AI & SOCIETY
https://doi.org/10.1007/s00146-025-02461-0

OPEN FORUM

We need a Belmont report for AI

Kevin Patton[1]

Received: 1 April 2025 / Accepted: 23 June 2025

Abstract
The Belmont Report has a place of great importance in American biomedical research ethics. This paper argues that a similar kind of report, and the legal infrastructure that birthed it, is needed in the United States if we are to preempt a great many of the potential issues that are on the horizon with artificial intelligence (AI). What makes the Belmont Report so important is not just that it established a new basis for how medical professionals ought to treat their patients and experiment participants; it did so with the force of law. Establishing an equivalent legal framework for AI is going to take tremendous buy-in from a variety of private and public actors in the United States. The model afforded by the Belmont Report is well suited to generate such buy-in. While this may seem like a daunting task given various polarizing issues at play in society today, the context that produced the Belmont Report was quite fractious itself. It is the position of this paper that a similarly styled approach to AI regulation can succeed in proactively limiting the harms of AI's use (and abuse).

Keywords Artificial intelligence · Ai ethics · Ethics · Medical ethics · Principlism

1 Purpose

The AI landscape, as will be described shortly, is quite a messy one. Every few weeks, some new feature is announced by a technology company eager to attract new users in its quest for AI dominance. Regulatory frameworks have been sparse in the United States, while the European Union (to its credit) has been quite proactive in the regulatory framework that it has enacted. The purpose of this paper is to serve as a call to action for us in the United States to develop a legal and ethical framework for the regulation of AI. The thesis of this paper is that a workable federal regulatory framework can be arrived at by modeling it after the *National Research Act* and the *Belmont Report*. The power of this model is in its ability to fuse ethical principles with national law. It is the position of this paper that such a fusing is needed for AI as well. I will develop this idea by first sketching the AI landscape (which I call "the wild west"). Then, I will sketch the lead up to both the National Research Act and the Belmont Report.. I then close the paper by noting two practical benefits of utilizing this model for AI legislation.

✉ Kevin Patton
kevinpatton@unomaha.edu

[1] University of Nebraska at Omaha, Omaha, United States

2 The wild west

I don't think it is too far off to describe the current artificial intelligence landscape as a kind of "wild wild west". In terms of generative AI, the biggest players in the U.S. are well known: Google, OpenAI, and Anthropic PBC. In terms of AI more broadly, the biggest players also include Facebook, Amazon, Microsoft, IBM, and many others. Additionally, companies such as Baidu, Alibaba, and Tencent (BAT), and DeepSeek are some of China's most prominent companies. Each player in this space is seeking not only to capture monthly subscriptions from individual and corporate users, but are also seeking to establish themselves as *the* leader in AI. Being a leader in this space requires vast sums of capital, land, and access to a supply chain that can source, manufacture, and deliver tens of thousands of graphics cards, servers, and more. The electrical and cooling demands of these AI systems is on a scale that is equally vast. With regard to electricity alone, GoldmanSachs estimates these systems will require 24GW of power by 2030, whereas others have predicted that, with an increase in AI-specific chips, the power requirements could exceed 300GW of power.[1] To put that into perspective, California has a total current power capacity of 86GW. With regard to worries about climate change, Eric Schmidt (one of the founders of

[1] For a comprehensive report which looks at various projections, see: Pilz and Heim 2025.

Published online: 05 July 2025

Google), has said "We're not going to hit the climate goals anyway because we're not organized to do it."[2] The takeaway of this comment was that we should go all in on AI even if it accelerates climate change (in part because AI may help us solve climate change).

In addition to worries about the climate, many have worries related to bias and prejudice being "trained into" these powerful AI systems. An AI system is only as good as its training data, and if that training data has certain bias "built into" it, then the AI systems will perpetuate those biases at scale. That this is a problem can be seen by looking at the programs for various conferences on AI ethics. The 2024 *Artificial Intelligence, Ethics, and Society* (AIES) conference, for example, had over twenty speakers and over fifty poster presentations. The vast majority of those talks and posters focused on some specific way(s) in which AI systems either perpetuate negative stereotypes, would discriminate against certain racial or ethnic groups, or how current policies fail to address biased training data.[3]

In addition to all of this, there is a patchwork of legal frameworks at play. The European Union's *AI Act* offers, by my lights, the most robust attempt to implement standards and mechanisms of enforcement for AI systems. The Act breaks down AI systems by levels of *risk* where risk is understood as being a risk to human rights. The more rights that an AI system can impact, the more riskier it is. There are a total of four levels of risk, with each increasing level becoming more and more regulated. The final level involves risks too great to accept, and thus bans certain uses of AI. The United States, in contrast, has very little regulatory frameworks designed specifically for AI at the federal level. Certain existing laws regarding, for example, privacy in healthcare automatically apply to AI usage. Additionally, some states, such as California, have recently passed a series of laws regarding digital likeness and transparency. Until a full legal framework is passed at the federal level in the United States, however, the current patchwork will have serious gaps. American AI companies that provide their services to citizens of EU member countries, for example, are bound by the AI Act. This does nothing, however, to prevent those companies from acting in risky ways elsewhere. I will come back to this issue below. The AI companies themselves have offered their own codes of ethics and best practices, and organizations such as the *Partnership on AI* exist to try to bring various players in the AI space together to agree on responsible AI practices. The major issue with this particular kind of approach is that it is opaque to outsiders how such policies are enforced—if they are enforced at all. Without third-party (government or otherwise) assessments, such codes and agreements fail to ensure compliance.

My particular focus in this paper is the United States. We citizens, our politicians, and the AI companies which are based herein must draw some lines in the sand with regard to AI. While the EU's AI Act is impressive, most of the big names in AI cited above reside *outside* of the EU. In order to address this inherent limitation, the United States must take action. If the United States were to establish a novel legal framework at the federal level, then this would go a long way to taming the wild west. Companies based in the United States cannot circumvent federal law in the ways that they can when doing business outside of the US. As a first step toward this, I propose a clear but difficult goal: we must arrive at a set of ethical principles which can serve as the basis for good legislation. This might sound a bit like a platitude, but I do not think that it is. What I have in mind is not just some idealistic list that everyone agrees to, and then goes back to their respective companies and does whatever will make the most money. The best way to explain the kind of principles that I have in mind is by analogy with another domain that has utilized a principles-based approach in response to serious moral failings: biomedical ethics.

3 A brief history

In the later half of the twentieth century, a fairly stark shift occurred within western medical research away from what is called *paternalism*. Paternalism in medical research was, roughly, the view that the researcher occupied a role similar to that of a parent, whereas the subject occupied a role similar to that of a child. Under the paternalistic model, researchers were not required to properly inform subjects on diagnoses or the reasons for certain treatments. The researcher (parent) knows best, and the subject (child) needs to simply let the researcher do their work. Such a view of the researcher-subject relationship is not as common anymore, though it has not been fully eliminated in medical practice more generally.[4] Despite that, paternalism in research is largely a relic of the past. It was supplanted by an approach which prioritizes four ethical principles. These four principles are: respect for persons (autonomy), acting in the best interest of the patient (beneficence), avoiding harm or injury to the patient (non-maleficence), and a focus on the equitable distribution of healthcare resources while treating patients fairly (justice).

[2] The interview was part of the "Special Competitive Studies Project" which has posted the full recording of the interview on YouTube.

[3] The full program can be viewed online at the AIES webpage for the conference: https://www.aies-conference.com/2024/

[4] For example, see: Fleisje 2023.

The transition to the four principles approach did not happen overnight. The horrors of the Second World War played an enormous role in shocking not just medical professionals in allied countries, but the public at large. The egregious violations of consent and well-being by Nazi doctors, exposed at the Nuremberg Trials, lead to the *Nuremberg Code*. This ten-point document correspondingly focused heavily on consent and non-harm to patients. A few decades later, Peter Buxtun would expose the Tuskegee Syphilis Study.[5] Shortly after this exposure, the *National Research Act* was passed, and the *National Commision for the Protection of Human Subjects of Biomedical and Behavioral Research* was formed. This commission was charged with determining a set of basic ethical principles from which biomedical and behavioral research could continue. Four years later, in 1979, the commission would publish the Belmont Report. This report identified autonomy, beneficence, and justice as the core ethical principles which ought to govern future research. That same year, Tom Beauchamp and James Childress would also publish their influential book "Principles of Biomedical Ethics". In this book, Beauchamp and Childress argued that biomedical ethics ought to adopt four principles to govern medical ethics. Their list, like the Belmont report, included autonomy, beneficence, and justice. But, unlike the Belmont report, Beauchamp and Childress separated out a non-maleficence condition independent from the beneficence.[6] However, one feels about that particular move, the medical research community has, as a whole, adopted the four principles approach (often called *principlism*). The principlism of Beauchamp and Childress has morphed over the years given certain stinging criticisms. This has helped to both refine and strengthen the view, though the details of those debates are outside the scope of this paper.[7] Having now briefly described how the National Research Act and Belmont Report came into existence, I now need to offer some clarifications for my project.

To some, it may appear that many technology companies have already taken the first step towards a Belmont report for AI. Many have, for example, adopted some set of ethical principles that they claim helps to direct the development and deployment of their various AI systems. Google has their *AI Principles*, Anthropic has their *Claude Constitution* (Claude is the name of their AI product), and OpenAI has their *Safety and Responsibility* guidelines. This, however, is a far cry from Belmont. Without a mechanism for enforcement, such principles fail to compel adherence. Interestingly, the Belmont Report itself is a rather short document with very little detail regarding how to implement the principles espoused therein. The power of the document resides in the laws that were consistent with it. Ethical principles were given a legal counterpart such that an ethical infraction was now a kind of legal infraction. This enforcement piece is currently missing at the federal level. This point is often called *ethics washing*.[8] Ethics washing is typically characterized as a company invoking moral standards merely for the purpose of influencing public perception. To the degree that we want AI regulated effectively, we likely cannot trust technology companies to guard their own hen house. Hence, in order to properly regulate AI we need both ethical principles and legal consequences. The Belmont Report provides a model that satisfies this need nicely. To be clear, the US is not the only jurisdiction that needs to take action on this issue, but it is one of the most powerful. This is especially pressing since many leading AI and AI-related companies are based in the US.

4 Some clarifications

It is important to note that my proposed approach is preventative rather than reactive. The Belmont Report was generated *in response to* horrible events. No such events (to our knowledge) have occurred with AI. In fact, one may argue that since there are no standards for a base-level of decency that exist, then they cannot be ignored (as they were in the Tuskegee case). For this reason, it may be doubted whether anything like a Belmont report is needed.[9] This difference, however, is not a serious problem for my view. One of the motivations for this paper is that we *now* have the (fleeting) opportunity to *prevent* harms from AI rather than (as happened in biomedical research) to act after the fact. The chronological ordering of how the National Research Act and Belmont Report were produced is surely not essential to how we can model an AI equivalent of each. What is needed

[5] While the story caught mainstream attention due to Buxton contacting the newspapers *Washington Star* and the *New York Times*, it is quite incorrect (and shocking) to think that the United States government was intentionally hiding the Tuskegee Study. On the contrary, for nearly four decades, the Public Health Service (the precursor to the CDC) published papers on it every 4–6 years with the first report being in the Journal of the American Medical Association. For a more detailed history, see: Tobin 2022.

[6] The Belmont report, in contrast, states that a beneficent action is one that, in part, "does not harm". There is an interesting philosophical question regarding whether beneficence is logically independent from non-maleficence, or whether this distinction by Beauchamp and Childress is a more pragmatic one. As this is not related to the thrust of this paper, I will ignore this wrinkle.

[7] For a nice summary of some of the historically important developments, see: Arras 2010 and Childress 2007.

[8] See chapters 4, 6, and 7 in the *Oxford Handbook of Ethics of AI* for examples of this worry.

[9] My thanks to an anonymous reviewer for pressing me on this.

for AI is a set of enforceable moral principles. The National Research Act and Belmont Report can serve as a model for that goal, even if the motivations for each are quite different. There is, however, a stronger kind of skepticism which could claim that such preventative federal action is unlikely without sufficient social outrage. The National Research Act and the Belmont report, it will be remembered, arose in response to the public outrage over egregious abuse. Since there is no social outrage against AI, this line of reasoning goes, we should not expect any action in Washington. While I am sympathetic to this practical reality regarding the so-called "gridlock" in Washington, I think there is some reason to view the AI case more optimistically. I first would note that the general public is already aware that AI poses significant risks—even if many members of the public have not used (or don't realize that they are using) the technology themselves.[10] Moreover, current data indicate that most Americans view AI negatively.[11] I take this as evidence that the American public is, in some sense, primed for discussions about AI regulation in a way that may be akin to social outrage. There are certainly not the levels of frustration and outrage that preceded the Belmont Report, but there is a large enough discontent on the issue that the aforementioned pessimism regarding legislation seems more dogmatic than reflective of reality. For these reasons, I take this stronger skeptical worry to be, unless more data can be offered, misplaced.

Another worry about this approach starts by noting that there are literally hundreds of different principlist approaches to AI ethics.[12] The approach I am promoting, it may seem, is no different than those.[13] In response, I want to say two things by way of clarification. To be clear, I am not defending nor promoting any particular principle(s). On that topic, I am remaining neutral. My primary concern regarding that literature revolves around application. What I am promoting is the idea that a legally enforceable set of such guiding principles is a viable approach to AI regulation. The debate about which principles we hold important enough to guide our legislation is an important one, but it must also be recognized that those principles will be useful only to the degree that we can compel others to adhere to them. I view the existence of the National Research Act and the Belmont Report to be a model capable of wedding those two goals.

With those defenses and clarification in place, I now wish to offer two benefits of utilizing a principlist approach to this issue. With the ultimate goal being the passage of fairly sweeping federal law, achieving that goal will require large-scale buy-in by various corporate, political, and citizen actors. A principles-based approach is an excellent candidate to achieve such buy-in due to the following two features.

5 Theory neutral

The first key benefit of adopting a principles-based approach is that it does not require the truth of a larger theory. This feature is what philosophers call *theory neutral*. When something is theory neutral, then its plausibility is not dependent on the success of one theory over its rivals. Whichever moral theory ultimately wins out, ethical values which are theory neutral can be relied on. As an example, let us quickly look at two otherwise rival moral theories: Utilitarianism and Kantianism. Utilitarianism is an ethical theory that focuses on benefiting as many people as possible with our actions. A Utilitarian views the goodness or badness of an action as a function of the consequences of that action. A rival theory, that of Immanuel Kant, claims that only the rational will of the agent has moral worth. That they are rivals is hopefully clear: the Utilitarian focuses on the consequences of actions, and the Kantian focuses on the intentions of the agent. These two theories, while both well supported by a variety of moral intuitions and cases, are in conflict with one another. Given that each theory has a radically different account of moral worth, it would seem that we cannot know what is morally good unless we first determine which theory of goodness is true. This is especially worrying since both theories have been around for a long time, and no side shows any signs of being ready to concede to the other. Now, let us take as an example the principle of respecting autonomy from earlier. Both the Utilitarian and the Kantian will agree that it is good to have such respect. For the Kantian, this is because rational autonomy is the only intrinsically good thing. The Utilitarian, in contrast, will claim that it is because such respect *usually* produces good consequences.[14] The upshot of focusing on theory-neutral principles is that it does not matter which theory wins out.

Now, I am painfully aware that most people do not walk around explicitly utilizing Utilitarianism, or Kantianism, or any formal moral theory. I merely mention them as examples. Swap in whatever political, religious, or otherwise

[10] Pew Research Center 2023.

[11] Maese 2025.

[12] Corrêa 2023.

[13] My thanks to a (second) anonymous reviewer for pushing me on this.

[14] For many Utilitarians, there are cases where such respect can be violated when the consequences are intrinsically good enough, but such cases are claimed to be rare. For example, John Stuart Mill claimed in chapter two of his book "Utilitarianism" that, "It is not the fault of any creed, but of the complicated nature of human affairs, that rules of conduct cannot be so framed as to require no exceptions, and that hardly any kind of action can safely be laid down as either always obligatory or always condemnable.".

conflicting groups in society you wish. So long as there are some values that we can agree to with regard to how we want AI governed, then, despite whatever other conflicts we have, we should be able to make headway on national legislation. Leaders in AI, politicians, and regular citizens are going to need to agree (or at least to not object) to certain ethical values being of prime importance so that we can formulate laws to govern AI usage. The principlist approach can meet this challenge.

6 Morally flexible

The second key benefit of the principlist approach is that the duties which arise from our principles need not be conceived of as *absolute*. By this, I mean that given certain salient features of a given circumstance, certain duties can be overridden. The kind of principlism that I advocate for views the duties that arise from our principles as *prima facie* duties.[15] A duty is *prima facie* when the duty has, on the face of it, a claim to normativity (i.e. it seems like it is correctly telling you how to act). There are sometimes other reasons, however, that can override this *prima facie* normativity. To take an overly simplistic scenario, suppose an unconscious patient arrives at the emergency room. Such a patient cannot consent to the treatment prescribed by the doctor. Is the doctor violating the patient's consent in a problematic way? Typically, the answer to this is "no". Respect for autonomy does not vanish in this case, but the typical weight that the duty carries is now overridden due to these other features of the case. The doctor will likely act in a way that they think the patient would want them to, and that usually involves acting in a way to benefit the patient's overall health (i.e. to act beneficently). *All things considered*, then, in this case the duty of beneficence outweighs the duty to respect autonomy. Thus, a *prima facie* account of duties first arrives at a set of duties which are normative for the domain to which they apply, and then determines which duties weigh most heavily after the specific features of the case are known. You cannot, as it were, do armchair ethics with *prima facie* duties. The details matter.

While AI is not a researcher and the public are not its subjects, viewing the ethical principles that we need to agree on as *prima facie* is useful in the context of AI as well. Given that there will be so many varied applications of AI to our lives, it will likely turn out that the weighting we have of certain values in one area will not map neatly into another. This will be due to the specific details of each application of AI and the ways in which such applications will impact our lives. This is completely consistent with a principles-based approach. The move from *prima facie* to *all things considered* requires the incorporation of all circumstantially relevant details. How such values will be weighted in, for example, the implementation of AI into medical diagnosing is likely to be very different than how those same values will be weighted in cases where AI is utilized to determine employability. In fact, it may turn out that for different applications of AI, there are just different values that we agree to. The kind of cases that will arise due to AI being utilized in medicine, for example, may simply fall under the current principles already at play in contemporary medical ethics. With regard to AI being utilized in decisions about employment, however, principles of medical ethics may not be applicable. AI is malleable with regard to its uses since the data that it is trained on can come from nearly any domain. Hence, having different groupings of relevant ethical principles given different domains of applications is perhaps inevitable. This may make something like a Belmont Report for AI a more complicated and multifaceted project. All the more reason to get started now.

7 In conclusion

In this paper, I have argued that the National Research Act and Belmont Report can serve as a model form which AI regulation can follow. What we need for effective AI regulation is both ethical consensus among varying parties and external enforcement. With regard to the former, there is no shortage of candidates on offer in the literature. With regard to the latter, however, either we trust the companies to regulate themselves or we must rely on legislation. I take that most are hesitant to trust these companies to self-police effectively, and so we must focus on legislation. Ethics and legislation thus comprise the two sides of the AI regulation coin. To that end, I propose that we should establish a National AI Act with a corresponding document expounding the relevant ethical principles at play.

Author contributions K.P. was solely responsible for the conception and writing of this paper. K.P. also approved the final version of the manuscript to be published and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

[15] For the classic treatment of *prima facie* duties in ethics, see: Ross 1930.

References

AI, Ethics, and Society — home. (n.d.). https://www.aies-conference.com/2024/

Arras J. (2010) Theory and bioethics. https://plato.stanford.edu/archives/sum2013/entries/theory-bioethics/. Accessed Mar 2025

Beauchamp TL, Childress JF (2009) Principles of biomedical ethics, 6th edn. Oxford University Press, New York

Childress JF (2007) Methods in bioethics. In: Steinbock B (ed) The Oxford handbook of bioethics. Oxford University Press, Oxford, pp 15–45

Corrêa NK et al (2023) Worldwide AI ethics: a review of 200 guidelines and recommendations for AI governance. Patterns 4(10):100857–100857

EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act. (n.d.). https://artificialintelligenceact.eu/

Fleisje A (2023) Paternalistic persuasion: are doctors paternalistic when persuading patients, and how does persuasion differ from convincing and recommending? Med Health Care Philos 26(2):257–269. https://doi.org/10.1007/s11019-023-10142-2

Maese Ellyn. (2025) "Americans Use AI in Everyday Products without Realizing It." Gallup.Com, Gallup, 15 Jan. news.gallup.com/poll/654905/americans-everyday-products-without-realizing.aspx.

Partnership on AI - Home - Partnership on AI. (2025). Partnership on AI. https://partnershiponai.org/. Accessed Mar 2025

Pew Research Center. (2023) Public Awareness of Artificial Intelligence in Everyday Activities: Limited enthusiasm in U.S. over AI's growing influence in daily life. [PDF]. https://www.pewresearch.org/wp-content/uploads/sites/20/2023/02/PS_2023.02.15_AI-awareness_REPORT.pdf. Accessed Mar 2025

Pilz K. F., Mahmood Y., & Heim L. (2025) AI's power requirements under exponential growth: extrapolating AI data center power demand and assessing its potential impact on U.S. competitiveness. RAND. https://www.rand.org/pubs/research_reports/RRA3572-1.html. Accessed Mar 2025

Ross WD (1930) The right and the good. Clarendon, Oxford

Special competitive studies project. (2024) Fireside - Eric Schmidt [Video]. YouTube. https://www.youtube.com/watch?v=oC46CzxT750

Tobin MJ (2022) Fiftieth anniversary of uncovering the tuskegee syphilis study: the story and timeless lessons. Am J Respir Crit Care Med 205(10):1145–1158. https://doi.org/10.1164/rccm.202201-0136SO