by AI, through which the subject detaches from the limitations specific to the human condition—mortality, locatedness, finitude—that bioethics had always presupposed. Where medicine has historically been concerned with the diseases of the body, an augmented bioethics will have to acknowledge that the technology designed to assist medicine generates, in its own turn, the pathologies of the next clinical epoch.

4.6. Towards an Augmented Bioethics and Polycentric Governance

Taken together, these transformations call for what we propose to term an *augmented bioethics*. Because the expression is liable to three distinct readings, and because the difference between them is not terminological but substantive, we state at the outset which of them we intend.

The term does not designate, first, the delegation of moral judgement to algorithmic systems—an automated ethics in which the arbitration between competing imperatives is itself computed. We have argued against precisely this in Section 3.3 of this paper: a system that infers its axiology from prior practice does not deliberate, but sediments, and the appearance of neutrality it thereby acquires is the very thing bioethics exists to contest. Nor does it designate, second, the mere extension of the four principles to a new class of objects, the algorithm taking its place beside the ventilator and the transplanted organ as one more technology to which an unaltered normative framework is applied. This second reading is the more insidious, since it preserves the appearance of continuity while conceding nothing:

it presupposes intact the unitary, transparent, culturally situated agent whose dissolution our three partial arguments have traced.

What we intend is a third aspect, and it is reflexive rather than instrumental. An augmented bioethics is a bioethics that has taken into its own deliberative structure the recognition that moral agency, epistemic access and distributive outcome are now co-produced with non-human systems, and that revises its ontological presuppositions accordingly rather than merely enlarging its field of application [78]. The sense of ‘augmented’ is that of augmented reality rather than of augmented intelligence: what is instrumented is not the deliberating subject but the frame through which the object is apprehended, and the object—the medical act, and the human person who undergoes it—is disclosed, not replaced. The proposal is therefore continuous with the diagnosis of Mittelstadt [79], who has shown that the transposition of principlism from medicine to artificial intelligence fails for want of the professional structures, the shared aims and the mechanisms of accountability that gave the principles their purchase in the clinic; our response differs from his in that we locate the deficiency not in the transposition alone but in the ontology of the agent that the principles presuppose on both sides of it.

Five commitments follow, each already established in the analysis above. One: responsibility is attributed across a distributed architecture of human and non-human contributors, and the responsibility gap is treated as a structural feature to be governed rather than an anomaly to be closed by locating a culprit (Section 3.1). Two: explicability is treated as a dimension of performance and not as an external constraint upon it, so that a system which cannot render its grounds intelligible is by that fact deficient rather than merely opaque (Section 3.3). Three: equity is treated as a systemic property of the socio-technical ecosystem rather than as an aggregate of individual outcomes, and algorithmic bias is accordingly an object of institutional design and not of technical correction alone (Section 3.2). Four: the legitimate variation in axiological priorities across jurisdictions and communities is provided for constitutively, not tolerated as a residue, on the grounds developed in Section 3.3 of this paper. Five, and this is what makes the framework reflexive in the strict sense: bioethics submits its own instruments to the scrutiny it directs at clinical ones, for the deliberation that judges algorithmic mediation is itself increasingly algorithmically mediated—a condition we have declared, in respect of the present study, in the back matter. Table 3 sets out the contrast in summary form.

The ethical governance of medical AI requires, correspondingly, multi-stakeholder structures that include not only developers and physicians but also patients, ethicists, jurists and the representatives of the affected communities. Ethics committees ought to proceed from the ex ante evaluation of protocols to the continuous monitoring of deployed systems, of their capacity for intervention and of the problematic patterns identified within them; also, controlled spaces—regulatory sandboxes—ought to be established for the testing of innovations under intensive ethical supervision, permitting responsible experimentation. An approach organised around the entire life cycle of the technology must recognise that the relevant stakeholders and the dominant ethical challenges vary significantly according to the stage of development, from conceptualisation and training, through clinical validation and implementation, to continuous monitoring and, where necessary, withdrawal from use. Ethical certifications for medical algorithms, analogous to FDA approvals for pharmaceuticals, could standardise the requirements of validation, transparency and equity.

Two limit-scenarios frame the interval within which the outcome will fall. In one of them, algorithmic medicine becomes predictive, preventive and participative, avoidable error is eliminated, and the patient becomes an informed partner in the management of their own health. In the other, biomedical surveillance becomes total, social life is medicalised to the level of the physiological parameter, stratification is organised around the genetic risk

score, and the irreducibly subjective dimension of the experience of illness is optimised away. Neither will be realised; the interval between them is the space of governance, and it is to the unresolved questions that structure it that we now turn.

Table 3. Classical bioethics and augmented bioethics: a structural contrast.

Criterion of Comparison	Classical (Principlist) Bioethics	Augmented Bioethics
Unit of moral agency	The unitary human agent who decides, knows, and may be held to answer.	A distributed architecture of human and non-human contributors; authorship irreducibly diffuse.
Locus of harm	The individual clinical encounter; error as a discrete event.	The socio-technical ecosystem; error as a scalable and systematic property.
Status of transparency	An external requirement of consent, satisfied by disclosure.	A dimension of performance: a system that cannot render its grounds intelligible is deficient as a system.
Conception of equity	The equitable distribution of a given benefit among individuals.	A systemic property, extending upstream to whose bodies furnish the training corpus.
Mode of arbitration between principles	Implicit, contextual, exercised in the clinical judgement of the practitioner.	Necessarily explicit and codified ex ante, and therefore itself an object of ethical scrutiny.
Relation to cultural variation	Presumed universality; variation treated as a residue to be accommodated.	Contextual ethical pluralism provided for constitutively; universality treated as a claim requiring justification.
Temporal form of evaluation	Ex ante approval of a protocol or device.	Continuous evaluation across the technology's life cycle, including withdrawal from use.
Reflexive status	The framework is presumed exterior to the practice it evaluates.	The framework acknowledges its own instrumentation and submits it to the same scrutiny.

4.7. Unresolved Controversies and a Research Agenda

The framework advanced here would be poorly served by a conclusion that suppressed what it cannot settle. Four difficulties remain open, and we set them out in the conviction that a conceptual proposal is tested by the questions it renders formulable.

The divergence of regulatory regimes. The comparative sketch offered in Section 4.6 of our paper requires qualification in the light of very recent developments, which illustrate the difficulty better than any abstract statement could. The European Union's Artificial Intelligence Act [41] established an architecture of ex ante conformity organised by risk class; yet the obligations governing high-risk systems, originally applicable from August 2026, were deferred by the amending Regulation adopted in July 2026—to December 2027 for stand-alone systems and to August 2028 for artificial intelligence embedded in products already regulated under EU product-safety law, medical devices among them [65]. The stated ground of the deferral was the incompleteness of the harmonised standards and conformity tools on which the regime depends. What applies from August 2026 is, instead, the transparency regime: the obligation to disclose to a person that they are interacting with an artificial system rather than a human one—precisely the duty whose emergence in a single American jurisdiction we noted in Section 4.2 of our paper, an obligation now binding across the internal market of EU.

Three observations follow. The first is that the contrast between a precautionary European Union and a pragmatic American regulator is less sharp than the comparative literature supposes: both are, at present, governing medical artificial intelligence through instruments that arrived after the technology did. The second is that the deferral is not a failure of political will but of technical prerequisite—one cannot certify conformity

against standards that do not yet exist—and this suggests that the rate-limiting step in the governance of algorithmic medicine is metrological rather than normative. The third bears directly on our thesis: a regime which regulates disclosure before it regulates safety has, in effect, conceded that the relation between patient and system is the site of the more urgent risk. That is the anthropological singularity legislating for itself.

These asynchronies generate questions we cannot resolve here. A system trained under one jurisdiction's data regime and deployed under another's liability regime occupies a position for which no settled doctrine exists; and the temptation of regulatory arbitrage—to develop where obligations are lightest and deploy where markets are richest—is structurally available in a way that the territorial logic of medical regulation was never designed to prevent.

Health systems with constrained resources. The literature on trustworthy artificial intelligence tends to assume an institution capable of bearing the cost of trustworthiness: independent audit, continuous post-market monitoring, documented data governance, local validation. Each is a genuine safeguard, and each is also a fixed cost. The paradox we wish to record is that the apparatus designed to prevent the technology from stratifying populations may itself become a mechanism of stratification, admitting to the market only those providers and those health systems able to finance compliance. The guidance issued by the World Health Organisation on large multi-modal models in health is explicit that governance must be adapted to settings of differing capacity [80]; but adaptation is not costless either, and the question of who bears the cost of making a technology safe for those with the least is, so far as we can determine, unaddressed. Whether such systems will leapfrog the infrastructural stages that wealthier systems traversed, or will instead be locked into dependence upon models validated on populations unlike their own, is an empirical question of the first importance and one that our framework can pose but not answer.

The limits of distributed responsibility. We have defended the migration towards distributed frameworks, and intellectual honesty requires that the objection be stated at full strength. Responsibility distributed across many contributors is, on a familiar analysis, responsibility that no contributor bears: the problem of many hands, in Thompson's formulation, is precisely that collective causation dissolves individual answerability rather than apportioning it [81]; and the responsibility gap identified by Matthias arises exactly where a learning system's behaviour is no longer predictable from the intentions of those who built it [82]. To name the gap is not to close it. A framework which distributes responsibility without specifying the mechanism of its enforcement risks providing every participant with a reason to point elsewhere—an outcome worse, for the injured patient, than the crude attribution it replaced. We hold that the answer lies in the direction of institutional rather than individual attribution, and in strict rather than fault-based liability for certain classes of algorithmic harm; but we advance this as a direction of inquiry, not as a solution, and we note that the EU's own attempt to legislate in this area has not, at the time of writing, produced a settled instrument.

The plurality of moral traditions. Worth recording is that such deliberation is already under way, and that it is institutional rather than merely academic. The papal encyclical invoked at several points above has itself elicited an articulated response from within Eastern Christianity, addressed to the same questions of personhood, enhancement and technological mediation and reaching partly convergent, partly divergent conclusions [83]; and the engagement extends beyond the theological academy, national bioethics bodies in more than one European jurisdiction having convened the discussion directly. This is contextual ethical pluralism in its operative rather than its postulated form: not the abstract concession that traditions differ, but the observable fact that they are presently arguing with

one another about the same technology, and that the arguments are entering the institutions through which that technology will be governed.

4.8. *Strengths, Limitations, and Further Research*

The most important achievement of the present study is represented by the delineation of a new societal and cultural paradigm, one capable of adapting itself to the moment of anthropological singularity described in this work—transformations which, although they appear across all socio-cultural areas, have been presented here from the perspective of healthcare, more precisely of digitalised medicine employing artificial intelligence. We regard this paradigm as a systematic mutation that has already begun to encompass society, producing mutations similar to—yet far more rapid than—those introduced by the discovery of fire or of the printing press. Whereas the slave-based model of production rested upon the brute force of individuals, and modern society—being knowledge-based—was developed upon the force of knowledge, the new paradigm may be defined as one grounded in prompting, that is to say, in expressions of will.

The principal opening which we consider this work to possess for future research will be precisely the investigation of this post-work society and of its ethical and civilisational implications. The universe of public health is—we hold—more transparent to this virtualisation of social life, yet also more exacting as regards the limits imposed upon the human condition. Other domains, among which we name here education and culture—although they stand at the dawn of this virtualising transformation—are perhaps more conservative with respect to the requirement of implementing a technoethics applied to those respective domains, there existing rather a polarisation between techno-pessimists and techno-optimists. Bioethics thus becomes a framework-model that may be transposed into diverse aspects of technoethics, for any other area of human evolution, towards a particular form of anthropological singularity.

A final asymmetry deserves record, since it bears upon the very pluralism this study has defended. The traditions that have engaged the biological vector of the transformation have done so at some length: the genome question has received sustained treatment within Islamic ethics [84], within Hindu bioethics—where it is elaborated, notably, within a framework that is itself philosophically pluralist and ethically contextual [85]—and within Jewish medical ethics [86]. The algorithmic vector is another matter. Beyond the Christian material examined above, sustained engagement is at present confined to a small number of traditions: Islamic ethics has begun to address artificial intelligence directly [87], and a Buddhist account has been developed which is of particular interest here precisely because it does not converge with the personalist argument we have advanced—contesting the anthropocentrism of the debate itself and proposing compassion towards all sentient beings, rather than the dignity of the human person, as the governing criterion of technological design [88]. We record the disparity rather than resolve it. If the anthropological singularity is what we claim it is—a mutation in the ontology of the human—then the traditions that have thought longest about that ontology are indispensable interlocutors, and their comparative engagement with the algorithmic vector is among the more urgent items on the agenda we have set out.

A research agenda. Six questions follow directly from the framework and are, in our view, tractable. (1) Can indicators of algorithmic stratification be constructed and measured comparatively across health systems, distinguishing stratification of access from stratification of epistemic inclusion? (2) Does clinical de-skilling follow the trajectory that the endoscopic evidence suggests when observed across specialties and over longer intervals, and which institutional arrangements measurably attenuate it? (3) What competence does the solicitation of algorithmic systems in fact require of clinicians, and can it be specified

sufficiently to be taught and assessed as a formal element of medical curricula? (4) How do patients from differing cultural contexts construe the authority of an algorithmic recommendation, and does the disclosure now mandated under the European Union's transparency regime alter that construal? (5) What forms of liability attribution, tested against actual cases of algorithmic harm, succeed in closing the responsibility gap without suppressing the innovation the technology promises? (6) Can the conceptual proposals advanced here—the anthropological singularity, algorithmic stratification, augmented bioethics—be operationalised into observables, and thereby subjected to the empirical scrutiny that a conceptual paper exists in order to make it possible?

5. Conclusions

Artificial intelligence in contemporary medicine does not represent an incremental technological innovation but a profound paradigmatic transformation, one that imposes a fundamental reconsideration of the entire conceptual structure of bioethics. We have argued that this transformation is best understood as the medical face of a wider mutation—the anthropological singularity—the threshold at which humanity withdraws from natural evolution and enters a self-directed, artificial one, along three vectors: the virtualisation of social space, the cognitive-relational reconfiguration of the human agent, and the modification of the biological substratum. The traditional principles—autonomy, beneficence, non-maleficence and distributive justice—remain valid as normative foundations, yet they presuppose an ontology of the unitary, transparent, culturally situated human agent that this mutation renders revisable; they require, accordingly, a significant recalibration if they are to answer the challenges specific to algorithmic decision, to probabilistic prediction and to the automation of clinical processes.

To the three partial arguments developed in Section 3 of this paper we may now add those that emerged in the Section 4; each of the six analyses offered here is, in our reading, a partial argument for that single thesis. The migration from individual to distributed responsibility reveals that the unitary moral agent is no longer the basic unit of the medical act. The stratification of access reveals the social form of an artificial evolution that selects its beneficiaries before any biological intervention. The impossibility of arbitrating the conflicts of principle from within the principlist scheme reveals the dissolving ontology that the scheme presupposes. The de-skilling of the clinician reveals the directed evolution of human cognition itself; the virtualisation of the therapeutic relation and the limits of synthetic empathy reveal the transfer of the medical act into a space of symbolic rather than spatial distances. Germline editing reveals the anthropological singularity in its most literal form, and the human species become the object of its own design. The autonomy of the patient must therefore be re-interrogated through the prism of informed consent under conditions of algorithmic opacity, where the technical comprehensibility of intelligent systems enters into conflict with the right to accessible explanation; for it is not sufficient to know that the algorithm computes the best solution, but also which are the processes that generated that prediction.

An augmented bioethics, in the sense in which we have employed the term, need not draw its complementary criteria solely from within the medical-ethical tradition. A convergent set of orientations has emerged, in the contemporary public debate on AI, from the social doctrine of the Catholic Church, where Pope Leo XIV has restated as criteria of discernment for the digital age the inalienable dignity of the human person, the common good, the universal destination of goods, subsidiarity, solidarity and social justice [44]. We hold that these principles—formulated in a register at once philosophical and public—operate, in the medical-sociological reading proposed here, as an external corroboration of our central thesis: that an artificial evolution which is allowed to proceed without a

corresponding ethical and institutional maturation produces not a more capable humanity but, in the words of the same document, an increase in means without a growth in humanity, “having more” without “being more” [44]. The position we have defended is precisely that humanity, in its grandeur and in its woundedness, must neither be replaced nor surpassed by the technological apparatus that serves it, and that the augmentation of bioethics is at bottom the institutional form of that refusal [44].

The future of medical ethics in the age of artificial intelligence depends critically upon our collective capacity to construct effective mechanisms of ethical governance, to implement public policies that democratise the benefits of the technology and so prevent the amplification of existing inequities, and to maintain the profoundly human dimension of the medical act amid its hyper-technologisation. Only an integrative approach—one that balances innovation with precaution, efficiency with equity, technological autonomy with human control—can ensure that the transformative promise of artificial intelligence in medicine is realised in conformity with fundamental ethical values and with respect for the inalienable dignity of the human person. If the anthropological singularity is indeed under way, the task of bioethics is not to arrest it, which it cannot, but to ensure that what emerges on the far side of it remains recognisably, and worthily, human.

Author Contributions: Conceptualization: A.S. and P.P.N.; Methodology: A.S.; Writing—original draft preparation: A.S.; Writing—review and editing: A.S. and P.P.N. All authors have read and agreed to the published version of the manuscript.

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 analysed in this study. Data sharing is not applicable to this article.

Acknowledgments: During the preparation and the subsequent revision of this manuscript, the authors used Anthropic Claude (Opus 4.7) for the purposes of language translation and editing, structural reformulation of draft sections, and bibliographic verification. The authors have reviewed and edited the output and take full responsibility for the content of this publication. The conceptual content, the argumentative structure and the substantive interpretations are those of the authors. This declaration is made not only in satisfaction of the journal’s policy but as an application of the fifth commitment of the augmented bioethics set out in Section 4.6, according to which a bioethics that evaluates algorithmic mediation must submit its own instruments to the same scrutiny and to European Union Ai Act that requires disclosure of using AI in any intellectual product.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sandu, A. Fundamente filosofice ale consimțământului informat [Philosophical Foundations of Informed Consent]. *Rev. Filos.* **2022**, *LXIX*, 715–733.
2. Sandu, A. From the Philosophy of Bat to Nonduality. *J. Study Relig. Ideol.* **2025**, *24*, 16–31.
3. Beauchamp, T.L.; Childress, J.F. *Principles of Biomedical Ethics*, 8th ed.; Oxford University Press: New York, NY, USA, 2019.
4. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies*; Oxford University Press: Oxford, UK, 2014.
5. Kurzweil, R. *The Singularity Is Nearer: When We Merge with AI*; Viking: New York, NY, USA, 2024.
6. O’Lemmon, M. The Technological Singularity as the Emergence of a Collective Consciousness: An Anthropological Perspective. *Bull. Sci. Technol. Soc.* **2020**, *40*, 15–27. [[CrossRef](#)]
7. Veitas, V.; Weinbaum, D. A World of Views: A World of Interacting Post-human Intelligences. *arXiv* **2014**, arXiv:1410.6915.
8. Löw, M. *The Sociology of Space: Materiality, Social Structures, and Action*; Palgrave Macmillan: New York, NY, USA, 2016.
9. Ollinaho, O.I. Virtualization of the Life-World. *Hum. Stud.* **2018**, *41*, 193–209. [[CrossRef](#)]
10. Harris, J. *Enhancing Evolution: The Ethical Case for Making Better People*; Princeton University Press: Princeton, NJ, USA, 2007.

11. Savulescu, J. Procreative Beneficence: Why We Should Select the Best Children. *Bioethics* **2001**, *15*, 413–426. [CrossRef] [PubMed]
12. Jaakkola, E. Designing Conceptual Articles: Four Approaches. *AMS Rev.* **2020**, *10*, 18–26. [CrossRef]
13. Blumer, H. What Is Wrong with Social Theory? *Am. Sociol. Rev.* **1954**, *19*, 3–10. [CrossRef]
14. Sandu, A.; Cojocaru, D.; Oprea, L. Autonomia pacientului în contextul îngrijirilor medicale. In *Relația Medic-Pacient*; Oprea, L., Gavrilovici, C., Vicol, M.C., Astărăstoe, V., Eds.; Polirom: Iași, Romania, 2013; pp. 125–165.
15. Freidson, E. *Profession of Medicine: A Study of the Sociology of Applied Knowledge*; Dodd, Mead & Co.: New York, NY, USA, 1970.
16. Conrad, P. *The Medicalization of Society: On the Transformation of Human Conditions into Treatable Disorders*; Johns Hopkins University Press: Baltimore, MD, USA, 2007.
17. Clarke, A.E.; Shim, J.K.; Mamo, L.; Fosket, J.R.; Fishman, J.R. Biomedicalization: Technoscientific Transformations of Health, Illness, and U.S. Biomedicine. *Am. Sociol. Rev.* **2003**, *68*, 161–194. [CrossRef]
18. Armstrong, D. The Rise of Surveillance Medicine. *Sociol. Health Illn.* **1995**, *17*, 393–404. [CrossRef]
19. Timmermans, S.; Berg, M. *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care*; Temple University Press: Philadelphia, PA, USA, 2003.
20. Nettleton, S. The Emergence of E-Scaped Medicine? *Sociology* **2004**, *38*, 661–679. [CrossRef]
21. Lupton, D. *Digital Health: Critical and Cross-Disciplinary Perspectives*; Routledge: London, UK, 2018.
22. Link, B.G.; Phelan, J. Social Conditions as Fundamental Causes of Disease. *J. Health Soc. Behav.* **1995**, *35*, 80–94. [CrossRef]
23. Metzl, J.M.; Hansen, H. Structural Competency: Theorizing a New Medical Engagement with Stigma and Inequality. *Soc. Sci. Med.* **2014**, *103*, 126–133. [CrossRef] [PubMed]
24. Latour, B. *Reassembling the Social: An Introduction to Actor-Network-Theory*; Oxford University Press: Oxford, UK, 2005.
25. Bowker, G.C.; Star, S.L. *Sorting Things Out: Classification and Its Consequences*; MIT Press: Cambridge, MA, USA, 2000.
26. Winner, L. Do Artifacts Have Politics? *Daedalus* **1980**, *109*, 121–136.
27. Benjamin, R. *Race After Technology: Abolitionist Tools for the New Jim Code*; Polity Press: Cambridge, UK, 2019.
28. Veinot, T.C.; Mitchell, H.; Ancker, J.S. Good Intentions Are Not Enough: How Informatics Interventions Can Worsen Inequality. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1080–1088. [CrossRef] [PubMed]
29. Eubanks, V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*; St. Martin's Press: New York, NY, USA, 2018.
30. Couldry, N.; Mejias, U.A. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*; Stanford University Press: Stanford, CA, USA, 2019.
31. Kleinman, A. *Patients and Healers in the Context of Culture: An Exploration of the Borderland Between Anthropology, Medicine, and Psychiatry*; University of California Press: Berkeley, CA, USA, 1980.
32. Lock, M.; Nguyen, V.-K. *An Anthropology of Biomedicine*; Wiley-Blackwell: Chichester, UK, 2010.
33. Farmer, P. *Pathologies of Power: Health, Human Rights, and the New War on the Poor*; University of California Press: Berkeley, CA, USA, 2003.
34. Rose, N. *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century*; Princeton University Press: Princeton, NJ, USA, 2006.
35. Costa, D.; Serra, R. The Cultural “Blood Ties” Between Medical Anthropology and Cardiovascular Disease: A Narrative Review. *Societies* **2025**, *15*, 97. [CrossRef]
36. Hottois, G. *Le Signe et la Technique: La Philosophie à L'épreuve de la Technique*; Aubier: Paris, France, 1984.
37. Brownsword, R. *Rights, Regulation and the Technological Revolution*; Oxford University Press: Oxford, UK, 2008.
38. Andorno, R. Global Bioethics at UNESCO: In defence of the Universal Declaration on Bioethics and Human Rights. *J. Med. Ethics* **2007**, *33*, 150–154. [CrossRef] [PubMed]
39. Council of Europe. *Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine*; ETS No. 164, Art. 13; Council of Europe: Oviedo, Spain, 1997.
40. Obermeyer, Z.; Powers, B.; Vogeli, C.; Mullainathan, S. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science* **2019**, *366*, 447–453. [CrossRef] [PubMed]
41. European Parliament and Council. Regulation (EU) 2024/1689 of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), as Amended. *Off. J. Eur. Union* **2024**. Available online: <http://data.europa.eu/eli/reg/2024/1689/oj> (accessed on 10 August 2026).
42. Wong, A.; Otles, E.; Donnelly, J.P.; Krumm, A.; McCullough, J.; DeTroyer-Cooley, O.; Pesttrue, J.; Phillips, M.; Konye, J.; Penzo, C.; et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern. Med.* **2021**, *181*, 1065–1070. [CrossRef] [PubMed]
43. Zizioulas, J.D. *Communion and Otherness: Further Studies in Personhood and the Church*; T&T Clark: London, UK, 2006.
44. Leo XIV. *Encyclical Letter Magnifica Humanitas on Safeguarding the Human Person in the Time of Artificial Intelligence*; Libreria Editrice Vaticana: Vatican City, Vatican, 2026. Available online: <https://www.vatican.va/content/leo-xiv/en/encyclicals/documents/20260515-magnifica-humanitas.html> (accessed on 20 June 2026).

45. Mackin, S.; Major, V.J.; Chunara, R.; Newton-Dame, R. Identifying and Mitigating Algorithmic Bias in the Safety Net. *npj Digit. Med.* **2025**, *8*, 335. [CrossRef] [PubMed]
46. Wen, D.; Khan, S.M.; Xu, A.J.; Ibrahim, H.; Smith, L.; Caballero, J.; Zepeda, L.; de Blas Perez, C.; Denniston, A.K.; Liu, X.; et al. Characteristics of Publicly Available Skin Cancer Image Datasets: A Systematic Review. *Lancet Digit. Health* **2022**, *4*, e64–e74. [CrossRef] [PubMed]
47. Daneshjou, R.; Vodrahalli, K.; Novoa, R.A.; Jenkins, M.; Liang, W.; Rotemberg, V.; Ko, J.; Swetter, S.M.; Bailey, E.E.; Gevaert, O.; et al. Disparities in Dermatology AI Performance on a Diverse, Curated Clinical Image Set. *Sci. Adv.* **2022**, *8*, eabq6147. [CrossRef] [PubMed]
48. Morales-Forero, A.; Rueda, L.J.; Herrera, R.; Bassetto, S.; Coatanea, E. Predictive Representativity: Uncovering Racial Bias in AI-based Skin Cancer Detection. *arXiv* **2025**, arXiv:2507.14176.
49. Sjoding, M.W.; Dickson, R.P.; Iwashyna, T.J.; Gay, S.E.; Valley, T.S. Racial Bias in Pulse Oximetry Measurement. *N. Engl. J. Med.* **2020**, *383*, 2477–2478. [CrossRef] [PubMed]
50. Vyas, D.A.; Eisenstein, L.G.; Jones, D.S. Hidden in Plain Sight—Reconsidering the Use of Race Correction in Clinical Algorithms. *N. Engl. J. Med.* **2020**, *383*, 874–882. [CrossRef] [PubMed]
51. Cirillo, D.; Catuara-Solarz, S.; Morey, C.; Guney, E.; Subirats, L.; Mellino, S.; Gigante, A.; Valencia, A.; Rementeria, M.J.; Chadha, A.S.; et al. Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare. *npj Digit. Med.* **2020**, *3*, 81. [CrossRef] [PubMed]
52. Ahtari, M.; Salihu, A.; Muller, O.; Abbé, E.; Clair, C.; Schwarz, J.; Fournier, S. Gender Bias in AI's Perception of Cardiovascular Risk. *J. Med. Internet Res.* **2024**, *26*, e54242. [CrossRef] [PubMed]
53. Krieger, K.; Hameed, I.; Quer, G.; Mack, C.; Savic, M.; Mantaj, P.; Hirofuji, A.; Gregg, A.; Soletti, G.; Rossi, C.S.; et al. Generative pre-trained transformer reinforces historical gender bias in diagnosing women's cardiovascular symptoms. *Eur. Heart J. Digit. Health* **2025**, *7*, ztaf131. [CrossRef] [PubMed]
54. Mittelstadt, B.D.; Allo, P.; Taddeo, M.; Wachter, S.; Floridi, L. The Ethics of Algorithms: Mapping the Debate. *Big Data Soc.* **2016**, *3*, 2053951716679679. [CrossRef]
55. Wachter, S.; Mittelstadt, B.; Floridi, L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *Int. Data Priv. Law* **2017**, *7*, 76–99. [CrossRef]
56. Sheehan, M. Can Broad Consent be Informed Consent? *Public Health Ethics* **2011**, *4*, 226–235. [CrossRef] [PubMed]
57. Hofmann, B. Broadening Consent—and Diluting Ethics? *J. Med. Ethics* **2009**, *35*, 125–129. [CrossRef] [PubMed]
58. Wong, A.; Currey, D.; Schwinne, M.; Park-Egan, B.; Meyer, S.; Gutting, A.; Cao, J.; Khan, S.; Dantes, R.; Pan, T.; et al. Multicenter Prospective Validation of an Updated Proprietary Sepsis Prediction Model. *JAMA Netw. Open* **2026**, *9*, e260181. [CrossRef] [PubMed]
59. Topol, E.J. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nat. Med.* **2019**, *25*, 44–56. [CrossRef] [PubMed]
60. Budzyń, K.; Romańczyk, M.; Kitala, D.; Kołodziej, P.; Bugajski, M.; Adami, H.O.; Blom, J.; Buszkiewicz, M.; Halvorsen, N.G.; Hassan, C.; et al. Endoscopist Deskilling Risk after Exposure to Artificial Intelligence in Colonoscopy: A Multicentre, Observational Study. *Lancet Gastroenterol. Hepatol.* **2025**, *10*, 896–903. [CrossRef] [PubMed]
61. Natali, C.; Marconi, L.; Dias Duran, L.D.; Miglioretti, M.; Cabitza, F. AI-induced Deskilling in Medicine: A Mixed-Method Review and Research Agenda for Healthcare and Beyond. *Artif. Intell. Rev.* **2025**, *58*, 356. [CrossRef]
62. Seitz, L. Artificial Empathy in Healthcare Chatbots: Does It Feel Authentic? *Computers in Human Behavior: Artificial Humans. Comput. Hum. Behav. Artif. Hum.* **2024**, *2*, 100067. [CrossRef]
63. Ajeesh, K.G.; Joseph, J. The Compassion Illusion: Can Artificial Empathy Ever Be Emotionally Authentic? *Front. Psychol.* **2025**, *16*, 1723149. [CrossRef] [PubMed]
64. Yannaras, C. *The Freedom of Morality*; St Vladimir's Seminary Press: Crestwood, NY, USA, 1984.
65. European Parliament and Council. Regulation (EU) 2026/1744 of 8 July 2026 Amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as Regards the Simplification of the Implementation of Harmonised Rules on Artificial Intelligence (Digital Omnibus on AI). *Off. J. Eur. Union* **2026**. Available online: <http://data.europa.eu/eli/reg/2026/1744/oj> (accessed on 10 August 2026).
66. Greely, H.T. CRISPR'd Babies: Human Germline Genome Editing in the 'He Jiankui Affair'. *J. Law Biosci.* **2019**, *6*, 111–183. [CrossRef] [PubMed]
67. Serra, M.-À. Human Enhancement and Functional Diversity: Ethical Concerns of Emerging Technologies and Transhumanism. *Mètode Sci. Stud. J.* **2022**, *12*, 169–175. [CrossRef]
68. Ranisch, R. When CRISPR Meets Fantasy: Transhumanism and the Military in the Age of Gene Editing. In *Transhumanism: The Proper Guide to a Posthuman Condition or a Dangerous Idea?*; Hofkirchner, W., Kreowski, H.-J., Eds.; Springer: Cham, Switzerland, 2020.

69. Rueda, J. Genetic Enhancement, Human Extinction, and the Best Interests of Posthumanity. *Bioethics* **2024**, *38*, 96–106. [[CrossRef](#)] [[PubMed](#)]
70. Hart, D.B.; Chrysavgis, J. (Eds.) *For the Life of the World: Toward a Social Ethos of the Orthodox Church*; Holy Cross Orthodox Press: Brookline, MA, USA, 2020.
71. Engelhardt, H.T., Jr. *The Foundations of Christian Bioethics*; Swets & Zeitlinger: Lisse, The Netherlands, 2000.
72. Østergaard, S.D. Will Generative Artificial Intelligence Chatbots Generate Delusions in Individuals Prone to Psychosis? *Schizophr. Bull.* **2023**, *49*, 1418–1419. [[CrossRef](#)] [[PubMed](#)]
73. Pierre, J.M.; Gaeta, B.; Raghavan, G.; Sarma, K.V. “You’re Not Crazy”: A Case of New-onset AI-associated Psychosis. *Innov. Clin. Neurosci.* **2025**, *22*, 11–13. [[PubMed](#)]
74. Flathers, M.; Roux, S.; Torous, J. Beyond Artificial Intelligence Psychosis: A Functional Typology of Large Language Model-Associated Psychotic Phenomena. *Lancet Digit. Health* **2026**, *8*, 100974. [[CrossRef](#)] [[PubMed](#)]
75. Chung, V.H.A.; Bernier, P.; Hudon, A. Mass Media Narratives of Psychiatric Adverse Events Associated with Generative AI Chatbots: Rapid Scoping Review. *JMIR Ment. Health* **2026**, *13*, e93040. [[CrossRef](#)] [[PubMed](#)]
76. Proudfoot, A. Could a Conscious Machine Deliver Pastoral Care? *Stud. Christ. Ethics* **2023**, *36*, 675–693. [[CrossRef](#)]
77. Sandu, A. Autonomia în era digitală: Perspective etice și juridice asupra virtualizării existenței umane. In *Bioetica Finalului Vieții*; Ioan, B.G., Toader, E., Curcă, G.C., Eds.; Junimea: Iași, Romania, 2025; pp. 171–184.
78. Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds Mach.* **2018**, *28*, 689–707. [[CrossRef](#)] [[PubMed](#)]
79. Mittelstadt, B. Principles Alone Cannot Guarantee Ethical AI. *Nat. Mach. Intell.* **2019**, *1*, 501–507. [[CrossRef](#)]
80. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models*; WHO: Geneva, Switzerland, 2024.
81. Thompson, D.F. Moral Responsibility of Public Officials: The Problem of Many Hands. *Am. Political Sci. Rev.* **1980**, *74*, 905–916. [[CrossRef](#)]
82. Matthias, A. The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics Inf. Technol.* **2004**, *6*, 175–183. [[CrossRef](#)]
83. Girard, S.; Psimopoulos, C. AI and Healthcare: Orthodox Christian Theological Considerations. *Public Orthodoxy*, 29 June 2026. Available online: <https://publicorthodoxy.org/2026/06/29/ai-and-healthcare/> (accessed on 15 August 2026).
84. Ghaly, M. (Ed.) *Islamic Ethics and the Genome Question*; Brill: Leiden, The Netherlands, 2019.
85. Crawford, S.C. *Hindu Bioethics for the Twenty-First Century*; State University of New York Press: Albany, NY, USA, 2003.
86. Dorff, E.N. *Matters of Life and Death: A Jewish Approach to Modern Medical Ethics*; Jewish Publication Society: Philadelphia, PA, USA, 1998.
87. Ali, F.; Bouzoubaa, K.; Gelli, F.; Hamzi, B.; Khan, S. Islamic Ethics and AI: An Evaluation of Existing Approaches to AI using Trusteeship Ethics. *Philos. Technol.* **2025**, *38*, 120. [[CrossRef](#)]
88. Hongladarom, S. *The Ethics of AI and Robotics: A Buddhist Viewpoint*; Lexington Books: Lanham, MD, USA, 2020.

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.