From Principles to Process: The Principlist Approach to AI Ethics and Lessons from Internet Bills of Rights

***

Kinfe Yilma, PhD (University of Leeds)

Accepted for Publication in AI & Ethics (2025)

Abstract

This article examines possible lessons the principlist approach to AI ethics could learn from the relatively developed initiatives for Internet bills of rights. From intergovernmental organisations, governments to technology companies and civil society groups, a series of guidelines that lay out principles for the ethical design, development and use of AI technologies have emerged. Despite the prominence of this 'principlist' approach to AI ethics in the literature, its interplay with similar approaches to technology governance has not received any attention. This article seeks to fill this void by exploring possible lessons that the more nascent principlist approach to AI ethics could learn from Internet bills of rights. It argues that the future of the principlist approach hinges on averting factors that undercut the potential of the Internet bill of rights project while at the same time drawing upon its progressive elements. The article thus explores imperatives of grounding the principlist approach in a sustained normative and institutional process that enjoys a level of legitimacy and normative novelty. That would shift the focus in AI ethics discourse from articulating principles to setting in motion a process that would clarify, sustain and legitimise them.

Keywords: AI ethics, principlism, principlist approach, AI governance, Internet bills of rights, digital constitutionalism

1. Introduction

One of the effects of the unprecedented growth in the development and use of Artificial Intelligence (AI) in recent years has been the proliferation of AI ethics guidelines. From intergovernmental organisations, governments to technology companies and civil society groups, a series of guidelines that set out principles for the ethical design, development and use of AI technologies have emerged. Recent surveys have documented hundreds of initiatives that articulate AI principles (Corre^a et al, 2023; Jobin et al, 2019). This "AI ethics boom" has put to the spotlight the role of principles in the governance of AI. To that effect, the role of this "principlist" approach to AI ethics has been a focus of sustained scholarly attention in the past few years (Hagendorff, 2020; Field et al, 2020; Schiff, 2022; Seger, 2022).

Except for some consideration of possible lessons from public health ethics or sustainability (Nyrup, 2021; Auld et al, 2022), no attention is given to what lessons that this emergent approach may draw from other similar approaches to technology governance. This paper is an attempt to fill that void by exploring lessons the more nascent principlist approach to AI ethics could learn from the relatively developed but less successful initiatives for Internet Bills of Rights (IBRs). Sometimes referred to as a "movement", "project" or more recently "digital constitutionalism", IBRs represent the multi-level initiatives that sought to offer new ways of framing rights, norms and principles for the governance of the Internet and allied technologies.

There are three reasons why IBRs offer a unique vantage point from which to examine the role of the principlist approach to AI ethics in the governance of AI. First, there is an underlying similarity of objective between initiatives for IBRs and the principlist approach. IBRs typically embody rights as well as principles whereas AI ethics guidelines are principle-focused generally. Principles and rights are not conceptually identical, of course. Rights, for example, bestow entitlements—enforceable or otherwise—to individuals or groups but principles do not. According to Raz, rights are "grounds" of duties based upon an individuals' interest (Raz, 1984). That means most rights, if not all rights, involve correlative duty bearers who would be responsible to fulfil the legitimate interests guaranteed by the rights. A key feature of principles, on the other hand, is that they "leave their subjects unspecified", and thus "prescribe highly unspecified actions" (Raz, 1972). With this inherent generality, the central function of principles is to provide the overarching background against which rights (and other rules) are interpreted and applied (Raz, 1972).

But both "digital rights" and AI principles as conceived in the domains in question have a commonality of objective. The objective is to provide action-guiding standards, for example to lay out standards for the design of technologies in line with human rights and ethical standards. Recent years have, of course, seen a noticeable convergence between the principlist approach to AI ethics and IBRs. For instance, recent AI ethics initiatives are resorting to the "bill of rights" label. After the proposal for a crowd-sourced "bill of rights for AI-powered world" that would "clarify" constitutional rights for the 21st century, the US government has released a Blueprint for AI bill of rights followed recently by a wide-ranging Executive Order (Lander and Nelson, 2021; White House, 2022; Executive Order 14110, 2023). This emergent convergence lends weight to the aim of this article, i.e., to explore lessons that Internet bills of rights may offer to the principlist approach to AI ethics.

Second, the momentum towards IBRs appear to have waned in recent years. As shall be discussed further in Section 3, the IBRs project has been momentary that gains traction following certain events but dies down gradually. The Snowden revelations in mid-2013, for instance, was the major latest episode in the IBRs movement. In the wake of the revelations, a series of IBRs were launched by various stakeholders. But the hype generally died down in the following years. If the principlist approach—which some claim is now "methodologically adolescent" (Lundgren, 2023)—is to avoid a similar fate, it should learn from the shortcomings as well as the strengths of the rather mature IBRs. In discussing key aspects of IBRs, the aim of this article, then, is to contribute towards saving the principlist approach from the fate of IBRs.

Third, a growing body of scholarship in AI ethics increasingly seeks to move the focus from "what" principles to "how" to translate the principles into practice (Umbrello et al, 2021; Morley et al, 2020; Palmer and Schwan, 2024). But this comes at a time when the principlist approach faces a wide-ranging critique, including about pervasive conceptual ambiguities and tensions in AI principles (Munn, 2023). Moving to translate AI principles into technological design in the face of such concerns is problematic. Beside other efforts to address the concerns, seeking to draw lessons from initiatives of similar character, origin and objective offer a useful lesson. That is the other motivation for this article.

Against this backdrop, this paper argues that the future of the principlist approach hinges on averting factors that undercut the potential of IBRs while at the same time drawing upon its strength. The

paper explores four interrelated key lessons IBRs offer to the principlist approach, namely "durable normativity", "legitimacy", "institutional process" and "exercise in novelty". We critically discuss the imperatives of grounding the principlist approach in a sustained normative and institutional process that enjoys a level of legitimacy and novelty. In discussing these lessons, the aim is to inform ongoing discussions at various levels, including in academia, AI companies and policymakers. The hope is that the four lessons would help in making informed decisions not only in developing AI principles but also in translating AI principles in technology design or incorporating them into AI regulatory initiatives.

The rest of the paper proceeds in four sections. Section 2 explores the salient features, nature and scope of the principlist approach to AI ethics. Literature on AI principles have grown exponentially in the past few years, but this section seeks to distill central and cross-cutting themes in a fast-evolving field. A brief introduction to IBRs with a particular focus on aspects of IBRs that offer useful lessons to IBRs is provided in Section 3. Section 4 explores four key lessons that the IBRs movement offers to the principlist approach to AI ethics. Section 5 concludes the paper.

A few caveats are in order. First, we use the terms "initiatives for IBRs", "IBRs movement", "IBRs project", and "IBRs" interchangeably to refer to initiatives for digital rights put forward at various levels. By "principlists", the reference is simply to institutions that author AI ethics guidelines. Second, the paper does not provide the full mechanics of the proposed institutional process, not least for lack of space.

2. The principlist approach in a nutshell

Initiatives for AI ethics are known by several nomenclatures. They are sometimes referred to as the "principled approach" or "ethical principlism" in light of the role of medical ethics played in inspiring AI ethics principles (Mittlstadt, 2019; Palmer and Schwan, 2024; Stahl, 2021). Others have referred to them collectively as the "legalistic conception of AI ethics" (Resseguier and Rodrigues, 2020). This article follows Nyrup's terminology who refers to them as a "principlist" approach. Nyrup defines it as an approach that "seeks to articulate a small set of principles to guide ethical decision-making" (Nyrup, 2021).

The approach, in that sense, is a collective label for a series of initiatives aimed at rolling out principles meant to offer the building blocks for the governance of AI, including its ethical design, development and use. For purposes of this paper, the phrase the "principlist approach to AI ethics" refers to the initiatives for AI ethics advanced in the form of AI principles and ethics guidelines. Within that frame, we deploy the terms "principlist approach", "AI principles" and "ethical/ethics guidelines" interchangeably. This is helpful to make sense of an otherwise dispersed, multilevel set of initiatives of AI ethics.

The origins of the principlist approach traces back to the remarkable increase in AI research, development and use in the past decade. Many scholars credit the Asilomar Conference for Beneficial AI in 2017 for spearheading the first set of AI principles (Whittlestone et al, 2019; Floridi, 2023). But the principles of robotics launched by the Engineering and Physical Sciences Council in 2010 are perhaps an earlier iteration of AI principles (Principles of Robotics, 2010). But certainly, the recent uptick in the number of AI ethics principles is directly related to the significant advances in AI

research, development and deployment in the past decade. In that sense, while AI principles are not all new, their prominence has increased in recent years in the wake of advances in the field, and the worries over the attendant ethical risks.

A closer look at the trajectory of AI ethics initiatives in the past decade reveals that the principlist approach has largely been a fragmented and multi-level initiative. Pursued by various actors with particular institutional goals, it is yet to evolve into a structured mechanism of governance. At the time of writing in January 2024, Algorithm Watch's comprehensive inventory lists 173 AI ethics guidelines (Algorithm Watch, 2024). But one rarely sees any conversation between authors of each initiative. The sheer volume of documents launched in a span of a short period of time reinforces the point.

Based on the scope of the respective initiatives, Palmer and Schwan identify two versions of the principlist approach: generalist and domain-specific. The generalist approach establishes a set of principles applicable to many or all distinct domains of practice (Palmer and Schwah, 2024). A good case in point is the Ethics Guidelines for Trustworthy Artificial Intelligence of the European Union (EU) which is directed at almost all domains where AI may be applied (High-Level Expert Group on AI, 2019). What follows from this is that principles envisaged in such a framework are formulated in generic terms. That would make the principles generally applicable to all domains in which AI technologies apply. But an inherent limitation is that the principles do not offer more than high-level guidance. Given the niche ways in which AI might be applied in practice, general principles might fall short in providing nuanced action-guiding norms.

The domain-specific approach, on the other hand, applies to ethical decision-making involving advanced technologies in specific domains (Palmer and Schwah, 2023). Microsoft's principles on facial recognition technologies are an example of domain-specific ethics guidelines (Microsoft, 2018). Other examples include principles applicable in specific fields such as health care, law enforcement or education provision. A key advantage of domain-specific approach is that the principles provide tailored action-guiding norms. Ideally, that would make the principles better positioned to enable the ethical design, development or use of AI technologies in specific contexts. Yet, that AI-driven technologies tend to evolve fast, might mean the principles fault to remain abreast of technological developments. In contrast, generic principles would generally have the potential to apply across time regardless of rapid technological advances. That said, the existence of a diverse body of AI principles applicable for different contexts might be, to a degree, desirable. But it may also be an indication of varying priorities among authors of each initiative.

The proliferation of AI principles is generally taken as a response to a perceived governance lacuna created by the advent of AI technologies (Jobin et al, 2019). According to this viewpoint, AI principles emerged due to the governance and regulatory lacuna at various levels. No doubt, there is some validity in this argument. As a rapidly evolving suite of technologies, the gestation of AI regulatory initiatives by states or regional bodies is bound to take some time. The European Union's (EU) AI Act for instance was first proposed in April 2021 and came into force only in August 2024. AI principles, in this regard, could be said to have emerged to temporarily fill the regulatory void. The EU AI Act was, of course, preceded by the Ethics Guidelines for Trustworthy Artificial Intelligence. To that degree, AI principles were a temporary way station to permanent regulatory instruments such as the AI Act.

This conclusion does not however apply to all sorts of AI principles. Similar regulatory instruments are yet to emerge in other parts of the word. Instead of binding instruments such as the AI Act or domestic AI legislation, a number of countries are resorting to national AI strategies and white papers. In the UK for example, no AI legislation is anticipated in the near future. The government issued, instead, a White Paper that offers non-binding set of guidance on AI (UK AI Regulation White Paper, 2023). In Africa, the predominant approach has been the adoption of AI strategies. Close to a dozen countries have introduced AI strategies, and the African Union has recently introduced a continental AI strategy (AU Continental AI Strategy, 2024). Often, AI strategies embody AI ethics principles. So the point is that AI principles continue to play the role of filling the governance vacuum in parts of the world where AI legislation does not exist.

But the argument that AI principles are aimed at filling a governance vacuum does not offer a full explanation. A closer study of the AI ethics guidelines and the accompanying literature reveals a multitude of objectives dictated by the institutional priority of the respective author. Offering a starting point for a formal governance framework is widely recognised to be the chief driver of many of the AI principles. AI principles are, hence, seen as building blocks of future legislation at various levels (Coeckelberg, 2020). This role of AI principles is often associated with government-led initiatives (Fjeld et al, 2020), but it is now commonplace for civil society and private sector initiatives to call for legislation (Microsoft, 2018). At times, the sought-after legislation is internal, i.e., by-law, aimed at governing the internal development and use of AI within technology companies (Fjeld et al, 2020; Telia, 2019).

Offering guidance for the ethical design of AI technologies is the second recurring objective of AI principles (Nyrup, 2021). This is targeted at AI developers, and in that sense, the goal ultimately is to nurture a culture of responsibility among AI developers (Seger, 2022). This positions AI principles as an ethical blueprint for technologists. Implicit in this articulation of the role of AI principles is that AI principles are aimed at influencing industry practices. If taken as the sole objective of the principlist approach, this would however considerably limit the role of AI principles. Left for the whim and will of AI companies, AI principles are likely to have limited influence in the development of technologies.

Related to this is the 'advocacy' objective of AI guidelines by which they serve as advocacy tools towards shaping public policy on AI (Fjeld et al, 2020). In articulating the ways in which AI should be governed, proponents of AI principles essentially are seeking to shape law and policymaking. This appears to be the primary objective of most civil society-led AI ethics initiatives. AI principles have emerged in great numbers in recent years but there is no mechanism of achieving their objectives of shaping policy. In the absence of institutional processes, as shall be discussed further in Section 4, AI principles would not have meaningful impact in either public policy or industry practice. It is not clear, for instance, to what extent civil society led AI principles have shaped the making of AI legislation at the national or regional level.

The third envisioned role of AI principles is discursive by which the principles offer the norms to discuss the opportunities and risks of AI (Whittlestone et al, 2019). Envisioned this way, the release of AI principles by various stakeholders should be taken as part of a conversation on AI ethics. Through interactive processes, this discourse would ideally contribute towards the emergence of

consensus among stakeholders on AI ethics. But rarely has this "conversation" been part of sustained discourse that would have otherwise lent weight to the creation of consensus. Further compounding the concern is the lack of institutional process that would sift through the pile of AI principles or facilitate the conversation. We return to consider the issue of institutional process in Section 4.

Overall, the foregoing discussion illustrates the fragmented nature of the principlist approach to AI ethics. That it is fragmented means that the underlying rationales of AI ethics initiatives vary. But this dispersive approach would ultimately lessen the potential impact of AI principles in meeting the objectives alluded to above. In part, this is already illustrated in the public policy sphere. Outside Europe, legislation on AI is few and far between. If this short trajectory of the principlist approach is not reversed, it is bound to run into disuse as IBRs did. IBRs offer therefore some insight in making the principlist approach fit for purpose. Before we turn to exploring these insights or "lessons" from IBRs, what follows provides a brief overview of IBRs. We focus particularly on aspects of IBRs that stand to offer useful lessons to the principlist approach.

3. Internet bills of rights: A brief introduction

IBRs is a collective nomenclature for a series of initiatives that envision a set of rights, principles and governance norms to the new realities of the digital age (Gill et al, 2015). Prompted mainly by the new ways in which human rights are undermined or established legal principles fail to be fit for purpose, IBRs initiatives aim to translate or adapt pre-existing human rights and principles to the new digital realty (Musiani, 2009; Celeste, 2023). Article 19's Universal Declaration of Digital Rights offers a useful example. Parts of the Declaration's preamble read as follows (Universal Declaration of Digital Rights, 2017):

> Reaffirming the relevance of international human rights standards in the digital environment and the need to explore and expand new human rights guarantees for the future;
>
> We therefore now proclaim this Universal Declaration of Digital Rights as a common standard of digital freedoms for all […] (Emphasis added)

A key feature of Article 19's Declaration is that it envisages a broad range of rights which find no direct counterpart in the existing catalogue of human rights. Among such rights include "right to anonymity", "right to encryption", "right to be free from surveillance" and "right to hack" (Universal Declaration of Human Rights, 2017: Articles 4-6, 13). Similar novel and quasi-novel rights are introduced in a broad range of IBRs introduced in the past two decades. The Italian Declaration of Internet of Internet rights, for example, 'right to the inviolability of electronic systems, devices and domiciles' and 'right of anonymity' (Italian Declaration of Internet Rights, 2015: Articles 7, 10).

Similar to AI principles, IBRs are authored by governments, intergovernmental organisations, technology companies and civil society actors. The majority of IBRs initiatives had been launched by civil society groups as an advocacy tool, but in a few cases, governments had proposed or adopted some form of IBRs legislation (Italian Declaration of Internet Rights, 2015; Nigerian Digital Rights and Freedom Bill, 2018). Although its scope was broadened to address a broad range of issues on Internet regulation, Brazil's Marco Da Civil is often considered, or at least at its inception, as an example of a national bill of rights for the Internet (Marco Da Civil, Law No. 12.965, 2014). Not

only that it recognises widely advocated digital rights such as the "right of Internet access", but also it sets out governance norms on topics such as net neutrality and intermediary liability (Marco Da Civil, Law No. 12.965, 2014: Chapter II, Chapter III, Sections I & III.).

In terms of geographic distribution, most IBRs initiatives were proposed in the Global North. In the few cases where countries or organisations in the Global South rolled out IBRs documents, one sees a significant similarity of normative content with those launched in the Global North (African Declaration of Internet Rights and Freedoms, 2014). Among the rights and principles incorporated in the African Declaration include "right of Internet access" and "privacy and data protection" which are common in other IBRs. But curiously, the Declaration does not envisage other novel and quasi-novel rights which are common in other IBRs discussed above.

A feature of IBRs worth highlighting is the ways in which authors of IBRs sought to enhance the legitimacy of IBRs documents, be it the legitimacy of substance or process (Yilma, 2022). As shall be explored in more detail in Section 4, the IBRs project has valuable lessons in enhancing procedural and substantive legitimacy. In routinely invoking international human rights law as well as collating rights codified in domestic constitutional law, IBRs tends to enhance its substantive legitimacy. A number of IBRs are results of collaborative norm entrepreneurial endeavours whereby relevant stakeholders are open to the development of IBRs. We return to consider how this may offer a "positive" lesson to the principlist approach to AI ethics.

As the name readily suggests, IBRs seek to draw inspiration from national and international bills of rights (Yilma, 2022). As an essentially human rights movement, IBRs draws upon and invokes the "international bill of human rights" which consists of the Universal Declaration of Human Rights, and the two Covenants on Civil and Political Rights and on Economic, Social and Cultural Rights (Humphrey, 1976). The Magna Carta and the American Bill of Rights are examples of national bills of rights upon which the idea of IBRs is benchmarked (The Guardian, 2014). By drawing upon established legal standards, proponents of IBRs hope to enhance the legal standing and legitimacy of the initiatives. As highlighted above, Article 19's Universal Declaration aims to "reaffirm" international human rights law to new realities of the digital age. However, IBRs exhibit a level of novelty that departs from the structure of conventional bill of rights instruments in two ways.

One is that they tend to read into the existing human rights catalogue rights that find no direct counterpart either in international or national bills of rights. Beyond "reaffirming" existing international human rights in the digital environment, Article 19's Universal Declaration—for instance—recognises the need to explore "expanding" the existing catalogue of human rights. With respect to the right to privacy for example, some IBRs documents embody specific rights such as the "right to use encryption technologies" and "freedom from surveillance" (Universal Declaration of Digital Rights, 2017).

Envisioning technology companies as duty bearers is the other progressive aspect of IBRs initiatives. Private actors such as companies owe no direct human rights obligations in international law. Instead, they are subject to a non-binding framework of corporate social responsibility (Guiding Principles on Business and Human Rights, 2011). It is through states' positive human rights duty that the conduct of corporations may be held to account through domestic legislation. By envisaging technology companies as duty bearers, IBRs are seeking to address this longstanding gap. This exercise in novelty

is the second strength of the IBRs project that offers a "positive" lesson to AI principles. We return to this point in Section 4.

A characteristic feature of the IBRs project is the sporadic and fragmented nature of the initiatives. It has largely been a momentary movement that rises and falls with particular events. This is best illustrated by the proliferation of IBRs initiatives in the wake of the Snowden revelations in the summer of 2013. Edward Snowden, a former contractor at the US National Security Agency (NSA), provided millions of classified documents that unveiled unprecedent mass surveillance practices of the US and its allies (The NSA Files, 2013). Intelligence agencies of the so-called Five Eyes States (US, UK, Australia, Canada and New Zealand) tapped into and shared among themselves phone records, fibre optic cables and Internet servers, at times with the acquiescence of technology agencies and on a mass scale. Not only were the surveillance measures carried out on a mass scale but also had targeted world leaders such as leaders of German and Brazil.

The unprecedented nature and scale of the revelations generated worldwide condemnation, but importantly it also reignited the Internet bill of rights movement. A number of actors, mainly civil society groups, launched a series of initiatives that seek to uphold human rights in the face of mass digital surveillance practices. A memorable example is the call for a "digital magna carta" proposed by inventor of the web Tim Berners-Lee (The Guardian, 2014). This proposal was shortly followed by a campaign to draft the magna Carta through a crowdsourcing mechanism, and resulted in the launch of a 'contract the web' on 25 November 2019 on the occasion of the UN Internet Governance Forum in Berlin (Contract for the Web, 2019). Not only was the Contract a set of nine high-level principles addressed to governments, technology companies and citizens, but it appears to have faded into the background since.

But the post-Snowden hype did not linger as attention quickly shifted towards the next hype about emergent technologies such as AI. The foregoing examples illustrate the sporadic and momentary nature of the digital bill of rights movement. Part of the reason for this lack of "durable normativity" is the absence of some form of institutional process that would steer the initiatives towards a unified goal. We turn to consider in Section 4 that embedding institutional processes in the principlist approach to AI ethics. What follows turns to explore what the above features of the IBRs project teach to the principlist approach.

4. From principles to process: Lessons to AI ethics from IBRs

Despite recent calls for (Prem, 2023; Morley et al, 2020)—and even concrete initiatives (Seppälä et al, 2021)—to translate AI principles into technology design, the principlist approach remains, by and large, in a state of flux. Many of the ethics guidelines archived in Algorithm Watch's directory were adopted before April 2020. And curiously, only six new guidelines have been added since (Algorithm Watch, 2024). This may be taken to imply there is a notable saturation of AI ethics initiatives. But when one considers that almost all AI ethics guidelines—discussed further below, are from the Global North, this tendency of saturation is not necessarily a sign of maturity of the principlist approach. And of course, the repository cannot be taken to be definitive evidence to arrive at such a conclusion.

Indeed, initiatives that introduce principles for the ethical design, development and use of AI continue to emerge. Google, for instance, recently launched its updated catalogue of AI principles (Google AI

Principles, 2023). Of governments, the government of Saudi Arabia, for instance, recently issued AI Ethics Principles (Saudi Arabia AI Ethics Principles, 2023). In Africa, about a dozen countries have issued national AI strategies that, among other things, tend to envision AI principles as building blocks of AI governance. At the regional level, the African Union has adopted an AI Strategy that similarly frames AI principles as a starting point for continental framework of AI governance (AU Continental Strategy, 2024).

What this suggests is that the momentum around AI principles is not fading. Instead, it appears that the principlist approach is moving to parts of the world where frameworks for the governance of AI are yet to emerge. This in turn offers an opportune moment to explore the ways in which some of its fundamental flaws may be tackled before it develops into legal and technical standards, or even worse before it is overtaken by the next hype. In that spirit, this section discusses four lessons that the principlist approach may learn from IBRs.

4.1. Durable normativity

"Normativity" is a term of art in philosophy relating to the question of what "ought to" other than "what is" (Wedgwood, 2007). For purposes of this paper, "normativity" refers to the act of producing action-guiding norms by public bodies (e.g., governments), private actors (e.g., civil society groups and corporations), and at times, technologies (e.g. algorithms) (Fourneret and Yvert, 2020). When we, then, refer to "durable normativity", it simply means an act of norm generation that is sustained, targeted and coordinated over a period of time. Put differently, a durable normative process is one that does not rise and falls with changing circumstances. Formative normative processes would not, ideally, be abandoned and displaced by the next hype.

IBRs, as highlighted in Section 3, were largely a momentary and fragmented discourse. Initiatives often gain a short-lived momentum in the wake of a certain human rights scandal. Such lack of durable normativity has significantly limited its potential of spurring meaningful outcome in policy or practice. Apart from the adoption of legislation of broader scope in Brazil (Marco Civil da Internet, 2014), and a series of resolutions by the United Nations (UN) (UNGA Res 77/211, 2022), the post-Snowden momentum around IBRs has either faded or have collapsed into another discourse on emerging technologies (HRC Res 53/29, 2023). The principlist approach faces a similar fate unless steps are taken to ensure it forms part of a sustained normative process.

The discourse sparked by the emergence of AI principles has aptly been described as a "fractured global conversation" (Fjeld et al, 2020), partly aimed at dodging regulatory interventions (Piedra, 2023). As highlighted in Section 2, hundreds of initiatives that make up the principlist approach have been launched in a span of a few years, and by various stallholders. And this proliferation of AI principles has not abated to date. While there are inevitable overlaps between principles, one sees no conversations among authors of each initiative. Importantly, new initiatives appear to emerge regardless of or with little heed to pre-existing principles. What that essentially means is that the initiatives—and the principles they embody—are not part of a sustained normative process or discourse. Each initiative appears to be motivated by and pursue isolated objectives than to sustain conversation sparked by earlier initiatives.

Rights and principles unilaterally launched by every stakeholder group is what ultimately led to the fragmentation of IBRs. Rarely have actors of IBRs had a conversation (Yilma, 2017). As alluded to above, IBRs have—by and large—been tentative tools of venting frustration over problematic government digital policy or emergence of new technologies. That means the fervour behind most initiatives recedes with time, and actors behind the initiative tend to move to the next target. Indeed, the lack of sustained academic discourse might have contributed to the fragmentation of IBRs. It is only recently that IBRs have gained some attention in the literature where commendable attempts are made to make sense of an otherwise deeply dispersed set of initiatives (Celeste 2018; Celeste, 2022). But scholarship in the field might have arrived as the IBRs loses momentum.

The trajectory of the principlist approach, as alluded to above, appears to be similar with IBRs. The lack of durable normativity would have bearing in the ultimate effect of AI principles in the development and use of AI in the years to come. Palmer and Schwan argue that the net effect of the sporadic nature of AI initiatives is that the principlist approach cannot offer a comprehensive list of ethical principles that attend to the interest of all stakeholders (Palmer and Schwan, 2024). Their concern is that a fragmented principlist approach would fail to be comprehensive in its normative structure. But that is not the main aftereffect of the sporadic and fragmented nature of the principlist approach.

Perhaps a worrisome implication might be that the hype towards identifying principles may soon be replaced by a rush to translating them into product design. Recent AI ethics scholarship tends to turn attention towards on how best to translate principles to practice (Prem, 2023; Morley et al, 2020; Seppälä et al, 2021). Many countries are investing heavily towards the development of vibrant AI sectors (Government of UK, 2025; BBC News, 2025). AI companies are rushing to develop enhanced versions of large language models. But in the absence of a widely accepted set of AI ethics principles, AI companies would be able to co-opt which principles to bake into designs or to lobby into legislation. Worse, companies would be at liberty to define complex notions such as fairness and justice in a manner that suits their corporate interests.

To overcome this challenge, bringing forth some normative durability in the principlist approach is vital. Addressing the question of how this could be achieved is beyond the scope of this article. But some points that can inspire further discussion can be made. Actors with stake in the field hold a unique position in enhancing durability in the normative process of AI principles. One such actor is AI ethics scholars. Sustained scholarship, for instance, would be instrumental in distilling areas where consensus in the principlist approach is emerging. This would provide the epistemic basis for formal processes towards durable normativity.

We will return to discuss in greater detail the vitality of "institutional process" in Section 4, but global and transnational entities with mandates on AI ethics are the other key players in bringing forth durable normativity in AI ethics. Among such bodies include the United Nations Educational, Scientific and Cultural Organisation (UNESCO) and the Organisation for Economic Cooperation and Development (OECD). Both intergovernmental entities have launched AI ethics guidelines targeted at a wider audience but have been just two principlists among dozens of other stakeholders. But their institutional infrastructure as well as mandate put them in a unique position to initiate the steps towards durable normativity in the principlist approach to AI ethics.

4.2. Legitimacy

Legitimacy, or avoiding the lack of it, is the second lesson that the principlist approach may draw from IBRs. Earning legitimacy–both substantive and procedural–is a primary consideration in any project aimed at laying down a framework of governance. Not only the substance of AI principles but also the process through which they are crafted must earn legitimacy to achieve their underlying objective of setting forth action-guiding norms. Legitimacy is a contested notion among political scientists and legal scholars with competing definitions of what makes an act "legitimate".

In the context of this article, the notion 'substantive legitimacy' concerns the legitimacy or pertinence of the content or the "purposive direction" of AI principles (Barnard, 2001). It is about "what" makes the outcome right. Whereas "procedural legitimacy" relates to sources and nature of the process that validates the outcome (Yilma, 2022; Furendal, 2022). As such, it concerns as to "who" – and "how" – could rightly produce a substantively legitimate bill of rights. It is to be noted however that procedural legitimacy is sometimes seen as a pre-condition to the substantive legitimacy of a given outcome (Besson, 2005). When the output is a product of an illegitimate process, it might, then, be considered illegitimate. Taken all together, legitimacy broadly relates to the extent to which actions of the state or other power wielding entities are acceptable to and are compatible with the values of those affected or governed (Stillman, 1974; Cromartie, 2003). We follow this enunciation when considering the legitimacy of AI ethics guidelines. In light of the diversity of authorship, we particularly focus on the legitimacy of non-State actor led AI ethics initiatives.

As highlighted in the preceding section, IBRs initiatives routinely invoke human rights standards to enhance their legitimacy. One also finds a number of rights and principles in IBRs initiatives that have already found some judicial recognition at national and regional level. For example, the "right to the integrity and confidentiality of information systems" is one of the commonest rights in IBRs instruments (Italian Declaration of Internet Rights, 2015). Before it was widely featured in these instruments, it was introduced by the German Constitutional Court in a 2008 decision (BVerfG, NJW 2008, 822). Instead of building its content wholly from scratch, the IBRs project identifies rights, principles and governance norms that have drawn some level of consensus (Yilma, 2022).

Unlike IBRs—which by and large is a human rights movement, the principlist approach does not often deploy human rights language or invoke human rights to support its lofty claims, except for the Toronto Declaration (Toronto Declaration, 2018). The Declaration is focused on the protection of the rights to equality and of non-discrimination. Deploying human rights language will have a particular value for dispersed initiatives like the principlist approach. For one, it grounds claims on an established system of rules and platforms of advocacy. Human rights language offers a common vocabulary that would bring with it a level of authority that goes beyond mere professional aspirations of ethical principles. But the Toronto Declaration is an exception in that reference to human rights language is not very common in AI ethics initiatives.

AI ethics scholars are increasingly acknowledging the importance of tying AI principles with national and international law (Schiff et al, 2020), with a view—*inter alia*—to address the "fragile" nature of the censuses over AI principles (Morley et al, 2020). What this would really involve is however far from clear. Schiff and colleagues appear to note that AI guidelines should make "reference to specific laws and policies rather than merely mention the general importance of complying with law" (Schiff

et al, 2020). The distinction between "reference" and mere "mention" is not quite straightforward, but it is clear that AI principles should build on and engage with established legal standards. Missing in the analysis however is what those legal standards are: is the allusion to those enacted at the national level or regional and international levels? One of the strengths of IBRs is, as already noted, that established legal standards are routinely invoked. What principlists should, then, aim to do is to go beyond mere "reference" to laws, and align AI principles with principles that are firmly grounded in pertinent national and international legal standards.

Another dimension of substantive legitimacy relates to the degree to which principlists draw upon a diverse set of normative sources. Drawing upon values of diverse cultures would certainly enhance the legitimacy of AI guidelines. This is a domain where IBRs did not fare well, as highlighted in Section 3. The principlist approach seems to be headed the same way in this regard. At one level, it is said to draw exclusively upon liberal political thought and hence excludes non-western traditions (Nyrup, 2021; Adams, 2021; Mohammed et al, 2020; Hassan, 2022; Eke et al, 2023). Eke and Ogoh, for instance, argue that current AI ethics and governance discourse excludes 'African narratives' Eke and Ogoh, 2022). Indeed, none of the surveys of global AI initiatives have uncovered AI ethics guidelines from, for example, Africa and Latin America (Jobin, 2019; Corre^a et al, 2023). This widely perceived problem in AI ethics cannot easily be dismissed. It would ultimately reduce the substantive legitimacy of AI principles as building blocks of AI governance.

But the process seems to be equally problematic, thereby raising concerns of procedural legitimacy. That is also another area where IBRs offer useful "positive" lesson. This is particularly with regard to the experiment with the "multi-stakeholder" approach to the development of norms, reinforced by crowdsourcing mechanisms (Yilma, 2022). The IBRs project is often hailed for successfully experimenting an inclusive way of norm creation that would allow drawing of inputs from 'all' stakeholders in the process of norm creation (Kulesza, 2014). Neither multi-stakeholderism nor crowdsourcing are without limitations, however. Crowdsourcing might, for example, open the door for poor quality input in the absence of some form of quality control (Yilma, 2022). Likewise, none of the existing fora where multi-stakeholderism has been a modality of governance is perfect. A number of factors such as the digital divide, resource constraining and expertise hamper the realisation of a truly multi-stakeholder forum of global governance (Hofmann, 2016). But generally, the experiment, or at least the aspiration for, with an inclusive approach to the making of norms is a valuable lesson to the principlist approach to AI ethics.

AI ethics processes are widely criticised for not accounting for diverse voices (Erman and Furendal, 2022). Ó hÉigeartaigh and colleagues note that the global AI ethics discourse is taking place in, and between, countries with significant roles in the development of AI in the exclusion of countries to which these technologies are exported (Ó hÉigeartaigh et al, 2020; Hickok, 2021). The concern, as some scholars suggest, is not just that AI principles of western origin are not generalisable to the rest of the world (Stahl, 2021). The greater concern is that credibility of the principlist approach will be at risk as a whole, lending weight to radical claims about the advent of "ethical colonialism" (Piedra, 2023).

Except for legislative bodies in democratic societies, no normative process enjoys full legitimacy of course. Outside national legislatures, the closest to an adequately inclusive process exist in intergovernmental organisations, particularly at the UN, where every member State is represented.

The question of whether this is the direction AI ethics should follow is beyond the scope of this paper, but we flag below how an institutional process may help lessen the looming legitimacy deficit in AI ethics.

4.3. Institutional process

Problems of fragmented normativity and (il)legitimacy discussed above emanate mainly from the lack of what we refer to in this paper as an "institutional process". An "institutional process" broadly refers to institutional arrangements that help structure the global conversation on AI ethics expressed through principles. Ideally, an institutional process would help mitigate not only the problems of fragmentation and (ill)legitimacy but also keep the principlist approach relevant in face of changing developments. In the latter case for instance, an institutional process would help the principlist approach to draw lessons from other approaches to governance such as IBRs.

Dearth of an institutional anchor has been a major flaw of the IBRs project. Except for IBRs launched by national governments such as in Brazil, IBRs emerged largely outside formal institutional arrangements. The vast majority of initiatives for IBRs were launched by civil society organisations. And unsurprisingly, civil society IBRs initiatives have been largely hortatory in focus than aimed at proposing the blueprint for the institutional infrastructure needed for the realisation of rights and principles. That the IBRs project has generally been sporadic, momentary and fragmented movement means that institutional arrangements were unlikely to draw the attention of authors of the documents. Arguably, tendencies of fragmentation alluded to above in the principlist approach are products of the lack of similar institutional processes.

To a varying degree, imperatives of an institutional framework as a means of enhancing the value of AI principles has been espoused in the literature. Ferretii, for instance, emphasises the importance of "validating" AI principles by pertinent government bodies to enhance the democratic legitimacy of private sector-led AI principles (Ferretti, 2022). Jobin even calls for a "deliberative platform", perhaps to debate the content of AI principles (Jobin, 2019). Others vaguely called for an "ethical AI committee" that would monitor the implementation of AI principles in AI development and deployment (Zhou and Chen, 2023). But little is offered on the nature, scope and functions of the proposed institutional arrangements: it is not clear, for example, what "validating" principles would constitute; what would be role of the "deliberative" body; and who would create the proposed committee, and at what level (national, regional, international or industry level).

Some form of institutional process would be instrumental in at least four respects. At the most basic level, it would help clarify the mission of the principlist approach. As highlighted above, AI principles—just like IBRs—are emerging with each proponent advancing values that are aligned with its interests. The existence of an institutional anchor would help not only in defining a goal of the principlist project but also in overseeing whether that goal is being met. This would save it from the fate of IBRs which, as things currently stands, has faded into the background or has become a recent preoccupation of legal academics under the name of "digital constitutionalism".

Second, it would tackle one of the common limitations of AI principles, for example generality and tensions. Overly generic formulation of principles results in vagueness and ambiguity thereby opening the door for multiple interpretations (Munn, 2023). What compounds this concern is that AI

guidelines are riddled with concepts such as fairness and justice that are inherently prone to contentious understanding (Whittlestone et al, 2019). Add to that the inevitable tensions between competing principles such as transparency and privacy (Whittlestone et al, 2019). An institutional process of some sort would go a long way in addressing conceptual ambiguities and tensions. In this regard, the exercise in novelty in the IBRs project flagged above, explored further in Section 4.4. will be instructive. IBRs tend to frame rights in a rather elaborate manner which enables clearer stipulation of rights and principles. Corollary to this is defining or specifying the respective role of actors involved in the realisation of rights or implementation of principles. This would help to frame AI principles in a more nuanced and context-specific manner, thereby further enhancing their implementability.

Third, an institutional process would also ensure that AI principles attend to the most pressing ethical issues of AI. Recent surveys have highlighted omissions of important themes in the major ethics guidelines such as the impact of AI in reducing social cohesion and serving as a tool of political abuse (Hagendorff, 2020). Such omissions are inevitable when AI principles emerge in a fragmented manner. Related to this is that the proposed institutional process may help streamline emergent attempts to translate principles to practice. It may, for instance, help identify principles on which there is adequate consensus and conceptual clarity so as to move to the next step of embodying them into design or even legislation. This is corollary to the point made above that an institutional process would contribute towards addressing problems associated with ambiguities and vagueness in AI principles. In that sense, the institution(s) would distill principles on which there is settled and clear meaning so that they could be taken to the process of translating them into technological design and beyond.

Finally, an institutional process can be one way of addressing legitimacy deficits in AI ethics discussed above. Through such a process, a mechanism that gradually ensures meaningful representation and contributions of stakeholders in pertinent AI ethics platforms can be instituted. Better representation would gradually enable fashioning AI guidelines taking diverse cultures and stakeholders into account. Strengths of IBRs, in this regard, in enhancing procedural and substantive legitimacy discussed in the preceding section offer valuable lessons.

What remains is the question of who would lead the institutional process and at what level. Such bureaucratic details are beyond the scope of this article. Preliminary points can be made, however. Given the fact that AI is a global phenomenon, the ideal institutional arrangements should be set up at the global level. AI institutions exist at the national and regional level, and play crucial roles in overseeing the ethical design, use and governance of AI. In recent years, governments around the world are establishing national AI entities, be it in the form of "AI Institutes" (e.g. UK) or authorities (e.g. Saudi Arabia). But bringing forth a truly legitimate framework of AI principles would require the involvement of entities with the mandate and the expertise to lead global discussions on AI ethics. We have already alluded above the important role of UNESCO and the OECD, global entities with the mandate and domain expertise, in leading the charge. That way, the principlist approach to AI ethics could find the much-needed institutional anchor.

4.4. Exercise in novelty

The first and third lessons discussed thus far are "negative" lessons that the principlist approach should learn from but overcome in order to evolve into a meaningful mechanism of AI governance. The second lesson, i.e. one that relates to legitimacy, is a "positive" that principlists are advised to follow. There is also another aspect of the IBRs project from which principlists should draw useful lessons. The lesson relates to the exercise in novelty of the IBRs project flagged in Section 3. This relates to the ways in which IBRs creatively frame rights and principles in an elaborate manner and define the respective role of actors responsible for the realisation of rights and principles.

While most IBRs documents appear to "reaffirm" existing rights in the digital context, a set of novel rights that find no direct counterparts in the present human rights catalogue are also put forward. Specifying the scope of human rights is prompted by the long-established tradition of formulating rights in generic terms. Protagonists of the IBRs project maintain that there are contextual factors unique to cyberspace which require unpacking the normative contours of human rights in light of the new realities of the digital environment (Casacuberta and Senges, 2008). To specify generic rights and principles, many IBRs define rights and principles in an elaborate manner. This is particularly the case with respect of the right to privacy and freedom of expression.

In IBRs, the right to privacy is often specified into what are sometimes referred to as "subset of rights" such as the "right to use encryption technologies" and "freedom from surveillance" (Charter of Human Rights and Principles for the Internet, 2022: Arts 8-9; Universal Declaration of Digital Rights, 2017: Arts 5-6; Italian Declaration of Internet Rights, 2015: Arts 5-7). This stands in stark contrast with the rather generic formulation of the right to privacy in international law (Universal Declaration of Human Rights, 1948: Art 12; International Covenant on Civil and Political Rights, 1966: Art 17). Not only is the right to privacy in Articles 12 and 17 of the UDH and the ICCPR (respectively) formulated overly generally but also that the legitimate grounds for restriction are not stipulated (Yilma, 2023). By specifying the right into a subset of rights and envisaging the grounds for restriction, IBRs addresses the longstanding problems of international privacy law. Following this path, the post-Snowden privacy resolutions of the UN clearly recognise the widely accepted grounds for restriction, i.e. legality, necessity and legitimacy (UNGA Res 77/211, 2022).

Similarly, freedom of expression is elaborated into a set of sub-rights. For instance, the Charter of Human Rights and Principles on the Internet, an IBRs launched through one of the dynamic coalitions in the UN Internet Governance Forum, enshrines a set of rights under freedom of expression. Among the subsets of rights include "freedom of online protest", "freedom from censorship", "right to information" and "freedom from hate speech" (Charter of Human Rights and Principles for the Internet, 2022: Art 5). Doubtless, freedom of expression is formulated in a far more elaborate manner than the right to privacy in international human rights law (Universal Declaration of Human Rights, 1948: Art 19; International Covenant on Civil and Political Rights, 1966: Art 19). And indeed, the Human Rights Committee, a UN treaty body charged to oversee the ICCPR, has developed much elaborate jurisprudence around freedom of expression (General Comment 34, 2011). Yet, freedom of expression in international human rights law, and the attendant jurisprudence, do not come close to how elaborately that it is envisioned in IBRs.

As highlighted in Section 2, the high-level nature of principles and hence the attendant conceptual ambiguities and interpretive problems is a widely discussed limitation of the principlist approach (Munn, 2023); Mittelstadt, 2019). Indeed, this is a problem inherited by principlists from principlism

in medical ethics. The four foundational principles of medical ethics have long been criticised for being vague, abstract and "thin in content" (Lee, 2009; Clarke, 2009). It is here where exercise in novelty in IBRs offer a valuable lesson. Following IBRs' approach to elaborative and contextualised formulation of rights and principles is one way of addressing the over-generality AI principles. Drawing from the experience of IBRs, the next phase in the principlist approach to AI ethics should be, inter alia, to bring about clarity of concepts underlying AI principles. This would help AI ethics move past the path dependencies and conceptual constrictions created by medical ethics whose age-old general principles of beneficence, non-malfeasance, autonomy and justice defined the normative contours of AI principles.

AI principles widely enshrined in AI ethics initiatives should be reimagined to make them fit for purpose. Moving past generic formulation, the specific contexts in which they apply would need to be elaborated. Achieving this goal is not an easy exercise of course, and laying out the mechanics of how it would be achieved is beyond the scope of this paper. But one way in which this may be pursued is through the institutional process discussed above. An institutional anchor would ideally identify aspects of AI principles which require clarity and specificity. Tying the task of elaborating vague AI principles to institutions that enjoy some legitimacy is vital. Importantly, it would increase the prospect of such interpretations of principles and concepts such as fairness to be accepted by AI developers.

As discussed in Section 3, the other sphere of novelty in the IBRs project is the elaborative way in which it defines the role of state and non-state actors in the respect and protection of rights. Importantly, IBRs tend to directly address corporations, precisely technology corporations, as duty bearers alongside states. Contract for the Web offers an illustrative example. The Contract is an IBRs which resulted from the campaign led by 'inventor of the web' Tim Berners-Lee who, in the wake of the Snowden revelations, called for a "magna carta" for the digital space (Contract for the Web, 2019; The Guardian, 2014). Not only does it define the role of states but also technology companies and citizens. Principle 5, for instance, provides that technology companies "will respect and protect people's privacy and personal data to build online trust". This proviso is elaborated further in the Contract (Contract for the Web, 2019: Principle 5). This departs significantly from the current structure of international human rights law where corporations owe no human rights obligations, but subject merely to high-level and non-binding corporate social or moral responsibility. Perhaps due to the influence of IBRs, this approach is now adopted, to a degree, in post-Snowden privacy resolutions where technology companies are specifically and directly "called upon" to take a series of steps to uphold privacy rights (UNGA Res 77/211, 2022).

The principlist approach to AI ethics can benefit from this progressive feature of the IBRs project. Following the steps of IBRs, AI ethics should clarify who shall do what and when in the course of researching, developing, selling, acquiring, using and even governing AI. A recurring challenge in recent attempts to apply principles in, for example technological design, is difficulty of translating into code complex principles such as fairness and justice. Value-laden concepts such as fairness are inherently prone to divergent interpretation but also the respective role of AI companies as well as states and other stakeholders remain unspecified. A fairly elaborate and contextualised formulation of such principles is one way of tackling this challenge.

In practice, this would mean defining, in clear terms, the respective role of developers, users, sellers or other actors involved in the AI life cycle vis-a-vis a particular value enshrined in ethics guidelines. This would involve, for example with respect to privacy, defining what it means, scope and the respective responsibilities of AI developers and relevant state bodies. Such an elaborate approach would bring clarity not only to AI developers or companies but also entities tasked to oversee the implementation of the principles. To be sure, there are some signs of a move towards defining the role of technology companies in AI ethics initiatives. A notable example, in this regard, is the Toronto Declaration which lays out principles for the protection of rights to equality and of non-discrimination in machine learning systems (Toronto Declaration, 2018). It briefly considers the "responsibilities of private sector actors", but merely reiterates the non-binding UN based framework of human rights due diligence (Guiding Principles on Business and Human Rights, 2011). It does not, in that sense, introduce tailored duties owed by companies operating machine learning systems.

But who should lead the charge in making AI principles fairly elaborate, clear and amenable to effective technological design and governance? This, again, is a question that cannot be addressed in this article. But we contend that the institutional process discussed above would be critical in leading the effort towards making AI principles clearer. Considering the question of which institutions are best placed in this regard goes beyond the focus of this article. But entities such as UNESCO and OECD which, as already submitted with respect to the other lessons, have the mandate and experience in AI ethics are pertinent in facilitating the ways in which the principlist approach may initiate the conversation towards unpacking AI principles. This reinforces the common thread in this article that the principlist approach to AI ethics should move from articulating principles to setting in motion an institutional process.

5. Conclusion

In his defense of the principlist approach, Lundgren argues that there is still room for the "methodologically adolescent" AI ethics guidelines to overcome their shortcomings (Lundgren, 2023). And of course, there has been some effort to point towards overcoming the shortcomings in the past few years, including by drawing upon other cognate disciplines such as public health ethics. This paper was the first original attempt in that direction by exploring ways in which the principlist approach may learn from the rather niche but closely related field of IBRs.

We argued that the prospect of the principlist approach in bringing about meaningful impact in policy formulation or technology design could be enhanced by averting weaknesses and embracing strengths of IBRs. We explored four interrelated areas where IBRs offer valuable lessons to AI ethics. At the core of these is an institutional process that would be crucial not only to structure the rather fragmented AI ethics discourse but also to enhance the legitimacy and novelty of the principlist approach. Unless principlists learn from IBRs, an otherwise adolescent process may not only risk failing to reach maturity but fall gradually into obscurity.

Recent initiatives at the international level wield a potential to reverse this fate of the principlist approach, however. One such initiative is the ongoing discourse at the UN Human Rights Council under the label "new and emerging digital technologies and human rights" (HRC Res 53/29, 2023). On top of embodying widely accepted AI principles, the series of resolutions adopted as part of this initiative emphasise the importance of institutional arrangements in upholding human rights in the

conception, design, and deployment of AI technologies. Such initiatives might help shift the focus of the principlist approach to AI ethics from rolling out 'principles' to sparking a 'process' that would help not only implement, but also clarify, sustain and legitimise them.

_//_

Acknowledgements: The author grateful gratefully thanks Stephen Cave for helpful discussion on an earlier version of this paper.

Funding Information: The author's research for this work benefited from funding provided by Magdalene College, University of Cambridge and Standard Bank Africa.

References

Adams, R. (2021) 'Can Artificial Intelligence Be Decolonized?', 46 Interdisciplinary Science Reviews 176.

African Declaration of Internet Rights and Freedoms (2014) <https://africaninternetrights.org/> (Last accessed 25 January 2025).

AI Ethics Guidelines Global Inventory (Algorithm Watch Global, 2024) <https://inventory.algorithmwatch.org/about> (Last accessed 25 January 2025).

AI Principles: Progress Update (Google, 2023) <https://shorturl.at/C2unU> (Last accessed 25 January 2025).

An Online Magna Carta: Berners-Lee Calls for Bill of Rights for Web (The Guardian, 12 March 2014) <https://shorturl.at/aHS32> (Last accessed 25 January 2025).

Auld, G. et al (2022) 'Governing AI Through Ethical Standards: Learning from the Experiences of Other Private Governance Initiatives', 29 *Journal of European Public Policy* 1822.

Barnard, F., Democratic *Legitimacy: Plural Values and Political Power* (McGill-Queen's University Press, 2001).

Besson, S., *The Morality of Conflict: Reasonable Disagreement and the Law* (Hart).

BVerfG, NJW 2008, 822 (2008).

Casacuberta, D. and Senges, M. (2008) 'Do We Need New Rights in Cyberspace? Discussing the Case of How to Define Online Privacy in an Internet Bill of Rights', 40 *Enrahonar* 99.

Charter of Human Rights and Principles for the Internet, 8th edition (Internet Rights and Principles Coalition, 2022) <https://shorturl.at/nOaYZ> (Last accessed 25 January 2025).

Celeste, E. (2018) 'Digital Constitutionalism: A New Systematic Theorisation', 33 *International Review of Law, Computers and Technology* 76.

_____, *Digital Constitutionalism: The Role of Internet Bills of Rights* (Routledge, 2023).

Clarke, DL (2009) 'Moral Principlism Alone is Insufficient, and Traditional Moral Theories Remain Important for Practical Ethics, 2 *SAJBL* 12.

Coeckelberg, M., *AI Ethics* (MIT Press, 2020).

Continental AI Strategy of the African Union (July 2024).

Contract for the Web (2019) <https://contractfortheweb.org/> (Last accessed 25 January 2025).

Correˆa, N. et al (2023) 'Worldwide AI Ethics: A Review of 200 Guidelines and Recommendation for AI Governance', 4 *Patterns* 1.

Cromartie, A., 'Legitimacy' in R. Bellamy and A. Mason, *Political Concepts* (Manchester University Press, 2003).

Eke, D. et al, 'Introducing Responsible AI in Africa', in D. Eke et al (eds) *Responsible AI in Africa: Challenges and Opportunities* (Springer, 2023).

Eke, D. and Ogoh, G. (2022) 'Forgotten African AI Narratives and the Future of AI in Africa', 32 *International Review of Information Ethics* 1.

Erman, E. and Furendal, M. (2022) 'The Global Governance of Artificial Intelligence: Some Normative Concerns', 9 *Moral Philosophy and Politics* 267.

Ethics Guidelines for Trustworthy Artificial Intelligence (High-Level Expert Group on AI, 8 April 2019) < https://shorturl.at/3gKO8> (Last accessed 25 January 2025).

Executive Order (E.O.) 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (White House, October 2023) <https://shorturl.at/oqDEG> (Last accessed 25 January 2025).

Fjeld, J. et al, 'Principled AI Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI', Research Publication No. 2020-1 (2020).

Floridi, L., *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities* (Oxford University Press, 2023).

Fourneret, E. and Yvert, B. (2020) 'Digital Normativity: A Challenge for Human Subjectivation', 3 *Frontiers in Artificial Intelligence* 1.

General comment No. 34, Article 19: Freedoms of Opinion and Expression, Human Rights Committee (2011).

Gill, L. et al, 'Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights', Berkman Center Research Publication No. 2015-15 (2015).

Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework, UN Doc A/HRC/17/31 (21 March 2011) <https://shorturl.at/koSV0> (Last accessed 25 January 2025).

Guiding Principles on Trusted AI Ethics (Telia, 2019).

Hagendorff, T. (2020) ‘The Ethics of AI Ethics: An Evaluation of Guidelines’, 30 *Minds and Machines* 99.

Hassan, Y. (2023) ‘Governing Algorithms from the South: A Case Study of AI Development in Africa’, 38 *AI and Society* 1429.

Hickok, M. (2021) ‘Lessons Learned from AI Ethics Principles for Future Actions’, 1 *AI and Ethics* 41.

Humphrey, J. (1976) ‘The International Bill of Rights: Scope and Implementation’, 17 *William & Mary Law Review*, 527.

Italian Declaration of Internet Rights (2015) <https://shorturl.at/inzGU> (Last accessed 25 January 2025).

International Covenant on Civil and Political Rights, UNGA 2200A (XXI) (16 December 1966).

Jobin, A. et al (2019) ‘The Global Landscape of AI Ethics Guidelines’, 1 *Nature Machine Intelligence* 389.

Hofmann, J. (2016)‘ Multi-stakeholderism in Internet Governance: Putting a Fiction into Practice’, 1 *Journal of Cyber Policy* 29.

Kulesza, J., ‘Protecting Human Rights Online: An Obligation of Due Diligence’, New York University School of Law Jean Monnet Working Paper No. 24/14 (2014).

Lander, E. and Nelson, A., Americans Need a Bill of Rights for an AI-Powered World (Wired, 8 October 2021) <https://shorturl.at/gjvOS> (Last accessed 25 January 2025).

Lee, M., (2009) 'The Problem of ‘Thick in Status, Thin in Content’ in Beauchamp and Childress’ Principlism, 36 *Journal of Medical Ethics* 525.

Lundgren, B. (2023) ‘In Defense of Ethical Guidelines’, 3 *AI and Ethics* 1013.

Marco Da Civil, Law No. 12.965 (2014) <https://shorturl.at/GEswv> (Last accessed 25 January 2025).

Mittlstadt, B. (2019) ‘Principles Alone Cannot Guarantee Ethical AI’, 1 *Nature Machine Intelligence* 501.

Mohammed, S. et al (2020) ‘Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence’, 33 *Philosophy & Technology* 659.

Morley, J. et al (2020) ‘From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices’, 26 *Science and Engineering Ethics* 2141.

Munn, L. (2023) ‘The Uselessness of AI Ethics’, 3 *AI and Ethics* 869.

Musiani, F. (2009) ‘The Internet Bill of Rights: A Way to Reconcile Natural Freedoms and Regulatory Needs?’ 6 *Scripted* 504.

New and Emerging Digital Technologies and Human Rights, HRC Res 53/29 (18 July 2023).

Nigerian Digital Rights and Freedom Bill (2018) <https://shorturl.at/hoWLs> (Last accessed 25 January 2025).

Nyrup, R. (2021) ‘From General Principles to Procedural Values: Responsible Digital Health Meets Public Health Ethics’, 3 *Frontiers in Digital Health* 1.

Ó hÉigeartaigh, S. et al (2020) ‘Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance’, 33 *Philosophy and Technology* 571.

Palmer, A. and Schwan, D. (2024) ‘More Process, Less Principles: The Ethics of Deploying AI and Robotics in Medicine’, 33 *Cambridge Quarterly of Healthcare Ethics* 121.

Piedra, J. (2023) ‘Decolonizing the Ethics of AI’, 19 *Cosmos and History: The Journal of Natural and Social Philosophy* 467.

Prem, E. (2023) ‘From Ethical AI Framework to Tools: A Review of Approaches’, 3 *AI and Ethics* 699.

Prime Minister sets out blueprint to turbocharge AI: Press Release (Government of UK, 2025) <https://shorturl.at/fKgfj> (Last accessed 25 January 2025).

Principles of Robotics (Engineering and Physical Sciences Research Council, 2010) <https://t.ly/T2th6> (Last accessed 25 January 2025).

Raz, J. (1972) ‘Legal Principles and the Limits of Law’, 81 *The Yale Law Journal* 823.

Raz, J. (1984), ‘On the Nature of Rights’, 93 *Mind* 194.

Resseguier, A. and Rodrigues, R. (2020) ‘AI Ethics Should not Remain Toothless! A Call to Bring Back the Teeth of Ethics’, 7 *Big Data & Society* 1.

Saudi Arabia AI Ethics Principles (Data and AI Authority, 2023) <https://shorturl.at/PooNR> (Last accessed 25 January 2025).

Schiff, D. et al, ‘What’s Next for AI Ethics, Policy, and Governance? A Global Overview’, AIES’20, February 7–8, 2020, New York, NY, USA (2020).

Seger, E. (2022) ‘In Defense of Principlism in AI Ethics and Governance’, 35 *Philosophy and Technology* 1.

Seppälä, A. et al, ‘From Ethical AI Principles to Governed AI’, Forty-Second International Conference on Information Systems, 12-15 December 2021, Austin, USA (2021).

Six Principles for Developing and Deploying Facial Recognition Technology (Microsoft, December 2018) <https://shorturl.at/bcjvJ> (Last accessed 25 January 2025).

Stahl, B., *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies* (Springer, 2021).

Stillman, P. (1974) ‘The Concept of Legitimacy’, 7 *Polity* 32.

Tech giants are putting $500bn into 'Stargate' to build up AI in US (BBC News, 22 January 2025) <https://www.bbc.co.uk/news/articles/cy4m84d2xz2o> (Last accessed 25 January 2025).

The Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People (White House, October 2022) <https://shorturl.at/abjn1>.

The NSA Files (The Guardian, 2013) < https://www.theguardian.com/us-news/the-nsa-files> (Last accessed 25 January 2025).

The Right to Privacy in the Digital Age, UNGA Res 77/211 (17 December 2022).

The Toronto Declaration: Protecting the Right to Equality and Non-discrimination in Machine Learning Systems (Access Now and Amnesty International, 16 May 2018) <https://shorturl.at/mnEGQ> (Last accessed 25 January 2025).

UK A Pro-innovation Approach to AI Regulation: White Paper (August 2023).

Umbrello, S. et al (2021) 'Value Sensitive Design to Achieve the UN SDGs with AI; A Case of Elderly Care Robots', 31 *Minds and Machines* 395.

Universal Declaration of Digital Rights (Article 19, 2017) <https://shorturl.at/huDGI> (Last accessed 25 January 2025).

Universal Declaration of Human Rights, UNGA 217 A (10 December 1948).

Wedgwood, R., *The Nature of Normativity* (Oxford University Press, 2007).

Whittlestone, J. et al, 'The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions', AIES'19, January 27–28, 2019, Honolulu, HI, USA (2019).

Yilma, K. (2022), 'Bill of Rights for the 21st Century: Some Lessons from the Internet Bill of Rights Movement', 26 *International Journal of Human Rights* 701.

_____ (2017), 'Digital Privacy and Virtues of Multilateral Digital Constitutionalism—Preliminary Thoughts', 25 *International Journal of Law and Information Technology* 115.

_____ (2023) *Privacy and the Role of International Law in the Digital Age* (Oxford University Press, 2023).

Zhou, J. and Chen, F. (2023) 'AI Ethics: From Principles to Practice', 38 *AI Ethics and Society* 2693.