CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

More Process, Less Principles: The Ethics of Deploying AI and Robotics in Medicine

Amitabha Palmer[1]* and David Schwan[2]

[1]University of Texas, MD Anderson Cancer Center, Houston, Texas, USA
[2]Department of Philosophy and Comparative Religion, Central Washington University, Ellensburg, Washington, USA
*Corresponding author. Email: apalmer1@mdanderson.org

Abstract
Current national and international guidelines for the ethical design and development of artificial intelligence (AI) and robotics emphasize ethical theory. Various governing and advisory bodies have generated sets of broad ethical principles, which institutional decisionmakers are encouraged to apply to particular practical decisions. Although much of this literature examines the ethics of designing and developing AI and robotics, medical institutions typically must make purchase and deployment decisions about technologies that have already been designed and developed. The primary problem facing medical institutions is not one of ethical design but of ethical deployment. The purpose of this paper is to develop a practical model by which medical institutions may make ethical deployment decisions about ready-made advanced technologies. Our slogan is "more process, less principles." Ethically sound decisionmaking requires that the process by which medical institutions make such decisions include participatory, deliberative, and conservative elements. We argue that our model preserves the strengths of existing frameworks, avoids their shortcomings, and delivers its own moral, practical, and epistemic advantages.

Keywords: artificial intelligence; robotics; ethics; deployment; participatory democracy

Introduction

Imagine that a group of physicians and administrators at a local nursing home is considering purchasing several social robots (henceforth "carebots") from Robocorp. Among other things, these carebots will take over menial tasks but also provide companionship and interactive engagement with residents. Administrators point to the safety, economic, and efficiency benefits. However, some staff have raised concerns given the profound impact that this technology may have on the care environment and relationships between caregivers and patients. They worry that such technologies reduce opportunities for caring interactions, compassionate listening, and human touch—all of which are at the core of good care. Should the nursing home purchase and deploy these carebots? How should they decide? If they decide to do so, how can they ensure that this technology is deployed ethically?

Given their potential to significantly alter foundational aspects of human life, discussions regarding the ethics of artificial intelligence (AI) and robotic technology have increased substantially over the past two decades.[1] As noted in the United Nation's Resource Guide on AI Strategies:

> "AI-based technologies blur the boundary between human subjects and technological objects [and] not only have societal implications [...] but they also affect the central categories of ethics: our concepts of agency and responsibility, and our value frameworks."[2]

Within medicine, a growing literature explores the ethical dimension of how such technologies can be designed[3] and applied[4] in a variety of care contexts like nursing[5] or eldercare.[6] Researchers explore how these technologies impact values such as patient autonomy, dignity, welfare, nonmaleficence, privacy, safety, transparency, human capabilities (e.g., bodily integrity, bodily health, control over one's environment), social isolation, and the care relationship.[7]

Although much of this literature examines the ethics of designing and developing AI and robotics, the majority of medical institutions in the United States must make purchase and deployment decisions about technologies that have already been designed and developed.[8] Given the frequent amalgamation of AI and robotics systems, we jointly refer to these overlapping technologies as "advanced technologies."[9] A central practical problem facing medical institutions is not one of ethical design but of ethical deployment. The purpose of this paper is to develop a practical model by which medical institutions may make ethical deployment decisions about ready-made advanced technologies.

Currently, a variety of governmental, industrial, and scientific organizations have drafted principles and ethics guidelines for developers of AI and robotic technologies.[10] Although there appears to be some convergence on broad ethical principles (e.g., respect for autonomy, harm prevention, fairness, and explicability),[11] critics have argued that this apparent agreement at the general level "obscures deep political and normative disagreement."[12] Furthermore, the available evidence suggests that such guidelines do not influence the behavior of professionals developing these technologies.[13]

Despite these problems, the continued articulation and evaluation of broad ethical frameworks by the government and industries is important in the ethical development of advanced technologies. Some guidance and regulation are likely better than none. Our proposal is consistent with the continued need for value-sensitive design as well as research on systems and structures that incentivize ethical design in the engineering of advanced technologies.[14] Nevertheless, we also share the additional practical concern noted by Mark Coeckelbergh in his recent analysis of AI policy guidelines:

> [I]t remains a huge challenge to build a bridge between, on the one hand, abstract, high-level ethical and legal principles and, on the other hand, the practices of technology development and use in particular contexts, the technologies, and the voices of those who are part of these practices and work in these contexts. This bridging work is left to the addressees of the proposals. Can and should more be done, at the earlier stage of policymaking? At the very least, more work on the "how" is required alongside the "what": the methods, procedures, and institutions we need for making AI ethics work in practice. *We need to pay more attention to process.*[15]

One response might be to argue that concerns about ethical deployment can be addressed through ethical design. It is true that design is a value-laden activity and that design explicitly and implicitly builds in values. However, building in values assumes that a particular set and ranking of values can be applied universally across all contexts in which advanced technology will be applied. As we discuss later in the paper, this is unlikely to be the case.

Taking our cue from Coeckelbergh, we aim to achieve two objectives: (1) Identify shortcomings in current decisionmaking processes regarding the deployment of advanced technologies in medicine; and (2) Develop a decisionmaking process for the ethical deployment of advanced technology in medicine that functions in a wider variety of value contexts. Our slogan is "more process, less principles." We call our model *Participatory Deliberative Conservatism* (PDC).

This model has four central features that we believe are necessary for ensuring that deployment decisions and the processes by which they are generated are ethical. First, unlike current generalist approaches, which purport to apply across all domains of human activity, PDC is *domain-specific* to medicine. Medical practice involves its own unique values, traditions, and goals.[16] We will argue that an adequate ethical model for the deployment of advanced technology must be sensitive to the unique normative and teleological features of this practice. Second, our model is fundamentally *participatory* in that, in addition to healthcare workers and administrators, it includes local and lay stakeholders such as patients and caregivers. Third, our model is *deliberative.* The deliberative element ensures that abstract values and stakeholder concerns are appropriately understood, weighed, and applied in their local

concrete contexts. Finally, our approach retains a *conservative* element since the outcome of this participatory deliberative process is not overriding but constrained by established legal, regulatory, and medical codes of ethics. Although patient values and concerns inform deployment, the moral and legal responsibility for deployment (and its consequences) rests with medical practitioners and administrators.

In the next section, we evaluate prominent ethical models and decisionmaking processes for the deployment of advanced technologies in medicine. We then present our alternative model, *PDC*, and explain how it overcomes challenges to existing views and has additional practical, moral, and epistemic virtues over its competitors.

Principles and Processes: Two Distinctions

Current approaches to the design, development, and deployment of advanced technologies favor *ethical principlism*, that is, the idea that a limited set of normative principles and values appropriately governs specific decisions.[17] Ethical principlism comes in two general forms: generalist and domain-specific. *Generalist* approaches aim to establish a set of ethical principles that ought to govern decisions about advanced technologies across many distinct domains of practice. *Domain-specific* approaches aim to establish a set of ethical principles appropriate to specific domains of human practice.

Similarly, the processes by which institutions make decisions about design, development, and deployment can be divided into two broad approaches: *top-down proceduralism* and *bottom-up proceduralism*. Although they exist on a continuum, *top-down proceduralism* favors investing decisionmaking power in those who hold higher positions of formal institutional authority in an organization. *Bottom-up proceduralism* accords more decisionmaking power to stakeholders affected by the relevant policies and practices regardless of their formal institutional authority. With these distinctions in mind, we briefly sketch the strengths and weaknesses of these common approaches to the ethics of advanced technologies.

Principlist Approaches to the Ethics of Advanced Technologies

Generalist Approaches

The most common approach to addressing the ethics of advanced technologies involves formulating and applying a set of broad ethical principles to guide their design and development.[18] These principles are intended to apply across all human domains of activity. For example, consider the recent proposal from the European Commission's Expert Group on AI (AIHLEG). In their *Ethical Guidelines for Trustworthy AI*,[19] they argue that AI should respect human autonomy, minimize harm, be fair, and be explicable "throughout the system's entire life cycle" and that these principles are presented "without a hierarchy."[20] These general principles are translated into concrete requirements involving human agency and oversight, technical robustness and safety, privacy, transparency, diversity and fairness, societal/environmental well-being, and accountability. The AIHLEG notes that these concrete requirements will be applied by developers and deployers, with end users requesting that they are properly upheld.

Although generalist principlisms are useful for foregrounding broad ethical values and concerns, they present theoretical and practical challenges. Our point in this section is not that generalist principlisms are irredeemably flawed but rather that current models are incomplete with respect to concrete deployment processes. We outline three challenges with these approaches, many of which have also been variously articulated by other authors: the ambiguity challenge, the ranking challenge, and the tacit exclusion challenge. We discuss each in turn.[21]

The *ambiguity challenge* arises as a consequence of the abstract nature of the ethical concepts applied to concrete situations by diverse populations. The concept of justice, for example, contains multiple conceptions like distributive, egalitarian, desert-based, social, and equity-based accounts. Within each account, there are further divisions. For example, not all egalitarians agree on the nature of equality or in what respects people ought to be equal. The ambiguity of the concept lends itself to being differently understood by different people and within different contexts.

To illustrate the ambiguity challenge in the context of AI, consider the recommended principles from the Department of Defense (DoD) on the "Ethical Use of Artificial Intelligence by the Department of Defense." They argue that defense AI should be responsible, equitable, traceable, reliable, and governable.[22]

These resemble other generalist ethical frameworks for AI, but a closer examination reveals how differently the terms are understood in specific cases. Although "equitable" in this context suggests that the DoD should "avoid unintended bias…that would inadvertently cause harm to persons," they also note that their understanding of this term does not follow the standard concept of "fairness" as it is "cited in the AI community."[23] This is because from the perspective of military engagement (within the relevant norms of warfare), "fights should not be fair, as DoD aims to create the conditions to maintain an unfair advantage over any potential adversaries."[24] Effective military action may be lethal and involve inflicting deliberate harm on targets or justify deception and deceit in ways that would be wrong in other contexts.

The ambiguity challenge raises further questions about justification: Whose understanding of the principles ought to apply and why? But this is also a practical problem. The practical needs of policy-makers require that, if they want a new technology to promote or preserve some set of values, they need to understand those values in a sufficiently precise, naturalistic, and implementable way.[25] Principlist models, however, offer nonhierarchical sets of abstract principles that admit multiple interpretations. Such approaches can fail to provide clear context-sensitive and therefore action-guiding interpretations of values and principles. Furthermore, even if they could provide context-sensitive action-guiding formulations of values, they do not provide rankings.

The *ranking challenge* occurs any time two or more values in a nonhierarchical set conflict.[26] For example, existing generalist models do not provide an account of what would justify prioritizing autonomy above beneficence in one context but beneficence above autonomy in another. Principled resolutions of value conflicts across cases require a metatheory for resolving conflicts, which current models lack. Thus, decisionmakers appealing only to generalist models lack clear methods for determining and justifying normative priorities in particular cases. Critically, both the ranking and ambiguity challenges leave unresolved the important moral and practical questions regarding *whose* normative interpretations and rankings govern decisions.

Finally, it is unlikely that any concise list can capture all values relevant to all situations. The *tacit exclusion challenge* occurs when important values are not given their due because they are not specified by a particular generalist model. Proponents of such models might reply that specified values should be considered necessary but not sufficient ethical criteria. Nevertheless, by formulating a limited list of general principles, policymakers risk attending less to unlisted values that may be salient in particular situations. For example, no current generalist model contains the value of caring. However, it would be a mistake not to give caring high priority when considering the effects of deploying carebots in nursing homes. Models that do not explicitly enumerate a value tacitly allow decisionmakers to either omit or diminish the weight of that value—especially if it conflicts with enumerated values.

The exclusion problem also stems from the fact that lists of values have been developed primarily by academics and policymakers who are likely distant from the local contexts in which decisions will be made.[27] Hence, local and lay values or values relevant only to particular contexts risk being underappreciated in decisions. Furthermore, as we will discuss later, the tacit exclusion problem is amplified when combined with top-down proceduralism.

Our point, once again, is not that generalist principlisms are inherently flawed or that principlists cannot respond to the three challenges but rather that ethical deployment requires that these challenges be addressed. Generalist approaches to the ethics of AI are important for clarifying relevant ethical values and identifying potential stakeholders. They also serve as accessible heuristics for decisionmakers seeking guidance on complex problems. Further, the AIHLEG has developed these broad principles (and concrete considerations) into an operationalized framework for addressing each of the ethical areas of concern (e.g., data privacy) and they continue to seek feedback from users in specific areas of industry or society to clarify this framework. However, they also rightly note that "the implementation of the [guidelines] needs to be adapted" to particular contexts and that "the necessity of an additional sectoral

approach, to complement the more general [framework]…should be explored."[28] We agree. As we discuss below, there are a variety of reasons to favor a domain-specific approach.

Domain-Specific Approaches to Advanced Technologies

As far back as Aristotle, thinkers and policymakers have recognized that particular domains of human activity have their own goods, which are set by the *telos* of that practice.[29] The goods of the art/craft (*techne*) of shipbuilding, for example, depend on what it is for a ship to be good. Similarly, the goods of medical practice are set by the preservation and promotion of health. Hence, *contra* the generalist principlist approach to the ethics of advanced technologies, one might advocate for a domain-specific approach.

On this view, the set of normative values that govern the ethical implementation of advanced technologies is fixed by the goals and values internal to a particular domain of practice. Medicine, as a distinct domain of practice with an identifiable *telos*, appears to be well suited to this approach.[30] This is evidenced by the fact that domain-specific principlism has come to dominate ethical analysis in medicine.[31] For our purposes, we will follow the consensus view.

A growing body of research takes a domain-specific approach to the ethical dimensions of introducing advanced technologies into a variety of clinical settings. Some literature examines caregiving broadly,[32] whereas others focus on specific issues within subdomains like eldercare,[33] or the potential impact of technology on specific practices like nursing.[34]

This raises an ontological question regarding the nature and scope of different subdomains of medicine and the ways in which they are related. For example, one might argue that psychiatry, surgical oncology, palliative care, and eldercare are substantively distinct and require governance by different ethical norms and values. There are two claims worth noting here. First, these subdomains are similar in that they all aim at the same *telos*, that is, they all seek to protect and promote health. As such, we should expect that the same broad normative approach governs these activities. Second, these subdomains are different in that they contribute to the *telos* in distinct ways. For example, following Aristotle's analogy, although the specific end of sail-making (e.g., to build sturdy sails that capture wind well) or bow-making (e.g., to build sturdy bows that break waves) may be distinct, both are subdomains of shipbuilding because they serve the general *telos* of shipbuilding (i.e., sturdy, swift ship construction).[35]

A strength of the domain-specific approach is that, unlike the generalist approach, it more clearly specifies the contextual features of the relevant clinical domain and helps illuminate some of the specific goods and values that are inherent in that discipline's practices.[36] For example, human contact is vital in many areas of medical practice and is particularly important in some subdomains like long-term eldercare where issues of loneliness may be more salient. As such, introducing robotic technologies into this particular domain may produce distinct challenges and opportunities for achieving a range of relevant goods.[37] Further, it is possible that introducing new technologies and practices may even *alter* or displace the nature of the goods that we want to promote.[38]

The key insight of the domain-specific approach is that it is essential to attend to a practice's specific goals and contextual features when evaluating the permissibility of employing new technologies in medicine. However, when domain-specific approaches are principlist, they inherit the same three challenges as generalist principlism: the ambiguity challenge, the ranking challenge, and the tacit exclusion challenge. Even if we narrow the scope of discussion to subdomains within medicine, the use of abstract ethical concepts may still suffer from problems of ambiguity and variance across contexts. Further, like generalist perspectives, domain-specific approaches require a non-*ad hoc* basis for ranking relevant values and avoid excluding relevant values of stakeholders impacted by the deployment of advanced technologies.

Two Forms of Proceduralism

In addition to ethical frameworks, decisions regarding the deployment of advanced technology may be addressed with two kinds of procedures that exist on a continuum. At one end, *top-down proceduralism* rests decisionmaking power in those who occupy the upper echelons of an institution or group's formal

hierarchy, whereas *bottom-up proceduralism* takes a broader view of stakeholders and derives decisions from those further down a hierarchy.

Top-down proceduralism exemplifies current institutional approaches to decisions regarding the deployment of advanced technologies. The process often begins when a group of physicians or administrators become aware of a new technology. Any decisions regarding deployment must satisfy, at minimum, institutional procurement, safety, quality, and data security oversight committees. These institutional bodies are themselves constrained by federal and state regulatory frameworks. Satisfying these various oversight groups often requires running a pilot study to demonstrate that the technology meets the various institutional requirements and demands.

Medical institutions are driven by innovation and patient outcomes, but as scholars have noted this "needs to be achieved within limited budgets" and "deciding which [technologies] will deliver clinical and cost advantages is fraught with difficulty."[39] Given that institutional actors make purchasing decisions, many new medical technologies are developed primarily with narrow institutional values in mind, that is, economic, efficiency, and safety benefits. However, these are obviously not the only values relevant to medicine. Hence, the ethical design and deployment of advanced technologies requires decisionmaking processes that identify and weigh all values relevant to those meaningfully affected by the new technology. As such, relevant stakeholders—including patients or patient groups—must be involved in the decisions about whether to adopt and how to deploy new technology.

Defenders of current practices might reply that hospital administrators and physicians can represent and advocate for patient interests. However, without substantive patient representation in top-down models, it is less likely that patient values and concerns can be accurately identified and ranked. Even if they can, top-down proceduralism raises concerns regarding whether they will receive sufficient weight —especially when they conflict with institutional and physician interests. Insufficient patient representation thus raises both ethical and epistemic concerns, both of which may undermine patient-centered care—the purported lodestar of modern medical practice.

Modern medicine is characterized by a movement toward greater patient-centered care and shared decisionmaking. On this model, healthcare providers are "encouraged to partner with patients to co-design and deliver *personalized* care."[40] Yet, when it comes to having a voice in the wider technological and institutional structure of the medical environment, patients typically have much less input. This means that the nature of the care environment and the choice architecture in which patients find themselves can emphasize the values and desires of physicians and hospital administrators. Insofar as institutional decisions to deploy technology aim to represent patient values and desires, top-down processes imply that these values and desires are filtered through the (not impartial) imaginations of physicians and administrators. Although they may sometimes converge, a genuine commitment to patient-centered care would include patient populations in decisionmaking to ensure that their contextually situated values and desires are accurately represented.

Relatedly, there is a consensus commitment to shared decisionmaking in medicine, and yet this practice is largely absent in decisions that reshape care and care environments with advanced technology. Fundamental to the shared decisionmaking model is the recognition of epistemic asymmetries between participants and, hence, the value of deliberation among stakeholders. Patients possess important knowledge that physicians lack, and vice versa. Coherent integration of this disparate knowledge requires a deliberative process among interested actors.

We suggest extending the existing commitment to shared decisionmaking to the deployment of advanced technologies in care environments. In short, decisions about whether and how to deploy advanced technology in medical institutions should be the outcome of a deliberative process among diverse stakeholders that include patient groups and/or their advocates, healthcare workers, and administrators. As such, we propose a model that is consistent with the ideals of patient-centered care and shared decisionmaking for decisions to deploy advanced technologies in medicine.

Participatory Deliberative Conservatism

In this section, we present our approach to decisionmaking about the deployment of advanced technologies in medicine. We call it *PDC.* In what follows, we demonstrate that PDC preserves the strengths of existing frameworks, avoids their shortcomings, and delivers its own moral, practical, and epistemic advantages. Importantly, our example below is only one of many possible instantiations of a decision process that incorporates the relevant elements of PDC. Medical institutions, depending on their size and scope of practice, may develop decision processes that instantiate participatory, deliberative, and conservative elements differently.

To illustrate this model, consider the case with which we began: An assisted living facility is considering introducing carebots. Employing these robots will likely be cheaper (in the long run) and perform certain tasks more efficiently than human nurses. Should this facility purchase these carebots? How should they decide? And if they decide to purchase them, how should they determine the way in which the carebots are deployed?

Step 1: Preliminary Identification of Values and Concerns

The goal of the first step is to generate a list of values and concerns from a small group of primary stakeholders. This is the work of the *pilot committee*, a group composed of diverse primary stakeholders that include nurses, administrators, caregivers, physicians (from various disciplines), patients, and/or former patients. In addition, the pilot committee should include representatives from the development team of the relevant advanced technologies.[41] This first step provides an initial account of any practical and normative considerations relevant to whether and how to deploy the new technology. Practical considerations are primarily pragmatic, like whether gloved hands can use a touchscreen on the carebot. Normative concerns involve value-laden features of concrete situations. For example, rather than expressing abstract concerns for privacy, patients might articulate apprehension about whether a carebot's presence violates their sense of privacy or whether it enhances it.[42]

Since the pilot committee is small, its conclusions risk being unrepresentative of the various stakeholder subgroups. Therefore, through appropriate interview and survey techniques, the pilot committee compiles information from a larger set of each of the primary stakeholder groups.[43] Possible survey techniques include interviews (structured, semi-structured, and unstructured), focus groups, panel sampling, surveys (telephone, mail-in, kiosk, and online), and other sociological methods. The particular methods employed will be determined by the properties of the issue, the target population, and their environment. Because so much will be context-sensitive, we do not advocate for any particular information-gathering method but rather suggest that the context will determine the methods. For example, methods used to gather information from an unhoused population will be different from those used to gather information from institutionalized populations. Appropriately drawing from larger samples increases the likelihood that the values and concerns relevant to deployment decisions adequately represent those of all stakeholder groups.

Step 2: Deliberation About Values, Concerns, and Ranking

Clinical situations are "saturated with values, obligations, responsibilities, character traits [and] virtues."[44] But, in particular situations, concrete normative concerns often represent more general values. The purpose of the second step is for stakeholders to deliberate about their collective contextually understood concerns, identify the underlying values that motivate them, and generate a provisional ordering of these values. Unlike a top-down principlist approach that begins with a list of abstract values to be imposed on the adoption decision, an ordered list of relevant values emerges organically from stakeholders' deliberations about contextually situated concerns.

For example, in Step 1, nurses may have expressed concerns that carebots will replace valuable face-to-face time with patients. Nurses are articulating a context-specific instantiation of the more general value of caring. The deliberative process allows stakeholders to identify the more general values that

underlie their concrete concerns. This accomplishes two important desiderata for decisions about the ethical deployment of advanced technologies.

First, it provides clarity since initially stakeholder concerns are not always clearly articulated, understood, or consistent with other normative concerns. Deliberative processes allow stakeholders to clarify both for themselves and to others the precise nature of their concerns.[45]

Second, the group identifies the underlying general values motivating their particular concerns. Values may be instantiated in multiple ways in concrete situations. When people are committed to a particular instantiation of a value (e.g., not enough face-to-face time) rather than the value that underlies that commitment (e.g., caring), reconciling disagreement can be difficult. However, by focusing deliberation on shared underlying values, deliberators can more easily rank those values and generate conciliatory policies when they understand that the same value may be instantiated in multiple ways. In other words, identifying and focusing on underlying values provides more avenues for conflict resolution.[46]

Step 3: Formulation of Initial Deployment Policy

In Step 3, the pilot committee decides whether to adopt the new technology and, if so, how. Consider again a nursing home's decision about whether to adopt carebots. In Step 2, the committee may have identified the following set of values: caring, privacy, autonomy, independence, safety, economic cost, and efficiency. The committee must evaluate whether introducing carebots protects and promotes these values better than current arrangements. Such a decision requires evaluating trade-offs between the various values. Importantly, decisions on whether and how to adopt will depend on contextual features and value rankings as they are locally understood.

If the pilot committee decides to implement the new technology, they generate, via a deliberative process, a means of deployment that satisfactorily protects and promotes the values from Step 2. In addition, the committee will develop success criteria, assessment tools by which implementation is measured, and the intervals at which they will be evaluated. The deliberative process is creative in that it generates novel ways of representing in policy the values identified in earlier steps. Stakeholders engage in the deliberative process to discover models of deployment that mutually satisfy diverse stakeholder concerns.

Although the process is participatory, it is not populist. Patient values and concerns from Step 2 are not *de facto* overriding. The final authority and responsibility regarding deployment (and its consequences) rests on medical practitioners and administrators, not patients. This is because, as we noted, sole reliance on patient preferences to guide policy may lead to practices that undermine important established medical values or legal frameworks, provide only marginal medical benefit, or insufficiently weigh local economic considerations.

Step 3, therefore, involves this model's conservative element. Since this model is specific to the domain of medicine, it aims to preserve and advance the values intrinsic to medical practice to which medical professionals are bound.[47] Further, it leverages important epistemic asymmetries between lay people, medical experts, and administrators. For example, patients know best their own medical needs, medical providers have special access to training and medical protocols to satisfy those needs, whereas administrators often have better access to economic and relevant legal/regulatory frameworks.

Step 4: Deployment and Iterative Evaluation of the New Technology

The purpose of Step 4 is to deploy the new technology and perform iterative evaluations of how well it protects and promotes the relevant values and concerns from Steps 2 and 3.[48] The pilot committee: (1) deploys the new technology according to the previously established goals and values; (2) monitors the deployed technology for unintended consequences; (3) meets regularly to assess and deliberate on how well the new technology achieves the various goals and objectives set out in Step 3; and (4) makes adjustments to the deployment policy commensurate with the assessments.

The Virtues of Deliberative Participatory Conservatism

Practical Virtues: Specifies Meaning of Values in a Local, Contextually Specific Way

Recall that most existing ethical frameworks for advanced technologies face the *ambiguity challenge* because they are principlist. This challenge arises due to the abstract nature of ethical concepts and how they are variously understood by stakeholders in different contexts. PDC's participatory and deliberative elements, however, ensure that stakeholder values and concerns are understood and applied in ways that reflect local cultural, economic, and social values and are sensitive to concrete contextual features.

Existing frameworks also face the *ranking challenge* which occurs any time two or more values in a nonhierarchical set conflict and the model contains no guidance or mechanism for conflict resolution. The PDC model generates rankings when diverse stakeholders engage in a deliberative process (Steps 2–4). We elaborate on the ethical significance of this point below. The point here is that, unlike existing principlist frameworks, PDC provides a non-*ad hoc* process by which values are ordered and conflicts between them are reconciled.

Moral Virtues: Justification of Values and Value Orderings

There is a tension between the ability to be sensitive to the normative richness of medicine and the practical need to narrow the scope of normative concerns. Approaches that insist on considering all and every normative value are unwieldy and inappropriate for the practical needs of institutional decision-makers, hence the popularity of principlist models. Approaches that narrow the scope of relevant normative considerations, such as some forms of ethical principlism, risk excluding important values in particular contexts—especially when the relevant values are imposed *a priori.* Hence, any approach to normative decisionmaking must balance sensitivity to the normative richness of medicine with the practical need for a limited and tractable set of normative concerns.

Any model of normative decisionmaking must justify why some normative considerations are included and others are not. Similarly, it must justify why some of the included values are weighted more heavily than others. PDC provides this justification since the relevant values for each decision emerge as a result of deliberation between those affected by and responsible for the new technology.

Those who experience the direct effects of a new technology ought to have a say in how it is applied to them. Unlike standard institutional approaches to deployment decisions, PDC includes patient voices. Doing so not only serves to justify value rankings but also is consistent with patient-centered care. Administrative/top-down policymaking risks failing to *respect* the very persons whose medical situation it is. It is *their* circumstances, *their* problems, *their* values, and *their* lives that are at stake.[49] Nevertheless, administration and medical professionals need to be included, too, since they also experience and bear responsibility for a new technology's effects, hence the conservatism in our approach.

The deliberative process serves to balance the otherwise narrow concerns of each constituent group. For example, there is long-standing consternation that if deployment decisions rest primarily with hospital administrators, economic concerns will be overweighted.[50] Patients and long-term caregivers, therefore, must have a voice. However, sole reliance on patient values and concerns can disregard genuine economic considerations and appropriate medical practice.[51] After all, as a goal-oriented human practice, medicine already aims at certain goods and values which constrain the values and rankings of this participatory process. Hence, physicians and nurses too must have a voice to ensure that the technology is medically beneficial, conforms with good practice, and does not displace practices containing goods internal to them such as caring, touch, and humanistic interactions. We view this model as analogous to a constitutional democracy. The public subject to laws and policies has a say in what those policies are, but their policy preferences are constrained by a preexisting and more fundamental legal and normative framework.

The deliberative and participatory elements of our model allow it to be sensitive to the local cultural context in which deployment decisions are made. For example, a decision whether or how to deploy carebots for eldercare will look different in rural Thailand compared with New York City. These differences in outcomes are a function of the model's sensitivity to local economic conditions and social,

cultural, religious, and political values. This sensitivity to these broad contextual features enhances the justificatory power of the deliberative outcomes in ways that models without a deliberative process cannot.

Epistemic Virtues: Complexity and Interaction

In addition to these practical and moral virtues, PDC also overcomes numerous epistemological challenges arising in decisions to deploy advanced technologies in medical contexts. One central problem with medical and hospital ecosystems involves their sheer variability and complexity. Even within a single hospital, there will be numerous divisions, each with distinct priorities, specialists, administrative staff, and technological needs. The complexity of gathering and factoring the preceding variables in decisionmaking is amplified when taking into account further contextual features within which they are embedded such as geographic location, specialization and availability of resources, and technology.

It is nothing new for medical institutions to introduce new technologies. For example, a hospital might introduce new bandages with advances in materials science. In these familiar kinds of cases, new introductions often serve as functional equivalents of whatever they are replacing. However, the institutional effects of these sorts of technologies are qualitatively different than those emerging from advanced technologies. The introduction of the latter often comes with unexpected complex down-stream effects on institutional structures, values, and human-to-human interactions.

For example, some deployment decisions may require information about practical concerns. This might include hiring technicians and providing relevant training to the staff. Others involve complex relationships between hospital divisions and healthcare workers. Machine learning systems in radiology, for example, may alter staffing and internal communication patterns.[52] In other cases, introducing a new technology can inadvertently eliminate or displace a good that is internal to a particular medical practice.[53] For example, suppose a hospital deploys a new technology that allows immobile patients to feed themselves.[54] In doing so, the technology displaces the nurse who otherwise would have fed the patient. Both cases are functionally equivalent in that the patient gets fed. However, "caring" is inadvertently removed along with the nurse's displacement. In short, advanced technologies can eliminate or displace goods and values internal to human practices when those technologies eliminate or displace the humans involved in those practices.

The epistemological challenges of introducing advanced technologies into complex human institutions present two challenges for generalist principlist and "top-down" decisionmaking models. The first problem relates to the way principlism can tacitly exclude important values. As we have argued, certain forms of principlism attempt to identify a limited set of values *a priori* and apply them to situations as they arise; however, following widely held views in clinical ethics, we believe that it is "not possible to know in advance, beyond common themes, just which moral issues are actually presented by any specific situation."[55]

Unlike generalist principlism, on the PDC model, the relevant values and concerns that inform a particular medical situation are generated (in Step 2) through discussion and deliberation with diverse groups of stakeholders who will have direct experience with the technology's impacts on the care setting and treatments. The composition of the decisionmaking body influences which values guide deployment policy. People are most acutely aware of the normative concerns that most directly impact their interests and lived experiences. By including a variety of stakeholders, the PDC model answers the question, "which are the values relevant to whether and how we deploy this technology?" For this reason, decision processes which draw from diverse stakeholder perspectives hold epistemic advantages over those that rely primarily on administrators or physicians.

The second problem relates to the epistemic disadvantages of employing top-down decisionmaking when introducing a new advanced technology into a complex and dynamic human practice. Top-down decisionmaking bodies are often deprived of relevant factual information necessary to make good decisions. This problem is analogous to a common critique of centralized economic planning, which is that centralized policymakers are typically epistemically insulated from the complex stream of market information required to accurately coordinate supply and demand.[56]

The PDC model solves this issue in two ways. First, by involving a diverse set of stakeholders, decisionmakers ensure that as much relevant information as possible is available to them at the outset. Excluding diverse stakeholder involvement deprives decisionmakers of rich sources of relevant information.

Second, even if decisionmakers had a full grasp of all relevant information at the time of deployment, introducing advanced technologies into dynamic medical settings may produce unexpected (desirable and undesirable) effects. For example, employing carebots in one context might exacerbate isolation and loneliness but empower residents with more independence and privacy in another.[57] In short, well-informed deployment policies require regular monitoring over time of all the areas touched by the new technology.

The iterative nature of the PDC's evaluation process (Step 4) ensures that new information is continuously integrated into the deliberative process. Without more direct epistemic access to these kinds of effects, decisionmakers risk deploying advanced technologies in ways that adversely impact the values and practices involved in good medicine.

Conclusion

In this paper, we have argued that decisionmaking about the ethical deployment of advanced technologies ought to emphasize process over principles. Contemporary models and methods for decisionmaking rely heavily on generalist forms of principlism and top-down decisionmaking. We have argued that these approaches present a variety of practical, moral, and epistemic challenges. In response, we presented a new approach, PDC, which addresses the challenges inherent to existing models and provides practical guidance for medical institutions. The participatory element ensures that a broad set of stakeholder values and concerns influences policy outcomes, whereas the deliberative element ensures the quality of policy outcomes. Finally, the conservative element recognizes the goods and values inherent to the practice of medicine and the legal and moral responsibility of physicians and administrators for the effects of policy decisions.

Importantly, we do not see PDC and domain-specific principlism as necessarily mutually exclusive. Throughout, we have emphasized the idea that, beyond broad themes, the contextual features of deployment decisions severely limit the possibility of knowing *a priori* stakeholder values and their rankings. However, institutions or decisionmaking bodies may still elect to adopt one of the many possible domain-specific principlisms for guidance. That is, the enumerated principles might be used more as suggestions and reminders rather than systematically deployed top-down. Whether or not institutions elect to adopt a principlist model, ethically sound deployment decisions in concrete situations will nevertheless still require the processes we have outlined throughout the paper.

Notes

1. See Coekelbergh M. *AI Ethics*. Cambridge, MA: MIT Press; 2020; Turkle S. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books; 2011; Vallor S. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press; 2016.
2. United Nations. *Resource Guide on AI Strategies*; 2021, 6–7; available at https://sdgs.un.org/documents/resource-guide-artificial-intelligence-ai-strategies-25128 (last accessed 12 Aug 2022).
3. Van Wynsberghe A. Designing robots for care: Care centered value-sensitive design. *Science and Engineering Ethics* 2013;**19**(2):407–33.
4. Topol E. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books; 2019.
5. See, Barnard A, Sandelowski M. Technology and humane nursing care: (Ir)Reconcilable or invented difference? *Journal of Advanced Nursing* 2001;**34**(3):367–75; Etzioni A, Etzioni O. The ethics of

robotic caregivers. *Interaction Studies* 2017;**18**(2):174–90; Archibald M, Barnard A. Futurism in nursing: Technology, robotics and the fundamentals of care. *Journal of Clinical Nursing* 2018;**27**(11–12):2473–80; Briganti G, Le Moine O. Artificial intelligence in medicine: Today and tomorrow. *Frontiers in Medicine* 2020;**7**(27):1–6; Stokes F, Palmer A. Artificial intelligence and robotics in nursing: Ethics of caring as a guide to dividing tasks between AI and humans. *Nursing Philosophy* 2020;**21**(4):e12306.

6. Sparrow R, Sparrow L. In the hands of machines? The future of aged care. *Minds and Machines* 2006;**16** (2):141–61; Sharkey A, Sharkey N. Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology* 2010;**14**(1):27–40; Coeckelbergh M. Care robots and the future of ICT-mediated elderly care: A response to doom scenarios. *AI & SOCIETY* 2016;**31**(4):455–62.

7. Vandemeulebroucke T, Dierckx de Casterlé B, Gastmans C. The use of care robots in aged care: A systematic review of argument-based ethics literature. *Archives of Gerontology and Geriatrics* 2018;**74**:15–25.

8. There are approximately 300 university hospitals and academic medical centers representing almost 5% of hospitals in the U.S. American Hospital Association, *Academic Medical Centers*; available at https://www.ashe.org/advocacy/orgs/amc (last accessed 12 Dec 2022). Most hospitals, therefore, likely do not have personnel and resources directed toward developing novel AI algorithms.

9. Although these domains are sometimes treated independently, we follow Müller in that "robotics and AI can…be seen as covering two overlapping sets of systems." Müller V. Ethics of artificial intelligence and robotics. In: Zalta EN, ed. *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University; 2020; available at https://plato.stanford.edu/entries/ethics-ai/#AIRobo (last accessed 15 Aug 2022).

10. Hagendorff T. The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines* 2020;30:99–120. For a comprehensive database of AI ethics guidelines, see Algorithm Watch, *AI Ethics Guidelines Global Inventory*; available at https://inventory.algorithmwatch.org/ (last accessed 12 Dec 2022).

11. European Commission, High Level Expert Group on AI. *Ethics Guidelines for Trustworthy AI*; 2019; available at https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf (last accessed 16 Aug 2022). For the remainder of the paper, we will refer to the High Level Expert Group as "AIHLEG."

12. Mittelstadt B. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 2019;**1** (11):501–7.

13. McNamara A, Smith J, Murphy-Hill E. Does ACM's code of ethics change ethical decision making in software development? In: Leavens G, Garcia A, Păsăreanu CS, eds. *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. New York NY: Association for Computing Machinery; 2018; available at https://dl.acm.org/doi/proceedings/10.1145/3236024 (last accessed 16 Aug 2022).

14. We appreciate the comments of an anonymous reviewer for helping to clarify this point.

15. See note 1, Coekelbergh 2020, at 170.

16. Jonsen A. *Bioethics Beyond the Headlines: Who Lives? Who Dies? Who Decides?* Lanham, MD: Rowman & Littlefield; 2005.

17. Although we refer to this view as "principlism," strictly speaking, this family of views is committed to multiple *values*. The commitment to these values is often articulated in principles. For example, most principlist views include the value of autonomy but articulate this commitment as a principle; for example, "one ought to ensure that new technologies respect autonomy."

18. For a recent review, see note 10, Hagendorff 2020. Since robotics codes of ethics are in their infancy relative to those for AI ethics, we focus on the latter. For example, the UNESCO COMEST Report on Robotic Ethics (2017) notes that although "ethical codes specifically written for roboticists seem to be still in their infancy," there have been several institutional initiatives for ethical regulation of robotics (p. 38). The COMEST report also summarizes the core principles and values that should guide the ethical development and deployment of robots: dignity, autonomy, privacy, harm avoidance, responsibility, beneficence, and justice (p. 49). See UNESCO, Commission on the Ethics of Scientific Knowledge and Technology, *Report on Robotic Ethics*; 2017; available at https://unesdoc.unesco.org/ark:/48223/pf0000253952 (last accessed 27 February 2023).

19. See also Université de Montréal, *Montreal Declaration for a Responsible Development of AI*; 2018; available at https://www.montrealdeclaration-responsibleai.com/ (last accessed 16 Aug 2022).

20. See note 11, European Commission 2019, at 2, 11.

21. See note 16, Jonsen 2005; see also Huxtable R. For and against the four principles of biomedical ethics. *Clinical Ethics* 2013;**8**(2–3):39–43; McMillan J. *Methods of Bioethics: An Essay in Meta-Bioethics*. Oxford: Oxford University Press; 2018; Friedman B, Hendry D. *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press; 2019; Flynn J. Theory and bioethics. In: *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University; 2021; available at https://plato.stanford.edu/entries/theory-bioethics/ (last accessed 16 Aug 2022).

22. Department of Defense, Defense Innovation Board, *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence*; 2019; available at https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF (last accessed 16 Aug 2022).

23. See note 22, Department of Defense 2019, at 31.

24. See note 22, Department of Defense 2019, at 31.

25. Sparks J. Privacy as informational health. Unpublished manuscript; 2022.

26. Timmerman T, Cohen Y. The limits of virtue ethics. In: *Oxford Studies in Normative Ethics*. Vol. **10**. Oxford: Oxford University Press; 2020: 255–82.

27. There have been a variety of cases where deployment decisions were made without wider community consultation which resulted in numerous moral harms. For a recent review on the issue of predictive policing, see Alikhademi K, Drobina E, Prioleau D, Richardson B, Purves D, Gilbert JE. A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law* 2022;**30**:1–17; see also Think Tank, European Parliament, *Artificial Intelligence in Healthcare: Applications, Risks, and Ethical and Societal Impacts;* 2022; available at https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729512 (last accessed 1 Jan 2023).

28. See note 11, European Commission 2019, at 6.

29. Aristotle. *Nicomachean Ethics.* Reeve CDC, trans. Indianapolis–Cambridge: Hackett; 2002.

30. The alternative would be to allow for technology to be deployed in ways that undermine the protection and promotion of health.

31. See note 21, Huxtable 2013; Taylor R. Ethical principles and concepts in medicine. *Ethical and Legal Issues in Neurology* 2013;118:1–9; see note 21, McMillan 2018; see note 21, Flynn 2021.

32. See note 5, Etzioni, Etzioni 2017.

33. See note 6, Coeckelbergh 2016.

34. See note 5, Stokes, Palmer 2020.

35. Further, even if it were true that each subdomain of medicine had its own set of values, this would not aid clinical decisionmaking in complex care situations. Modern hospital care is typically delivered by multidisciplinary teams. Some metaprinciple would be required to sort out which subdomain's values should take precedence. This is precisely the problem our model seeks to address by rejecting the notion that one can enter complex clinical situations with *a priori* commitments to certain values and their ranking.

36. Barach P, Johnson JK. Understanding the complexity of redesigning care around the clinical microsystem." *Quality and Safety in Health Care* 2006;**15**(1 Suppl):i10–i16.

37. See note 1, Vallor 2016.

38. Coeckelbergh M. Personal robots, appearance, and human good: A methodological reflection on roboethics. *International Journal of Social Robotics* 2009;**1**(3):217–21.

39. Campbell B, Knox P. Promise and plausibility: Health technology adoption decisions with limited evidence. *International Journal of Technology Assessment in Health Care* 2016;**32**(3):122–5.

40. Santana M, Manalili K, Jolley RJ, Zelinsky S, Quan H, Lu M. How to practice person-centred care: A conceptual framework. *Health Expectations* 2017;**21**(2):429–40.

41. Some authors have argued for a department of clinical artificial intelligence. We remain neutral on this debate, however were this to become a general practice, members of the clinical AI team should also participate in the deliberative process for deployment decisions. See Cosgriff CV, Stone DJ,

Weissman G, Pirracchio R, Celi LA. The clinical artificial intelligence department: A prerequisite for success. *BMJ Health & Care Informatics* 2020;27:e100183.

42. Baek C, Choi JJ, Kwak SS. Can you touch me? In: *Proceedings of the Second International Conference on Human–Agent Interaction*; 2014. doi:10.1145/2658861.2658909; Choi JJ, Kim Y, Kwak SS. Are you embarrassed? In: *Proceedings of the 2014 ACM/IEEE International Conference on Human–Robot Interaction*; 2014. doi:10.1145/2559636.2559798.

43. See note 21, Friedman, Hendry 2019; see also note 19, Université de Montréal 2018. Very broadly, this resembles processes recommended by the Value Sensitive Design approach and employed in the recent Montreal Declaration for Responsible Development of AI, though our approach here focuses on *deployment* of advanced technologies rather than *development*.

44. Zaner RM. *Conversations on the Edge: Narratives of Ethics and Illness*. Washington, DC: Georgetown University Press; 2004.

45. Gastil J, Levine P. *The Deliberative Democracy Handbook: Strategies for Effective Civic Engagement in the Twenty-First Century*. New York: John Wiley & Sons; 2011.

46. Best practices in professional bioethics mediation and healthcare ethics consultation (via American Society for Bioethics and Humanities) are that value-based conflict resolution is best achieved by focusing on underlying values because it provides more avenues for conflict resolution. We follow these established best practices for value conflict resolution. See Vollmann J, Dubler N, Liebman C. Bioethics mediation. A guide to shaping shared solutions. *Ethik in der Medizin* 2007;**19**:161–2; *Core Competencies for Healthcare Ethics Consultation.* 2nd ed. Chicago, IL: American Society for Bioethics and Humanities; 2011.

47. Consider some of the core values at work in the "Principles of Medical Ethics" outlined by the AMA: Patients should be provided with "competent medical care…compassion and respect for human dignity and rights." Physicians should act in the "best interests of the patient" and should "continue to study, apply, and advance scientific knowledge" and "maintain a commitment to medical education." In addition, the physician should contribute to and improve "public health." See American Medical Association. *AMA Principles of Medical Ethics*; 2016; available at https://www.ama-assn.org/about/publications-newsletters/ama-principles-medical-ethics (last accessed Aug 16 2022).

48. If at Step 2 or 3 the committee decided that there was no value-concordant way of deploying the technology, the process terminates. Step 4 only occurs if the committee determines that there is a value-concordant way of deploying the advanced technology.

49. Zaner RM. Listening or telling? Thoughts on responsibility in clinical ethics consultation. *Theoretical Medicine* 1996;**17**(3):255–77.

50. Greenberg D, Pliskin J. Adoption and use of new medical technology at the hospital level. *Health Management* 2008;**10**(1); available at https://healthmanagement.org/c/hospital/issuearticle/adoption-and-use-of-new-medical-technology-at-the-hospital-level (last accessed 16 Aug 2022).

51. Administrative and legal perspectives will also be required in order to ensure local and federal regulatory and legal compliance.

52. See note 4, Topol 2019.

53. See note 38, Coeckelbergh 2009.

54. SECOM, My Spoon: Publications; available at https://www.secom.co.jp/english/myspoon/publication.html (last accessed 16 Aug 2022).

55. Bliton M, Finder S. Traversing boundaries: Clinical ethics, moral experience, and the withdrawal of life supports. *Theoretical Medicine and Bioethics* 2002;**23**(3):233–58, at 239.

56. Hayek FA. The knowledge problem. The use of knowledge in society. *The American Economic Review* 1945;**35**(4):519–30.

57. Gawande A. *Being Mortal*. Toronto, ON: Anchor; 2014.