Dean&Francis
ISSN 2959-6149

Negative Obligations in Policymaking: Contractualist Approaches to Generative AI

Zi Ye

Abstract:

The future of generative AI (GenAI) presents various ethical challenges for lawmakers, as effectively legislating a rapidly evolving technology can be difficult when so much of its implications are still unknown. By introducing a conceptual analysis of what positive and negative obligations are appropriate for present and future people, the article will argue that we owe NO to both present and future peoples; that is, we bear the duty of ensuring human survival and taking precautionary action on behalf of all generations. However, while we owe PO to living people, we do not owe those same obligations to future generations due to the future's temporal remoteness and moral vagueness. Policymakers, therefore, should also operate based on these principles of positive and negative obligation, an approach which is justified by contractarianism. Towards the end, it will engage with a case study on the EU AI Act and show how it links with a suggested policy making process.

Keywords: negative obligation, temporal remoteness, contractualism, AI regulation

The charm of time traveling in any sci-fi story has always been in its ability to show us the future, something we lack and desperately yearn for as mortal beings with limited foresight. Any difficult choice in life can be boiled down to the unpredictability of what will happen next, specifically how our actions might affect those in the future. But in the real world, time travel does not exist (yet) and the moral asymmetry between present and future becomes more nuanced than any ficticious time travel adventure. Thus, it is important to consider what we owe to present people that we do not owe to those in the future.

This essay will begin with two essential concepts—negative obligations (NO) and positive obligations (PO)—and which are appropriate for present and future people. For brevity, NO refers to the responsibility to refrain from harmful actions and take precautionary methods that would hinder human dignity, while PO refers to the responsibility to "act" to resolve issues and improve conditions that strengthen human dignity (UNODC, 2019). The second half will engage with a case study on the EU AI Act and show how it links with my suggested policy making process.

Thus, this essay argues that we owe NO to both present and future peoples; that is, we bear the duty of ensuring human survival and taking precautionary action on behalf of all generations. However, while

ISSN 2959-6149

we owe PO to living people, we do not owe those same obligations to future generations due to the future's temporal remoteness and moral vagueness. Policymakers, therefore, should also operate based on these principles of positive and negative obligation, an approach which is justified by contractarianism.

Negative vs. Positive Obligations

Moral obligation is the requirement to pursue what we believe is right and act accordingly. All possible moral obligations that we could owe to present and future generations can be broadly divided into two categories of PO and NO. First, I will establish that the mutual obligations that we owe to both present and future are only NO.

By refraining from harmful actions and making precautionary policies, the fulfillment of NO ensures the baseline survival of humans, no matter present or future. Our negative obligations can include, but are not limited to, refraining from deforestation, excessive resource depletion, mass murder, and increasing our carbon footprint. Thus, our NO becomes an essential and irrevocable duty, insofar they hold a "baseline" for human dignity and preserve our energy by preventing future problems from forming in the first place.

For example, refraining from deforestation is a NO preserves biodiversity and ecological balance, decreasing the possibility of flood or mass extinction of animals. By preventing future disasters, our collective energy can further be pivoted towards ongoing prevention efforts, rather than intervention, which is a more sustainable long-term strategy. Thus, the fulfillment of our NO becomes a necessary duty in every timeframe.

Temporal Remoteness and Moral Vagueness

While we have negative NO to ensure baseline for human dignity, we also owe PO to present generations, but not to future ones, due to the latter's lack of temporal immediacy, otherwise known as temporal remoteness, where the extended duration between actions and their consequences undermines the grounds for positive obligations of contribution. An action that results in immediate negative consequences will cause a greater increase in moral urgency than an action whose consequences are delayed (Lincoln & Holmes, 2011).

The present generation benefits from temporal immediacy as they most urgently receive consequences from our actions, while future generations do not experience the same urgency. This is because our actions have an immediate effect on the present, as being "present" means incidents and effects are happening at this moment. This is a characteristic that future events or consequences do not have: the length of time between the action done in present and its consequence is infinitely extended, resulting in temporal remoteness.

Imagine a moral agent has to make a decision to prevent a potential murder occurring today versus two years in the future. The today's murder is more temporally immediate since it is closer to the current moment than the one in two years, and naturally, the moral agent will choose to save the first victim as that decision is more urgent. For the same moral decision that only differs in temporal immediacy, an increase in moral urgency, intuitively, increases the moral burden on the actor who is obligated to act on this issue. Conversely, the less moral urgency, the less moral burden, namely burdens for PO, falls on the moral agent. Thus, temporal remoteness results in little to no PO on them.

Another factor that extinguishes living people's PO to future generations is the moral vagueness of the future. There are three categories of consequence recipients in the future: (1) non-existent people, referring to future people whose existence is not determined, but can be affected by our present actions; (2) distant present people, referring to current people pictured in a distant future, such as children or babies that already exist; and (3) hypothetical people, referring to possible not-existent people that can help to hypothesize scenarios, but they are not causally related to our actions (Bruneau, 2018).

The PO we owe to each category of future persons is different. Non-existent people deserve no PO, because we are unsure if they will even exist, diminishing the moral urgency we have to solve their problem. Distant present people deserve weak PO, as they do exist in future, just with uncertain prospects, which undermines our level of PO towards them. Hypothetical people also deserve no PO, as they do not tangibly relate to our actions, extinguishing our moral duty to them.

Here, non-existent people and hypothetical people cannot consent to or participate in decision making, nor can they suffer harm directly, thus our PO towards them is morally vague. This is because the difficulty in distinguishing recipient groups and lack of specificity in our weak PO to distant people. Hence, the future is morally vague.

Contractualist Policy Making

Policy making must prioritize the rights, safety, and autonomy of the people who the policy is concerned with. It ought to prioritize immediate or short term benefits (ie. no longer than a few decades), while not harming humanity's future prospects. In other words, we must follow our PO

(ie. prioritizing benefits, rights and safety) to current people and NO (ie. refraining from actions that harm humanity) to both present and future people.

This policymaking approach takes inspiration from contractualist political philosophy. According to social contract theory, the State should act based on the needs of a rational agent; due to the moral ambiguity and lack of consent from non-existent and hypothetical people, such rationality does not exist—intuitively, it is hard to seek rationality from non-existent beings (Friend, n.d.). Therefore, non-existent and hypothetical people are not within the range of consideration of the state, which removes the state's burden of fulfilling PO to them.

For distant present people who certainly have some degree of rationality, the State only owes them a very weak PO. Thus, the policy making process should center rational agents in the present and only make precautions to refrain present people from harming the future. Anything beyond precautionary measures would impose on the autonomy of future rational agents, thus violating the State's social contract with them.

Negative Obligations in Action: The EU AI Act

Generative AI (GenAI) presents a uniquely powerful challenge for moral ethicists and policymakers alike—its consequences will stretch far into the future, affecting unborn generations. But it also creates immediate, observable harm. These range from biased hiring algorithms to black-box policing tools, and social media systems that impact mental health, particularly among youth. For example, GenAI systems has already negatively influenced teenagers' identity formation, spread misinformation, and amplified false stereotypes (Blouin, 2023).

Furthermore, copyright and intellectual property laws are ill-equpped to handle issues of liability and accountability for harmful or misleading GenAI content (Rajappa, 2024). Typically, such laws would only find developers or users as culpbable, but some philosophers have started questioning whether GenAI, which is becoming more autonomous by the day, will soon need legal personhood (Rajappa, 2024). However, these are big questions that might not be answered relatively soon and any delays in legal reforms will only create a bigger chasm between regulations and GenAI.

Thus, ethical frameworks for GenAI must begin with the structures and harms visible in the present. The EU AI Act is an exemplar of how to appropriately legislate within this framework.

According to the act, systems should be "designed and developed in such a way that natural persons can oversee their functioning […] and that their impacts are addressed over the system's life cycle" (European Union, 2024). This is a policy rooted in NO—ensuring that today's technologies do not lead to irreversible harms tomorrow. It wisely avoids placing unnecessary positive burdens on the present by attempting to anticipate every future harm GenAI will cause. Instead, it responds to concrete existing harms by mandating transparency, accountability, and human oversight in high-risk GenAI applications, maintaining a moral bottom floor: do not build systems that will destroy autonomy, safety, or fairness—even if we cannot fully predict their long-term effects.

Objections and Rebuttals

As a defender of moral equity might say, future people have the same moral worth as present people, and if we employ a utilitarian framework, temporal distance would be irrelevant. In a sense, harming someone in 2200 AD would be just as wrong as harming someone today. Under this perspective, discounting future interests due to "vagueness" or "lack of immediacy" would count as a type of arbitrary discrimination—akin to valuing nearby strangers less than distant family (Parfit, 1984). This viewpoint would want for all humans to be valued equally.

However, while future people certainly matter, our duties to them differ due to epistemic limits and causal asymmetry. Due to moral vagueness, there is no way for us to perform PO on them, imposing our preference on theirs and claim we are contributing. Imposing a utilitarian framework onto the future is definitely not respecting their human rights when we are ignorant if they are willing to choose such moral framework; instead, freedom should be granted. We, in the current generation, are completely clueless about the specific preferences and values of future ones. Diminishing PO and only performing NO to the future grants them more autonomy. NO ensures their freedom of choice: by taking precautionary policies and refraining from destructive actions, we make humanity capable of choosing the way they prefer society to be.

This is further complicated by the non-identity problem: since future people's very existence depends on our present choices, we cannot coherently claim to act "for their benefit" without circularity. What harms or helps them is contingent on decisions that also determines who comes to exist. Thus, paternalistic interventions (like imposing utilitarian optimization) are not only epistemically ignorant but also conceptually unstable, as there is no fixed future population to whom we owe specific positive obligations.

ISSN 2959-6149

Conclusion

Ultimately, understanding the difference between what we should do for others (PO) and what we should not do (NO) helps clarify our duty to the future. We owe people alive today both kinds of duties—actively helping and avoiding harm—based on our direct connections and shared society. But for future generations, our core duty is simpler: don't cause irreparable harm. Avoid poisoning their earth, do not leave them unsolvable disasters, and certainly do not destroy what they cannot replace.

This does not mean the future does not matter. It means our responsibility is about responsible restraint, not fruitless guesswork. Morality here is not about designing their world; it is about not breaking it. Our job is to leave them a world where they can choose their own path—not one where they are stuck fixing the messes we have left them.

The real ethical test is not dreaming up perfect futures; it is making sure we do not make things worse. Oftentimes, the most responsible choice is knowing when to leave things alone.

References

Blouin, L. (2023). *AI's mysterious "black box" problem, explained*. Dearborn. https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained

Bruneau, G. (2018). Do we have moral obligations towards future people? Addressing the moral vagueness of future environmental scenarios. *VERITAS*, *40*, 49-65. https://redae.uc.cl/index.php/veritas/article/download/62599/50151

European Union. (2024). *The AI act explorer*. EU Artificial Intelligence Act. https://artificialintelligenceact.eu/ai-act-explorer/

Friend, C. (n.d.). *Social contract theory*. Internet Encyclopedia of Philosophy. https://iep.utm.edu/soc-cont/. Accessed July 2, 2025.

Lincoln, S., & Holmes, E. (2011). Ethical decision making and the influence of moral intensity. *Journal of Healthcare, Science and the Humanities, 1*(1), 55-69. https://www.usna.edu/Ethics/_files/documents/ethical%20decision%20making%20and%20the%20influence%20of%20moral%20intensity.pdf

Parfit, D. (1987). *Reasons and persons*. Clarendon Press.

Rajappa, S. (2024). Existing legal frameworks and gaps in generative AI. *Forbes*. https://www.forbes.com/councils/forbestechcouncil/2025/09/09/enterprise-ai-is-moving-fast-can-it-operations-keep-up/

United Nations Office on Drugs and Crime [UNODC]. (2019). Positive and negative obligations of the state. In *Trafficking in persons & smuggling of migrants.* https://www.unodc.org/e4j/zh/tip-and-som/module-2/key-issues/positive-and-negative-obligations-of-the-state.html