27
GUIDED CHOICES
The ethics of using algorithmic systems to shape philanthropic decision-making

Rhodri Davies

1 Introduction

Discussion of the potential impact of artificial intelligence (AI) in the field of philanthropy has, to date, largely focused on three areas:

- The potential for harnessing machine learning (ML) and other tools to deliver new social and environmental interventions and how civil society organizations (CSOs) and philanthropic funders could play a role in realizing this potential;
- How AI might affect CSOs and philanthropic funders themselves through new opportunities to transform internal processes or through altering the nature of the broader financial and regulatory systems within which they operate;
- The impact that AI may have on individuals and communities and what this might mean in terms of the need for new approaches or increased advocacy from funders and CSOs.

What has received far less attention is the question of how AI may affect the philanthropic choices made by individuals, about when, where, and how to give. This may reflect a broader lacuna in our knowledge since, as Susser notes, "for several years, scholars have (for good reason) been largely preoccupied with worries about the use of artificial intelligence and machine learning (AI/ML) tools to make decisions *about us*," but that "only recently has significant attention turned to a potentially more alarming problem: the use of AI/ML to influence our *decision making*" (Susser, 2019). However, even if this is true across a broader set of domains, it is a particular problem for philanthropy, as individual choice plays a fundamental role in this context. Philanthropy's inherently dual nature (Reich et al., 2016) means that we have to understand it at both a macro level – as a systemic mechanism for allocating and redistributing resources within society at a scale that positions it alongside both the state and the market – and a micro level, where it reflects the myriad choices of individuals to use their "voluntary action for the public good" (Payton & Moody 2008). Hence, the capacity of AI to affect how we make individual choices has profound implications for the future of philanthropy.

This chapter will look at these implications and how we might respond to them. It will consider the nature and role of choice in philanthropy, what we know about the capacity of AI to influence

DOI: 10.4324/9781003468615-32

and shape our individual choices, and what this suggests about the potential impact of AI on philanthropic choice and decision-making. The implications will be considered across three key groups of actors that have the potential to employ AI to shape the decisions of potential donors, namely:

- General-purpose search and recommendation services (i.e., search engines, conversational interfaces, text-based generative AI tools);
- Giving platforms (i.e., dedicated donation/lending or crowdfunding platforms and commercial platforms that facilitate giving, such as payment providers or social media platforms);
- Individual cause-based organizations (both formal and informal).

The chapter will highlight key ethical questions that emerge from considering the use of AI by each of these different actors and suggest changes in policy or practice that could help to address them.

2 Choice and philanthropy

Choice plays a fundamental role in philanthropy since our ability as individuals to decide whether and how to give away private assets for public benefit is one of the defining characteristics of philanthropy when set against other redistribution methods. However, there may be a danger that in making this assumption, we beg an important question about the nature of philanthropy since one of the fundamental philosophical questions about philanthropy throughout the ages has been whether it should be understood as a choice of duty. Is giving something you are entirely free to choose whether to do or not to do or are there moral or societal obligations of some kind that compel you to give (Martin, 1994; Schneewind, 1996)? To take one example of how this debate has evolved and the influence it has exerted, the predominant view among medieval Catholic scholars was that God determined the unequal distribution of wealth and resources within society and that part of his plan was for redistribution to take place through charity and almsgiving; so the "haves" had a duty to give to support the "have-nots" (Roberts, 1996). The Enlightenment then saw a shift in thinking about the nature of property that caused views to diverge. For some, like Locke or Grotius, people had a "natural right" to ownership of property that they had amassed through their efforts, and it was up to them whether to choose to share it (Winfrey, 1981). For others, like Wollstonecraft and Kant, the unequal distribution of resources reflected the fact that society was unjust, and addressing this injustice demanded that those with wealth give back as a matter of duty rather than of choice. To the extent that this duty was a "perfect" one (i.e., a duty where it is clearly defined who the recipients are and what they are due), it was usually thought best discharged through taxation. However, many argued that philanthropy represents an additional "imperfect" duty, where the requirement to give is clear, but the exact nature of how the money must be given or to whom is not specified (Schneewind, 1996).

The extent to which philanthropic giving represents a choice or a duty remains a matter of debate even today. Those who prioritize individual liberty may feel that philanthropy must be seen entirely as a matter of choice (Nozick, 1974; Salmon, 2023). Others who place more emphasis on justice and the equitable distribution of resources will argue that we must acknowledge an element of duty when it comes to philanthropy, even if this impinges upon our individual freedom. For instance, the moral philosopher Peter Singer has argued that there is a moral duty on those with sufficient wealth to give some of it out to help people in extreme poverty (Singer, 1972, 2006). Cordelli (2016) goes further, arguing that "philanthropy should be understood foremost as a duty of reparative justice" and that "affluent donors should, as a matter of moral duty, exercise no personal discretion when deciding how to give and to whom. Indeed, they should regard their donations as a way of returning to others what is rightfully theirs." In this view, philanthropy becomes solely a matter of duty, and there is no room for choice.

A reasonable position between these two extremes is to allow that both duties and choices guide philanthropic giving. We pay taxes as part of an agreed social contract, but this does not necessarily fully discharge the duty we owe to others, so an element of duty may also apply to our giving. In part, this may be simply an imperfect *duty to give* in an unspecified way, but it may also be a more specific duty to give *in a particular way* or *to particular causes* (MacAskill, 2015; Singer, 2015). (A duty that would obviously only have moral rather than legal force, of course.) Most people would stop short of claiming that *all* our philanthropy is driven by duty, however. They would allow the possibility of some giving that is *supererogatory* (i.e., above and beyond that which is demanded by duty). This portion of giving, at least, would be governed solely by choice in terms of our decision to give in the first place and in terms of what we choose to give to (Gewirth, 1987).

According to most current views, choice remains a vital ingredient of philanthropy. For this reason, it is unsurprising that a substantial body of research aims to shed light on the key factors affecting the choices we make when giving. Several literature reviews have attempted to survey this body of research and synthesize findings across it (Bekkers & Wiepking, 2011; Allen, 2018; Saeri et al., 2023), although they note that this presents challenges because the research is found across a wide range of disparate disciplines – including economics, political science, anthropology, neurology, psychology, and marketing – reflecting the inherently cross-disciplinary nature of philanthropy as a field of study. Despite this variety, however, it is possible to identify common themes across research from different fields. Bekkers and Wiepking (2011), for instance, postulate eight mechanisms that drive giving based on their analysis of the available literature: (1) awareness of need; (2) solicitation; (3) costs and benefits; (4) altruism; (5) reputation; (6) psychological benefits; (7) values; (8) efficacy.

Within these broad categories, it is possible to identify a wide range of specific factors that may affect philanthropic decision-making. Some relate to macro-level considerations, such as religious and ethnic diversity among donors (Andreoni et al., 2016), social class and socioeconomic status (Piff et al., 2010), the impact of government funding (Andreoni & Payne, 2011), the presence of match funding (Karlan & List, 2007), the availability of "social information" on what other donors have given (Alpizar et al., 2008; Croson & Shang, 2008; Shang & Croson, 2009), or the effect of media coverage of disasters (Brown & Minty, 2008). Various studies have also attempted to identify relevant cultural factors by exploring children's attitudes to giving and prosocial behavior, finding that children appear to have a natural inclination to act pro-socially even from a very young age (Zahn-Waxler et al., 1992; Warneken & Tomasello, 2006), that younger children are less prone to "free-riding" behavior than adults (Harbaugh & Krause, 2000), and that as children get older, they are more likely to give to someone when they believe that the other person might reciprocate their generosity (Sebastián-Enesco & Warneken, 2015).

Many micro-level factors affect decisions made about giving at an individual level. Some of these are relatively unsurprising, such as whether the person soliciting donations is a member of their peer group (Meer, 2011), whether the person soliciting makes an audible request or not (Andreoni et al., 2017), the level of information the donor has about the cause being fundraised for (Eckel et al., 2007), whether that information concerns an identifiable individual or is in the form of statistical evidence (Jenni & Loewenstein, 1997; Slovic, 2007), and what kind of information the donor is given about the performance of the organization asking for funds (Butera & Horn, 2020). Other factors are perhaps less obvious, such as the influence of our body chemistry:

a number of studies have found, for instance, that the release of oxytocin results in a greater willingness to give (Zak et al., 2007; Barraza et al., 2011), while others have found that the brain's dopamine reward center may respond similarly when we give to how it does when we receive other kinds of rewards (Moll et al., 2006). There are also more esoteric factors that studies have found may have a bearing on our willingness to give and the choices we make: for instance, being exposed to "awe-inspiring" content (Rudd et al., 2012), having a prior conversation that primes you with the notion of a God or Supreme Being (Shariff & Norenzayan, 2007), thinking about death (Jonas et al., 2002), listening to "chill-inducing music" (Fukui & Toyoshima, 2014), listening to music with "prosocial lyrics" (Greitemeyer, 2009), or being in the presence of images of eyes (Ekström, 2012; Sparks & Barclay, 2013; Fathi et al., 2014).

As more and more of us give online − either through dedicated nonprofit platforms or, increasingly, through commercial social media or payment platforms that have added giving functionality to their offering – the nature of philanthropic choice and decision-making is changing rapidly. For one thing, the range of options available to us is far more significant because we are no longer constrained by physical proximity: many platforms enable individuals to give to groups and causes worldwide without ever needing to contact potential recipients or fundraisers. We are also less constrained by a reliance on formal organizations, as a large volume of online giving goes to informal groups, grassroots social movements, or individuals (Bernholz, 2021). But perhaps the most profound impact of the shift toward giving online is that the digital environment offers new opportunities to design "choice architectures," that is, how choices are presented to us so they are tailored and responsive to our personal preferences and can be highly targeted to produce desired outcomes. Knowledge of the individual − and often unconscious − factors that shape our giving decisions will be an important competitive advantage when harnessing such approaches' potential. Some have already questioned whether this raises ethical concerns and whether it is appropriate to exploit insights about our subconscious behavioral drives to "nudge" us toward specific actions in this way when it comes to philanthropy and other prosocial behavior (Schulz et al., 2017; Ruehle et al., 2021). Such concerns are only likely to become more acute as the impact of AI is felt more widely, as we shall see.

3 Choice and AI

AI already significantly impacts how choices and decisions are made in a wide range of fields. As noted at the start of this chapter, there has been a particular emphasis on how algorithmic systems are used to make decisions *about* us, with a growing body of literature examining the opportunities this presents – in terms of bringing new capabilities and efficiencies across the public, private, and nonprofit sectors (Wirtz et al., 2019; Kanter & Fine, 2022) – as well as the challenges it brings, such as new risks of "machine bias" against marginalized groups (Wachter-Boettcher, 2017; Eubanks, 2018; Noble, 2018) and concerns about insufficient transparency and accountability (de Fine Licht & de Fine Licht, 2020; Loi & Spielkamp, 2021). In this chapter, however, our primary focus is not on how AI may be used to make decisions about us but how it might affect our decisions as individuals when it comes to philanthropic giving. To that end, we will consider three key areas:

- The use of algorithms to determine responses to requests for information and how this shapes our choices and decision-making;
- The capacity of AI to enable personalization and "hyper-nudging," to drive particular actions and outcomes;
- The use of AI-generated content to prompt emotional responses and thus drive behavior.

3.1 AI and information provision

Our online world experience is heavily shaped by algorithms (Schmidt, 2021). While it is theoretically possible to navigate the internet without using intermediaries, in reality, the vast majority of web traffic comes via search engines (BrightEdge, 2019). More recently, social media platforms and conversational interfaces (either text-based, such as OpenAI's ChatGPT, or voice-based, such as Apple's Siri or Amazon's Alexa) have also evolved to fill a similar role in many users' online experiences (Huang, 2022; Perez, 2022). In both cases, the information we are provided is determined algorithmically in one of two main ways. The first is the *reactive* provision of information in response to a request, where a range of suitable answers may be algorithmically determined and then presented to the user either in the form of a ranked list (as in the case of a traditional search algorithm, such as that used by Google) or in the form of a single answer or small handful of answers (as tends to be the case with conversational interfaces). The second way algorithms may be used is for the *proactive* provision of information, where the user has not made an explicit request, but instead, suggestions or recommendations for content they may be interested in are provided to them based on data about their known interests, preferences, or online behavior. Recommender algorithms, which typify this latter approach, are at the heart of social media platforms such as TikTok, X, and Facebook, as well as content platforms like Spotify, Netflix, and YouTube, all of which seek to provide a steady stream of enticing content for us to click through to next to keep us on the platform for as long as possible (Schrage, 2020; Roy & Dutta, 2022).

When trying to understand how algorithms can shape our choices across both of these contexts, there are at least three things we need to consider: *what* information is provided (i.e., what the algorithm determines as the "correct" or "best" answer to a query, or how it decides what to recommend to us); *how* the information is provided (i.e., whether it comes in the form of a single answer or a list, and whether we can refine the information presented to us); and finally *our perception* of the information (i.e., whether we are aware that it represents the result of an algorithmic process).

The most basic question we must ask is: what is the nature of the algorithm that has produced the information we are presented with (with the immediate corollary questions: "Who designed it?" and "What are their motivations?")? Unfortunately, here we encounter an issue that will prove to cut across almost all our considerations, namely lack of transparency. In the case of many of the algorithms that shape our online experience – such as Google's ranking algorithm, TikTok's video recommender algorithm, or YouTube's content algorithm – the question of how they work largely remains a mystery. This is partly because these algorithms are closely guarded proprietary secrets. Still, it is also because many of them are by design "black boxes," whose inner workings are deliberately opaque and not always fully comprehensible even to those who created them (Pasquale, 2015; von Eschenbach, 2021). Scholars and campaigners have increasingly called for this situation to be remedied by introducing greater openness to the use of algorithms. Initially, this was focused on the idea that we need to make algorithmic systems more transparent, but more recently, the focus has shifted toward making algorithms "explainable," since transparency by itself is no longer felt by many experts to be a useful goal, as for most users being able to see the inner workings of an algorithm doesn't further their understanding of how it works (Mittelstadt et al., 2019; Rai, 2020).

How information is presented to us is also important. In particular, the emergence of conversational user interfaces (CUIs) as alternatives to traditional search engines to find information is potentially very significant (Liao et al., 2020). The design of CUIs may vary considerably, from basic chatbots that provide only limited responses through systems that can hold natural language conversations at a level that may be functionally indistinguishable from a human being; however, one feature that most of them have in common is that information is no longer presented in the form of a complete ranked list that can be inspected visually, but in the form of an "answer" (or small set of answers) to a question we have posed (Zamani et al., 2023). In practical terms, this may make it harder to identify and compensate for any biases. There are already concerns that traditional search engines perpetuate biases of various kinds (Espín-Noboa et al., 2022; Maillé et al., 2022), but the challenge may be more significant when it comes to information delivered through CUIs; both because of how we interpret the status of the information and because it is no longer possible to compensate for bias simply by looking further down the list of pages of results, as we might do with a traditional search engine. For these reasons, among others, some have questioned whether CUIs should be considered viable alternatives to conventional search engines (Gurdeniz & Hosanagar, 2023).

There are also deeper issues here. For one thing, we tend to see language as a marker of intelligence (Mahowald et al., 2023), so the more that automated systems can respond to our requests for information in credible natural language, the more likely we are to accept what they tell us and the less likely we are to assess that information critically. Another challenge is that the conversational nature of interfaces may reduce the *visibility* of the underlying technology, that is, our awareness that a process of digital intermediation and algorithmic determination of information is taking place. As scholars have noted, the potential *invisibility* of technology is a significant factor when it comes to practical and ethical concerns about how AI may shape individual choices, both because we are rendered far more susceptible to manipulation through technology when we are not even aware that this technology is being used (Van den Eede, 2011; Susser, 2017), and because the threat to autonomy is significantly higher (André et al., 2018; Vaassen, 2022; Bartmann, 2023).

The increasing invisibility of algorithmic systems and our willingness to accept the information they present us with less critically has led some scholars to claim that the distinction highlighted earlier in this chapter – between reactive information provision and proactive recommendation – is no longer that meaningful. For instance, Zamani et al. "do not make a strong distinction between *search* and *recommendation* tasks" because they see them as "closely related tasks that are becoming more closely related as time passes" and argue that in many cases, the same task can often be characterized as either search or recommendation depending on the device or interface being used and the context in which the task is taking place (Zamani et al., 2023). This blurring of boundaries between searches for information that we undertake actively and recommendations that we receive passively clearly has implications for the degree of autonomy we retain as users (Bartmann, 2023). However, we also need to bear in mind that the increasing reliance on algorithmically determined recommendations is primarily driven by our demands as consumers for highly tailored and personalized experiences (Arora, 2021). Should we, therefore, view this less as an unprovoked assault on our autonomy and more as a Faustian pact that we have entered into, in which we sacrifice autonomy for convenience?

3.2 Personalization, choice architecture, and hyper-nudging

The information provided to us is an important determinant of our choices, but it is by no means the only one. A range of factors relating to how options are presented to us also significantly affect our decisions.

Thaler and Sunstein (2009) coined the term "choice architecture" to describe this range of factors (and "choice architect" to refer to anyone who has control over them) (2009). Advertisers and marketers have long known that our subconscious motivations and drivers can be manipulated to sell us things (Samuel, 2010). Still, since the publication of Thaler and Sunstein's work, there has also been significant interest in how the design of choice architectures can be used in the public and nonprofit sectors to "nudge" people toward desirable prosocial actions (Behavioural Insights Team, 2013; Schulz et al., 2017; Capraro et al., 2019). (A "nudge" is defined by Thaler and Sunstein as "any aspect of choice architecture that predictably alters people's behavior without forbidding any options or significantly changing their economic incentives.")

Many early experiments with choice architecture focused on using the design of physical spaces or interactions to influence people's actions. However, it is in the digital world that choice architecture and nudging have really come into their own, as in this context, it becomes possible to tailor choice architectures to individuals in a way that is not feasible in the physical world (Thomas et al., 2013; Weinmann et al., 2016). AI has now added a significant further dimension, as the combination of big data and algorithmic processes makes it possible not only to personalize nudges but to adapt them iteratively in real time in response to user behavior. The legal scholar Karen Yeung has coined the term "hyper-nudge" for this phenomenon (Yeung, 2017) and argues that:

> unlike the static Nudges popularized by Thaler and Sunstein (2009), such as placing the salad in front of the lasagne to encourage healthy eating, Big Data analytic nudges are extremely powerful and potent due to their networked, continuously updated, dynamic and pervasive nature (hence "hypernudge").

There is an existing body of critical thought about the ethics of nudge approaches (Bovens, 2009; Hobbs, 2017; Ruehle et al., 2021), but hyper-nudges bring new challenges of their own. Allowing autonomous systems to act as choice architects may, for instance, raise significant issues regarding transparency, responsibility, and accountability (Mills & Sætra, 2022). There are also questions about whether the iterative and adaptive nature of hyper-nudges makes it harder for us to avoid them than traditional nudges, and whether they present a more significant threat to our autonomy (André et al., 2018; Mills, 2022).

3.3 Generative AI and emotive content

Generative AI is a term for a broad class of models or algorithms that produce new content – such as text, photos, illustrations, video, audio, or code – based on large sets of suitable training data. We are already seeing generative AI being used in efforts to shape our individual choices. This will likely become even more commonplace in the future since the ability to generate audio and visual content that seems genuine – while in fact, being carefully calibrated to convey specific messages or emotions – represents a hugely powerful new tool for inducing beliefs and influencing actions (Sætra, 2023). We will undoubtedly see an increase in malicious attempts to spread misinformation and disinformation to induce false beliefs; there have been widespread concerns, for instance, about the rise of "deepfakes" – synthetic media that have been digitally manipulated to replace one person's likeness convincingly with another – and their potential deployment in the context of elections and other political events (Chesney & Citron, 2018; Diakopoulos & Johnson, 2021; Pashentsev, 2023). Not all applications of generative AI will be deliberately malicious, of course, but some scholars argue that even the well-intentioned use of technology such as deepfakes is potentially harmful because it undermines our notions of authenticity in the online environment and thereby exacerbates the problematic erosion of trust (Hancock & Bailenson, 2021; de Ruiter, 2021).

4 AI and philanthropic choices

To understand how general considerations about how AI could affect individual choices apply in the specific context of philanthropy, we will consider three key actors that play a role in philanthropic choice-making in the digital environment:

- General-purpose search and recommendation services;
- Giving platforms;
- Individual cause-based organizations.

We will consider each actor's motivations for wanting to shape our philanthropic choices, what this means in terms of their likely adoption of AI tools, and the potential ethical raised.

4.1 General-purpose search and recommendation services

Many websites, platforms, and interfaces providing information to users – either in the form of reactive responses to requests or proactive recommendations – are not explicitly designed with philanthropy or the nonprofit sector in mind. The paradigm example is traditional search engines, but this category also includes any social media, content platforms, or conversational user interfaces that allow users to search for general information or to receive recommendations that could influence choices about giving. The primary motivation, in this case, is satisfying user demand for information – whether that is explicit (e.g., a search request) or implicit (e.g., a perceived openness to receiving a recommendation) – to ensure that users remain on a platform or return to a service in future. It is also possible that the company operating the search or recommendation service has its own corporate social purpose, which may be an additional factor in the motivation to provide information that can shape philanthropic choices.

In the case of providing information in response to a user request, it is unlikely the provider will exert any influence over the initial decision to give since this has most likely already been taken. However, there is substantial potential for influencing *where* and *how* the gift is made. A user may be seeking objective information – for instance, on which nonprofit organizations operate in a particular geographical location – or that user may be looking for more subjective information – for instance, on the "most pressing need" or which nonprofit organizations are "most effective." In both cases, it is important to interrogate the nature of the algorithms used to determine the information on offer and the goals and motivations of those designing and operating them. As highlighted earlier in this chapter, this is particularly important in those cases where responses to requests for information are presented in the form of a single answer or small selection of answers rather than a list, as in these cases, our ability to counteract the choice architecture imposed by the algorithm is much more limited.

In the case of proactive recommendations or unprompted information designed to act as a nudge, the user need not have taken any prior decision to make a philanthropic gift, so it would be possible to influence not only where and how they give but their choice to do so in the first place. In this situation, it would be important to examine what motivation the search provider or platform had for encouraging or nudging prosocial behavior in the form of giving. For instance, if it reflects a broader corporate social purpose, is that purpose explicit and known to the user? Is the aim merely to encourage giving in a generic sense, or is the information presented in such a way as to promote giving to specific types of causes or according to a specific philanthropic ideology?

In previous cases where generic service providers have decided to offer some form of giving functionality, such as the ability to make donations when using ATMs or paying at a checkout, concerns have been expressed about the limitations of the options available to users and the role the service provider plays in determining them (Caulfield et al., 2022). In the case of ATM giving or checkout donations, it is a relatively straightforward matter for users to bypass the nudge. In the case of nudges applied online, however, as we have already seen, the reduced visibility of the technology may lead to users being less aware that a choice architect is attempting to shape their behavior – which might make the nudge more effective but also brings a greater risk of eroding users' autonomy (Lv & Huang, 2022). In cases where hyper-nudges are implemented, the iterative and adaptive nature of the algorithms raises a further challenge in terms of an increased "burden of avoidance" (Susser, 2019), since it also becomes far more difficult for users to bypass the nudge.

4.2 Giving platforms

Many platforms are specifically dedicated to providing information and offering transactional services that encourage and facilitate giving. Some of these are themselves nonprofits or charitable organizations, while others may have for-profit organizational structures but with a stated social purpose. Some focus primarily on traditional philanthropic donations, but others offer opportunities that go beyond this, such as the ability to make loans or to contribute to crowdfunding campaigns in support of organizations or individuals (Lilly Family School of Philanthropy, 2023).

The majority of giving platforms aim (either stated or implicit) to try to increase levels of generosity by making it easier and more appealing for people to give. Some, however, may emphasize increasing the quality of giving rather than the quantity, and to that end, may focus on promoting a particular view of effectiveness or impact. Even where platforms profess to be "cause neutral" (i.e., not promoting any one cause above others), it is essential to recognize that the range of options they present to users does not represent the entire domain of potential gift recipients since all platforms have criteria that limit what users can give to. In some cases, these criteria are explicit – such as limitations on the geographic location of recipients or requirements that they must have specific legal or organizational structures – but increasingly, there are also concerns that platforms take additional decisions to remove particular organizations or cause areas that may meet legal requirements, but which have been deemed unsuitable for some reason. As more and more giving takes place online, the concern is that platforms play the *de facto* role of arbiters of the acceptable boundaries of civil society simply because of the criteria they impose about what their users can or cannot give to, but that this power comes with little in the way of transparency or accountability (Carlman, 2020; Wade, 2022).

In many cases, users will engage with a giving platform having already made the initial decision to give, so any use of AI will be more about informing and shaping their choices about which causes to give to and what methods to use. The key questions then are about how algorithms are being applied and what their purpose or goal is. One vital dimension here is visibility: are users aware that an algorithm is being used to determine the information they are presented with? In some cases, the involvement of AI may be obvious: for instance, where a chatbot is clearly and openly being used to respond to search requests or to provide advice and recommendations. In other cases, however, the algorithmic processes may be hidden within the platform's infrastructure, so users may not be aware of their existence. This will have a significant bearing on the degree to which we have concerns about whether the individual autonomy of users is being undermined. It may also have practical implications, as there is some evidence that people are less likely to donate when they feel as though their ability to choose has been limited, even when they are presented with options that appeal to them (Lv & Huang, 2022).

The visibility of any algorithms used to determine the information provided by a philanthropic giving platform is an important consideration, but even more important is the question of *what those algorithms are designed to do*. In the case of giving platforms that position themselves as neutral intermediaries, presumably, the goal is to maximize user satisfaction − in much the same way as the general-purpose search interfaces considered above. What does it mean in practice to "maximize user satisfaction" when it comes to philanthropic giving? We might assume that it means presenting users with information on organizations and causes likely to appeal to them, thereby increasing the likelihood that they will make repeat donations to those same organizations and causes or to others in the future. However, there are reasons to be cautious about adopting this as an acceptable goal. For one thing, some would question whether maximizing donor satisfaction should be seen as the sole or primary aim of fundraising or philanthropy advice and would argue that the ability to challenge donors – by questioning their assumptions or pushing them outside of their comfort zone − is an important part of both professions (Beeston & Breeze, 2023). In recent years, there has also been growing debate in the fundraising field over whether "donor-centric" or "community-centric" models are preferable (MacQuillin, 2022). In the context of using AI to shape giving choices, the adoption of donor satisfaction as the primary objective may prove particularly problematic due to the recognized tendency of ML systems to exhibit "algorithmic bias" where algorithms trained on data sets that contain existing statistical biases come to reflect and entrench those biases over time. To design an algorithm that could provide giving choices and recommendations that were likely to satisfy donors, you would need to consider information such as individual user's past donations and the giving behavior of others who share relevant demographic features. However, that data reflects the existing limitations of the philanthropic marketplace: such as the fact that large organizations receive the lion's share of donations (National Council for Voluntary Organisations [NCVO], 2023), the fact that "unpopular" or "unfashionable" causes a struggle to raise funds (Body & Breeze, 2016), or the fact that the system as a whole demonstrates significant biases in terms of race and gender (Dorsey et al., 2020; Damm et al., 2023). The concern would be that an algorithm designed to satisfy donors and trained on this data would lead to many of these known challenges becoming exacerbated, with smaller organizations, unpopular causes, and organizations led by women or people of color becoming increasingly marginalized in favor of well-known nonprofit brands and "safe" causes.

Suppose concerns such as these lead us to conclude that donor satisfaction should not be the sole criterion when designing philanthropy algorithms. In that case the obvious question is: what other goals should we specify? We might suggest that the aim should be to maximize effectiveness, by ensuring that resources are targeted toward areas of greatest need, or ensuring they are directed toward interventions with the highest level of impact (according to some preferred measure). However, this raises both practical and theoretical challenges. In practical terms, we do not have the current data that would allow us to say where the need is most acute or where the impact is most significant, even if we wanted to. And in theoretical terms, it is not apparent that we could define either of these things meaningfully even if we did have all the required data. Identifying specific needs as "most acute" or specific interventions as having the "greatest impact" would require us to have measures of need and impact that can be applied objectively across cause areas, and many would argue that this is impractical and undesirable. Some approaches to giving, such as Effective Altruism (EA), try to address this challenge head-on by promoting the idea of a utilitarian, cause-agnostic single measure of value that can be used as a benchmark across all philanthropy (MacAskill, 2015). These are far from universally accepted, however, and EA in particular has attracted heavy criticism from many for being unrealistically normative, for applying an overly simplistic framework to complex social problems, and for being an ideology that favors easily measurable interventions within the current status quo over harder-to-measure campaigning and advocacy that aims at fundamental reform of systems and structures that are the root cause of societal challenges (Srinivasan, 2015; Gabriel, 2017; Crary, 2023). It would certainly be possible to use the principles of EA as the basis for a cause-agnostic algorithm that could provide information and recommendations to people to maximize overall Expected Value in a utilitarian sense but to do so would reflect a clear bias toward one particular, and strongly ideological, view of philanthropy.

Since it is not possible to find a purely objective way to measure impact or need (or at least one on which there is majority consensus), any additional criteria we apply in designing the goal of a philanthropy algorithm must be subjective to some extent, so their legitimacy will derive from whatever authority we appeal to justify their adoption. In some cases, a platform may rely on its own authority if it has decided to take its own stance on what constitutes "good" giving and apply these to the designs of any algorithms it is using, but this will obviously leave it open to challenges. In other cases, a platform may choose to appeal to an external authority, such as the United Nations Sustainable Development Goals (UNSDGs). Clearly, this still does not provide a purely objective goal for giving, and the UNSDGs have their fair share of critics (Swain, 2018). Still, as a globally agreed framework for prioritizing needs and focusing actions, they also have a relatively high degree of legitimacy (and certainly more than a single giving platform would have by itself), so they may provide an appealing basis for setting the goals of a philanthropy algorithm.

Once a goal or set of criteria has been determined that can act as the basis for designing a philanthropy algorithm, there may still be a further challenge to ensure that the algorithm remains aligned with the creator's original intention. At one time, the only option available for those attempting to create AI systems would have been to map out all of the possible steps in a process and then program those directly into an algorithm (this model is sometimes colloquially known as "Good Old Fashioned AI" or GOFAI [Boden, 2014]). This presented a significant limiting factor because it requires that we understand the underlying nature of all the capabilities we are trying to emulate and capture these in symbolic form, but there are many functions, such as natural language conversation or image recognition, that are seen as important aspects of intelligence and which humans can usually perform easily, but which we are not able to explain fully. This is one of the main reasons that the emergence of machine learning approaches as an alternative to GOFAI has led to such enormous growth and evolution in AI: because designing algorithms that can "learn" by going through a process of repeated iteration and self-modification to improve performance concerning a specified measure allows us to create systems that can approximate (or even surpass) human performance in certain tasks *without us having to specify all of the steps involved in performing that task* (Smith, 2019). In cases where machine learning algorithms match or exceed human performance, one of the things that has been noted is that they often do so by solving problems in ways that never would have occurred to a human and perhaps aren't even fully understandable to us. A growing body of literature has explored this phenomenon in various contexts; for instance, algorithms that are designed to play video games, where it has been found that they often end up achieving their set goal of scoring highly or winning by engaging in "specification gaming" (i.e., looking for loopholes or weaknesses in how the task has been specified or in the video game's design) rather than by playing within the confines of the game in the way that a human would (Krakovna et al., 2020; Lehman et al., 2020). This makes it clear that our *goals* and *values* in designing ML algorithms are hugely important. If we may have little or no control over *how* algorithms evolve to achieve a particular result, it becomes absolutely vital that we can stipulate clearly what our desired goal is and what the acceptable parameters are when it comes to achieving it.

The idea of Value Alignment has come to prominence through the work of philosopher Nick Bostrom, who argues that one of the key challenges of AI development is what he has christened the Value Alignment Problem (VAP), i.e., ensuring that highly autonomous AI systems are designed to ensure that their goals and behaviors remain aligned with human value throughout their operation (Bostrom, 2014). Bostrom's work focuses on Artificial General Intelligence (AGI) and Superintelligence, i.e., AI systems that are capable of matching or surpassing human-level intelligence concerning any task, which, at this point, remains hypothetical. To demonstrate the potential risks of value misalignment, he created a thought experiment known as The Paperclip Maximizer, in which a highly intelligent AGI is given the simple task of producing as many paperclips as possible (Bostrom, 2003). The danger, Bostrom argues, is that without additional specification of constraints, the AGI may choose to maximize its performance of this task in ways that radically diverge from our original intent: perhaps it will decide, for instance, to wipe out the human race because it believes that human beings pose a threat (because they may order it to cease paperclip production at some point in the future) or simply because we represent valuable stores of raw materials that the AGI believes would be better used for making more paperclips. The Value Alignment Problem doesn't just apply to Superintelligent systems or AGI, however. It is also relevant to domain-specific AI systems of this kind we have today. And it is not limited to trivial examples of algorithms bending the rules win at computer games either: as AI systems are increasingly deployed by governments and companies in a wide range of contexts, they have the potential to affect many areas of our lives, so ensuring that they remain aligned with our values and intentions is crucial (Korinek & Balwit, 2022).

In the context of philanthropy algorithms being deployed by giving platforms, as in many other areas, the challenge will be to find ways of giving AI systems sufficient freedom and autonomy to bring benefits in the form of increased effectiveness and efficiency while minimizing the risk of negative unintended consequences. There is a growing body of literature focused on "AI safety" that proposes ways of achieving this balance by designing safeguards that enable suitable human oversight and corrective acts when the risk of unintended consequences do occur (Leslie, 2019; Houben et al., 2022), and it is important that those seeking to create philanthropy algorithms draw on this literature.

4.3 Cause-based organizations

The final context in which AI may impact philanthropic giving choices is that of individual cause-based organizations (e.g., charities, social enterprises, etc.). In these cases, the use of AI is unlikely to involve the design of philanthropy algorithms in the sense discussed above since the organization's interest (presumably) is not providing information or recommendations that maximize user satisfaction or facilitate giving in a generic sense but in maximizing their support. Cause-based organizations may, of course, still benefit from algorithms used by giving platforms or general-purpose information providers prioritizing them in search results or including them in recommendations. In this situation, do these cause-based organizations, as beneficiaries, bear any responsibility for concerns about algorithmic bias, loss of autonomy for users, or lack of transparency? Since we assume they neither designed the algorithms nor controlled their operation, these cause-based organizations are not directly responsible. However, if they are aware of the issues or have even engaged in practices that might exacerbate them (such as paying to be ranked higher in search results or participating in hyper-nudging initiatives where they stand to receive donations), then it might be argued that they are complicit to some extent and therefore bear a share of any moral responsibility.

Cause-based organizations may also face direct ethical issues if they use generative AI tools to produce content as part of their fundraising. Charities and nonprofits have always used emotive imagery and storytelling to appeal to people via the heart and the head, which has sometimes drawn criticism. In the context of international aid and development, for instance, there is a long-standing debate over whether depictions of aid recipients are patronizing and overly negative and whether this reflects problematic attitudes of "white saviorism" (Pieterse, 1992; Bhati, 2021). Similarly, disability rights campaigners have, at times, been vocally critical of the depiction of disabled people as objects of pity by nonprofit organizations (Longmore, 2015). These concerns will apply equally to content created using generative AI. In fact, they may even be exacerbated, as there are concerns that current AI image-generation tools demonstrate worrying levels of racial and gender bias, so their use may lead to the further perpetuation of problematic stereotypes (Lamensch, 2023; Small, 2023; Thomas & Thomson, 2023; Turk, 2023).

The capabilities of generative AI will bring other challenges, too. The fact that it is now possible to generate photo-realistic images or deepfake videos that are indistinguishable from real photos or video has led to concerns being raised about the potential impact this might have on notions of trust and authenticity (de Ruiter, 2021). For cause-based organizations, the risk of using generative AI content is that if done badly, it may have a significant negative impact on the perceptions of supporters and the wider public. In 2019, the UK charity Malaria No More UK successfully used deepfake technology to produce footage of the former footballer David Beckham reading out an appeal in nine different languages. It received broadly positive coverage (Davies, 2019). In early 2023, however, Amnesty International was heavily criticized for using AI-generated imagery to promote a report on police brutality in Colombia (Taylor, 2023). In the context of fundraising, there may also be reasons for caution when it comes to embracing generative AI: at least one study, for instance, has found that people are less likely to donate in response to a charitable giving advert if they become aware that AI-generated imagery has been used (Arango et al., 2023).

There are also wider ethical questions as to whether generative AI is inherently parasitic since it requires vast data sets of images, photographs, or text on which algorithms can be trained, which are only possible through the past efforts of human writers, artists, and creatives. Critics have accused companies at the forefront of the generative AI revolution of engaging in deliberate, large-scale copyright infringement to build their products, and this seems set to become a major issue for generative AI in the coming years, with several legal challenges already mounted and more likely to follow (Appel et al., 2023; Bearne, 2023). Again, this is not an issue for which charities or nonprofits bear direct responsibility, as they have no control over the development of commercial generative AI tools. They do, however, have a choice as to whether they use these tools or not. While broad concerns about copyright infringement and intellectual property rights are unlikely to be sufficient grounds by themselves for shunning generative AI, when added to the other ethical issues outlined already, it may be seen by some charities and nonprofits as sufficient grounds to question whether the use of generative AI is appropriate at all at this stage.

5 Conclusion

In this chapter, we have considered a range of ways in which AI could be used to shape individual choices and how these apply to different contexts in which decisions about philanthropic giving might be made. In doing so, we have identified several potential ethical concerns that funders and nonprofits need to be aware of as they contemplate using AI tools in this way. These ethical concerns and some of the key questions we need to ask as a result are summarized below:

Legitimacy, accountability, and transparency

- Is it clear to users that an algorithm determines the information or recommendations offered to them?
- Who designed the algorithm?
- What goal did they have in mind in designing the algorithm?
- Is this goal clear to the user?
- Where does their legitimacy of the goal derive from?
- Can the user challenge the algorithm or hold the designer to account?

Undermining agency

- Does the use of AI to create personalized choice architectures or hyper-nudges undermine the agency of individual donors?
- Is this an acceptable price to pay for encouraging prosocial behavior?

Bias

- Does using algorithms to provide information and recommendations that can shape philanthropic choice introduce the risk of bias against certain kinds of organizations or causes?
- Does the use of Generative AI bring the risk of bias (e.g., for race, gender, etc.), which could exacerbate existing concerns about how recipients of philanthropic funding are portrayed?

Erosion of authenticity and trust

- Does the use of generative AI to produce content (e.g., photographic imagery, video, text) for use in fundraising and campaigning bring the risk of contributing to an erosion of the notions of authenticity and trust in the online environment?

Intellectual property rights and training data

- Are current AI tools dependent on data sets that take advantage of prior human effort without suitable compensation or respect for intellectual property rights? Are nonprofit organizations that use these tools ethically compromised as a result?

Having identified these ethical issues, the question is what actions can be taken to address them. There is no single answer, but we can identify a range of current and future actions that can be taken by the various actors involved. Commercial platforms need to recognize the power they have to shape individual giving choices. They should seek to work with the nonprofit sector to minimize any potential harms that come from applying algorithmic processes to the provision of information and recommendations. Dedicated giving platforms are more likely to have pre-existing relationships with the nonprofit sector but should still recognize that their power as gatekeepers will increase as more and more giving takes places online and that they therefore have a responsibility to ensure that any applications of algorithms take into account the risks of undermining the agency of donors, or introducing biases that will adversely affect certain types of organizations or cause areas. Nonprofits and civil society organizations also need to be aware of the potential risks inherent in their own use of AI tools – and ensure they put in place measures to mitigate against them. This may seem daunting for organizations that are often resource-starved, which is why philanthropic funders also have a vital role in providing them with the infrastructure and support they need to engage with issues of AI ethics. This could be done through relationships with individual grantees or by creating new pooled funds (such as the European Artificial Intelligence and Society Fund). Philanthropic funders can also play a valuable role in broader efforts to ensure the ethical and responsible use of AI, by engaging in research and advocacy that enables the perspectives and insights of civil society organizations they support to be brought into the debates.

References

Allen, S. (2018). *The Science of Generosity* [White Paper]. Prepared for the John Templeton Foundation by the Greater Good Science Center at UC Berkeley. https://ggsc.berkeley.edu/images/uploads/GGSC-JTF_White_Paper-Generosity-FINAL.pdf

Alpizar, F., Carlsson, F., & Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics*, *92*(5–6), 1047–1060. https://doi.org/10.1016/j.jpubeco.2007.11.004

André, Q., Carmon, Z., Wertenbroch, K., Crum, A., Frank, D., Goldstein, W., Huber, J., Van Boven, L., Weber, B., & Yang, H. (2018). Consumer choice and autonomy in the age of artificial intelligence and big data. *Customer Needs and Solutions*, *5*(1–2), 28–37. https://doi.org/10.1007/s40547-017-0085-8

Andreoni, J., & Payne, A. (2011). Is crowding out due entirely to fundraising? Evidence from a panel of charities. *Journal of Public Economics, 95*(5–6), 334–343. https://doi.org/10.1016/j.jpubeco.2010.11.011

Andreoni, J., Payne, A. A., Smith, J., & Karp, D. (2016). Diversity and donations: The effect of religious and ethnic diversity on charitable giving. *Journal of Economic Behavior & Organization*, *128*, 47–58. https://doi.org/10.1016/j.jebo.2016.05.010

Andreoni, J., Rao, J. M., & Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy*, *125*(3), 625–653. https://doi.org/10.1086/691703

Appel, G., Neelbauer, J., & Schweidel, D. (2023, April 7). Generative AI has an intellectual property problem. *Harvard Business Review*. https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem

Arango, L., Singaraju, S. P., & Niininen, O. (2023). Consumer responses to AI-generated charitable giving ads. *Journal of Advertising*, *52*(4), 486–503. https://doi.org/10.1080/00913367.2023.2183285

Arora, N. (2021, November 12). *The Value of Getting Personalization Right—Or Wrong—Is Multiplying*. McKinsey. https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying

Barraza, J. A., McCullough, M. E., Ahmadi, S., & Zak, P. J. (2011). Oxytocin infusion increases charitable donations regardless of monetary resources. *Hormones and Behavior*, *60*(2), 148–151. https://doi.org/10.1016/j.yhbeh.2011.04.008

Bartmann, M. (2023). Reasoning with recommender systems? Practical reasoning, digital nudging, and autonomy. In S. Genovesi, K. Kaesling, & S. Robbins (Eds.), *Recommender Systems: Legal and Ethical Issues* (Vol. 40, pp. 129–145). Springer International Publishing. https://doi.org/10.1007/978-3-031-34804-4_7

Bearne, S. (2023, August 1). New AI systems collide with copyright law. *BBC*. https://www.bbc.co.uk/news/business-66231268

Beeston, E., & Breeze, B. (2023, May 16). *Disciplining Generosity*. Stanford Social Innovation Review. https://ssir.org/books/excerpts/entry/disciplining_generosity

Behavioural Insights Team (2013). *Applying Behavioural Insights to Charitable Giving* [White Paper]. Crown. https://assets.publishing.service.gov.uk/media/5a7516a1ed915d6faf2b228b/BIT_Charitable_Giving_Paper.pdf

Bekkers, R., & Wiepking, P. (2011). A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving. *Nonprofit and Voluntary Sector Quarterly*, *40*(5), 924–973. https://doi.org/10.1177/0899764010380927

Bernholz, L. (2021). *How We Give Now: A Philanthropic Guide for the Rest of Us*. Cambridge: MIT Press.

Bhati, A. (2021). Is the representation of beneficiaries by international nongovernmental organizations (INGOs) still pornographic? *Journal of Philanthropy and Marketing*, e1722. https://doi.org/10.1002/nvsm.1722

Boden, M. A. (2014). 4 GOFAI. In K. Frankish and W.M. Ramsey (Eds.). *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139046855.007

Body, A., & Breeze, B. (2016). What are 'unpopular causes' and how can they achieve fundraising success? *International Journal of Nonprofit and Voluntary Sector Marketing*, *21*(1), 57–70. https://doi.org/10.1002/nvsm.1547

Bostrom, N. (2003). Ethical issues in advanced artificial intelligence. In I. Smith et al. (Eds.). *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence* (Vol 2, ed. I, pp. 12–17). Institute of Advanced Studies in Systems Research and Cybernetics. https://philpapers.org/archive/BOSEII.pdf

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bovens, L. (2009). The ethics of nudge. In T. Grüne-Yanoff and S. O. Hansson (Eds.). *Preference Change: Approaches from Philosophy, Economics and Psychology* (pp. 207–219). Springer. https://philpapers.org/archive/BOVTEO-8.pdf

BrightEdge. (2019). *Channel Report 2019*. BrightEdge. https://videos.brightedge.com/research-report/BrightEdge_ChannelReport2019_FINAL.pdf

Brown, P. H., & Minty, J. H. (2008). Media coverage and charitable giving after the 2004 tsunami. *Southern Economic Journal*, *75*(1), 9–25. https://doi.org/10.1002/j.2325-8012.2008.tb00889.x

Butera, L., & Horn, J. (2020). "Give less but give smart": Experimental evidence on the effects of public information about quality on giving. *Journal of Economic Behavior & Organization*, *171*, 59–76. https://doi.org/10.1016/j.jebo.2020.01.011

Capraro, V., Jagfeld, G., Klein, R., Mul, M., & van de Pol, I. (2019). Increasing altruistic and cooperative behaviour with simple moral nudges. *Scientific Reports*, *9*(1), 11880. https://doi.org/10.1038/s41598-019-48094-4

Carlman, A. (2020, November 17). *Platform Neutrality Is Dead. Long Live Empathy*. Digital Impact. https://digitalimpact.io/platform-neutrality-is-dead-long-live-empathy/

Caulfield, J. L., Baird, C. A., & Lee, F. K. (2022). The ethicality of point-of-sale marketing campaigns: Normative ethics applied to cause-related checkout charities. *Journal of Business Ethics*, 1–16. https://doi.org/10.1007/s10551-020-04597-z

Chesney, R., & Citron, D. (2018, December 11). Deepfakes and the new disinformation war. *Foreign Affairs*. https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war

Cordelli, C. (2016). Reparative justice and the moral limits of discretionary philanthropy. In R. Reich, C. Cordelli & L. Bernholz (Eds.). *Philanthropy in Democratic Societies: History, Institutions, Values* (pp. 244–265). Chicago, IL: The University of Chicago Press.

Crary, A. (2023). Against "Effective Altruism". In C. J. Adams, A. Crary & L. Gruen (Eds.). *The Good It Promises, the Harm It Does: Critical Essays on Effective Altruism*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780197655696.003.0016

Croson, R., & Shang, J. (Yue) (2008). The impact of downward social information on contribution decisions. *Experimental Economics*, *11*(3), 221–233. https://doi.org/10.1007/s10683-007-9191-z

Damm, C., Dowrick, L., & Harris, C. (2023). *Mapping the UK Women and Girls Sector and Its Funding: Where Does the Money Go?* Rosa. https://rosauk.org/wp-content/uploads/2023/04/Women-and-Girls-Sector-Research-Mapping-Report-Amended.pdf

Davies, G. (2019, April 9). David Beckham 'Speaks'9 languages for new campaign to end malaria. *ABC News*. https://abcnews.go.com/International/david-beckham-speaks-languages-campaign-end-malaria/story?id=62270227

De Fine Licht, K., & De Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & SOCIETY*, *35*, 917–926. https://rdcu.be/dprkT

De Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology*, *34*(4), 1311–1332. https://doi.org/10.1007/s13347-021-00459-2

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, *23*(7), 2072–2098. https://doi.org/10.1177/1461444820925811

Dorsey, C., Bradach, J., & Kim, P. (2020). *Racial Equity and Philanthropy: Disparities in Funding for Leaders of Color Leave Impact on the Table.* The Bridgespan Group. https://www.bridgespan.org/getmedia/05ad1f12-2419-4039-ac67-a45044f940ec/racial-equity-and-philanthropy.pdf

Eckel, C., Grossman, P. J., & Milano, A. (2007). Is more information always better? An experimental study of charitable giving and Hurricane Katrina. *Southern Economic Journal*, *74*(2), 388–411. https://doi.org/10.1002/j.2325-8012.2007.tb00845.x

Ekström, M. (2012). Do watching eyes affect charitable giving? Evidence from a field experiment. *Experimental Economics*, *15*(3), 530–546. https://doi.org/10.1007/s10683-011-9312-6

Espín-Noboa, L., Wagner, C., Strohmaier, M., & Karimi, F. (2022). Inequality and inequity in network-based ranking and recommendation algorithms. *Scientific Reports*, *12*(1), 2012. https://doi.org/10.1038/s41598-022-05434-1

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Fathi, M., Bateson, M., & Nettle, D. (2014). Effects of watching eyes and norm cues on charitable giving in a surreptitious behavioral experiment. *Evolutionary Psychology*, *12*(5). https://doi.org/10.1177/147470491401200502

Fukui, H., & Toyoshima, K. (2014). Chill-inducing music enhances altruism in humans. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01215

Gabriel, I. (2017). Effective altruism and its critics. *Journal of Applied Philosophy*, *34*(4), 457–473. https://doi.org/10.1111/japp.12176

Gewirth, A. (1987). Private philanthropy and positive rights. *Social Philosophy and Policy*, *4*(2), 55–78. https://doi.org/10.1017/S0265052500000546

Greitemeyer, T. (2009). Effects of songs with prosocial lyrics on prosocial behavior: Further evidence and a mediating mechanism. *Personality and Social Psychology Bulletin*, *35*(11), 1500–1511. https://doi.org/10.1177/0146167209341648

Gurdeniz, E., & Hosanagar, K. (2023, February 23). Generative AI won't revolutionize search—yet. *Harvard Business Review*. https://hbr.org/2023/02/generative-ai-wont-revolutionize-search-yet

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

Harbaugh, W. T., & Krause, K. (2000). Children's altruism in public good and dictator experiments. *Economic Inquiry*, *38*(1), 95–109. https://doi.org/10.1111/j.1465-7295.2000.tb00006.x

Hobbs, J. (2017). Nudging charitable giving: The ethics of Nudge in international poverty reduction. *Ethics & Global Politics*, *10*(1), 37–57. https://doi.org/10.1080/16544951.2017.1312991

Houben, S., Abrecht, S., Akila, M., Bär, A., Brockherde, F., Feifel, P., Fingscheidt, T., Gannamaneni, S. S., Ghobadi, S. E., & Hammam, A. (2022). Inspect, understand, overcome: A survey of practical methods for AI safety. In T. Fingscheidt, H. Gottschalk & S. Houben (Eds.). *Deep Neural Networks and Data for Automated Driving: Robustness, Uncertainty Quantification, and Insights towards Safety* (pp. 3–78). Cham: Springer International Publishing Cham. https://doi.org/10.1007/978-3-031-01233-4_1

Huang, K. (2022, September 17). For Gen Z, TikTok is the new search engine. *The New York Times*. https://www.nytimes.com/2022/09/16/technology/gen-z-tiktok-search-engine.html

Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, *14*, 235–257. https://doi.org/10.1023/A:1007740225484

Jonas, E., Schimel, J., Greenberg, J., & Pyszczynski, T. (2002). The scrooge effect: Evidence that mortality salience increases prosocial attitudes and behavior. *Personality and Social Psychology Bulletin*, *28*(10), 1342–1353. https://doi.org/10.1177/014616702236834

Kanter, B., & Fine, A. H. (2022). *The Smart Nonprofit: Staying Human-Centered in an Automated World*. Hoboken, NJ: John Wiley & Sons, Incorporated.

Karlan, D., & List, J. A. (2007). Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *American Economic Review*, *97*(5), 1774–1793. http://dx.doi.org/10.1257/aer.97.5.1774

Korinek, A., & Balwit, A. (2022). *Aligned with Whom? Direct and Social Goals for AI Systems* [NBER Working Paper No. w30017]. National Bureau of Economic Research. https://dx.doi.org/10.2139/ssrn.4104003

Krakovna, V., Uesato, J., Mikulik, V., Rahtz, M., Everitt, T., Kumar, R., Kenton, Z., Leike, J., & Legg, S. (2020, April 21). *Specification Gaming: The Flip Side of AI Ingenuity*. DeepMind Blog, 3. https://www.deepmind.com/blog/specification-gaming-the-flip-side-of-ai-ingenuity

Lamensch, M. (2023, June 14). *Generative AI Tools Are Perpetuating Harmful Gender Stereotypes*. Centre for International Governance Innovation. https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/

Lehman, J., Clune, J., Misevic, D., Adami, C., Altenberg, L., Beaulieu, J., Bentley, P. J., Bernard, S., Beslon, G., & Bryson, D. M. (2020). The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities. *Artificial Life*, *26*(2), 274–306. https://doi.org/10.1162/artl_a_00319

Leslie, D. (2019). *Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector*. The Alan Turing Institute. https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

Liao, Q. V., Geyer, W., Muller, M., & Khazaen, Y. (2020). Conversational interfaces for information search. In W. T. Fu & H. van Oostendorp (Eds.). *Understanding and Improving Information Search: A Cognitive Approach* (pp. 267–287). Cham: Springer Cham.

Lilly Family School of Philanthropy (2023). *Digital for Good: A Global Study on Emerging Ways of Giving*. Lilly Family School of Philanthropy, Indiana University-Purdue University Indianapolis. https://scholarworks.iupui.edu/server/api/core/bitstreams/528c7dd7-4fb0-4af9-875a-1254a4e3034d/content

Loi, M., & Spielkamp, M. (2021). Towards accountability in the use of artificial intelligence for public administrations. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 757–766. https://doi.org/10.1145/3461702.3462631

Longmore, P. K. (2015). *Telethons: Spectacle, Disability, and the Business of Charity*. Oxford: Oxford University Press.

Lv, L., & Huang, M. (2022). Can personalized recommendations in charity advertising boost donation? The role of perceived autonomy. *Journal of Advertising*, 1–18. https://doi.org/10.1080/00913367.2022.2109082

MacAskill, W. (2015). *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference*. London: Guardian Faber Publishing.

MacQuillin, I. (2022). Normative fundraising ethics: A review of the field. *Journal of Philanthropy and Marketing*, e1740. https://doi.org/10.1002/nvsm.1740

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2023). *Dissociating Language and thought in Large Language Models: A Cognitive Perspective* (arXiv: 2301.06627). arXiv. http://arxiv.org/abs/2301.06627

Maillé, P., Maudet, G., Simon, M., & Tuffin, B. (2022). Are search engines biased? Detecting and reducing bias using meta search engines. *Electronic Commerce Research and Applications*, 101132. https://doi.org/10.1016/j.elerap.2022.101132

Martin, M. W. (1994). *Virtuous Giving: Philanthropy, Voluntary Service, and Caring*. Indianapolis: Indiana University Press.

Meer, J. (2011). Brother, can you spare a dime? Peer pressure in charitable solicitation. *Journal of Public Economics*, *95*(7–8), 926–941. https://doi.org/10.1016/j.jpubeco.2010.11.026

Mills, S. (2022). Finding the 'nudge' in hypernudge. *Technology in Society*, *71*, 102117. https://doi.org/10.1016/j.techsoc.2022.102117

Mills, S., & Sætra, H. S. (2022). The autonomous choice architect. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01486-z

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 279–288. https://doi.org/10.1145/3287560.3287574

Moll, J., Krueger, F., Zahn, R., Pardini, M., De Oliveira-Souza, R., & Grafman, J. (2006). Human fronto–mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences*, *103*(42), 15623–15628. https://doi.org/10.1073/pnas.0604475103

National Council for Voluntary Organisations (2023). *UK Civil Society Almanac 2023*. National Council for Voluntary Organisations. https://www.ncvo.org.uk/news-and-insights/news-index/uk-civil-society-almanac-2023/financials/#/

Noble, S. U. (2018). *Algorithms of Oppression*. New York: New York University Press.

Nozick, R. (1974). *Anarchy, State, and Utopia*. Hoboken, NJ: John Wiley & Sons.

Pashentsev, E. (2023). The malicious use of deepfakes against psychological security and political stability. In E. Pashentsev (Ed.). *The Palgrave Handbook of Malicious Use of AI and Psychological Security* (pp. 47–80). Cham: Palgrave Macmillan Cham. https://doi.org/10.1007/978-3-030-35746-7_3

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Payton, R. L., & Moody, M. P. (2008). *Understanding Philanthropy: Its Meaning and Mission*. Indianapolis: Indiana University Press.

Perez, S. (2022, July 12). *Google Exec Suggests Instagram and TikTok Are Eating into Google's Core Products, Search and Maps*. TechCrunch. https://techcrunch.com/2022/07/12/google-exec-suggests-instagram-and-tiktok-are-eating-into-googles-core-products-search-and-maps/

Pieterse, J. N. (1992). *White on Black: Images of Africa and Blacks in Western Popular Culture*. New Haven, CT: Yale University Press.

Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology*, *99*(5), 771–784. https://doi.org/10.1037/a0020092

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*, 137–141. https://doi.org/10.1007/s11747-019-00710-5

Reich, R., Cordelli, C., & Bernholz, L. (2016). *Philanthropy in Democratic Societies: History, Institutions, Values*. Chicago, IL: University of Chicago Press.

Roberts, S. (1996). Contexts of charity in the middle ages: Religious, social, and civic. In J.B. Schneewind (Ed.). *Giving: Western Ideas of Philanthropy* (pp. 24–53). Indianapolis: Indiana University Press.

Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, *9*(1), 59. https://doi.org/10.1186/s40537-022-00592-5

Rudd, M., Vohs, K. D., & Aaker, J. (2012). Awe expands people's perception of time, alters decision making, and enhances well-being. *Psychological Science*, *23*(10), 1130–1136. https://doi.org/10.1177/0956797612438731

Ruehle, R. C., Engelen, B., & Archer, A. (2021). Nudging charitable giving: What (if anything) is wrong with it? *Nonprofit and Voluntary Sector Quarterly*, *50*(2), 353–371. https://doi.org/10.1177/0899764020954266

Saeri, A. K., Slattery, P., Lee, J., Houlden, T., Farr, N., Gelber, R. L., Stone, J., Huuskes, L., Timmons, S., Windle, K., Spajic, L., Freeman, L., Moss, D., Behar, J., Schubert, S., Grundy, E. A. C., & Zorker, M. (2023). What works to increase charitable donations? A meta-review with meta-meta-analysis. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *34*(3), 626–642. https://doi.org/10.1007/s11266-022-00499-y

Sætra, H. S. (2023). Generative AI: Here to stay, but for good? *Technology in Society*, *75*, 102372. https://doi.org/10.1016/j.techsoc.2023.102372

Salmon, J. (2023). *Protecting Donor Intent Protects Giving*. The Philanthropy Roundtable. https://www.philanthropyroundtable.org/wp-content/uploads/2023/07/Protecting-Donor-Intent-Protects-Giving_.pdf

Samuel, L. R. (2010). *Freud on Madison Avenue: Motivation Research and Subliminal Advertising in America*. Philadelphia: University of Pennsylvania Press.

Schmidt, A. (2021). The end of serendipity: Will artificial intelligence remove chance and choice in everyday life? *CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter*, 1–4. https://doi.org/10.1145/3464385.3464763

Schneewind, J. B. (1996). *Giving: Western Ideas of Philanthropy*. Indianapolis: Indiana University Press.

Schrage, M. (2020). *Recommendation Engines*. Cambridge: MIT Press.

Schulz, J. F., Thiemann, P., & Thöni, C. (2017). Nudging generosity: Choice architecture and cognitive factors in charitable giving. *Journal of Behavioral and Experimental Economics*, *74,*139–145. https://doi.org/10.1016/j.socec.2018.04.001

Sebastián-Enesco, C., & Warneken, F. (2015). The shadow of the future: 5-year-olds, but not 3-year-olds, adjust their sharing in anticipation of reciprocation. *Journal of Experimental Child Psychology*, *129*, 40–54. https://doi.org/10.1016/j.jecp.2014.08.007

Shang, J., & Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal*, *119*(540), 1422–1439. https://doi.org/10.1111/j.1468-0297.2009.02267.x

Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming god concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*(9), 803–809. https://doi.org/10.1111/j.1467-9280.2007.01983.x

Singer, P. (1972). Famine, affluence, and morality. *Philosophy and Public Affairs*, *1*(3), 229–243. https://www.jstor.org/stable/2265052

Singer, P. (2006, December 17) What should a billionaire give and what should you? *New York Times*. https://www.nytimes.com/2006/12/17/magazine/17charity.t.html

Singer, P. (2015). *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically*. Melbourne: Text Publishing.

Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision Making*, *2*(2), 79–95. https://doi.org/10.1017/S1930297500000061

Small, Z. (2023, July 4). Black artists say AI shows bias, with algorithms erasing their history. *New York Times*. https://www.nytimes.com/2023/07/04/arts/design/black-artists-bias-ai.html

Smith, B. C. (2019). *The Promise of Artificial Intelligence: Reckoning and Judgment*. Cambridge: MIT Press.

Sparks, A., & Barclay, P. (2013). Eye images increase generosity, but not for long: The limited effect of a false cue. *Evolution and Human Behavior*, 34(5), 317–322. https://doi.org/10.1016/j.evolhumbehav.2013.05.001

Srinivasan, A. (2015). Stop the robot apocalypse: The new utilitarians. *London Review of Books*, *37*(18), 3–6. https://www.lrb.co.uk/the-paper/v37/n18/amia-srinivasan/stop-the-robot-apocalypse

Susser, D. (2017). Transparent media and the development of digital habits. In Y. Van den Eede, S. O. Irwin & G. Wellner (Eds.). *Postphenomenology and Media: Essays on Human-Media-World Relations* (pp. 27–44). Lexington Books. https://philarchive.org/rec/SUSTMA

Susser, D. (2019). Invisible influence: Artificial intelligence and the ethics of adaptive choice architectures. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 403–408. https://doi.org/10.1145/3306618.3314286

Swain, R. B. (2018). A critical analysis of the sustainable development goals. In W. L. Filho (Ed). *Handbook of Sustainability Science and Research* (pp. 341–355). Cham: Springer Cham

Taylor, L. (2023, May 2). Amnesty international criticised for using AI-generated images. *The Guardian*. https://www.theguardian.com/world/2023/may/02/amnesty-international-ai-generated-images-criticism

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. London: Penguin.

Thomas, A. M., Parkinson, J., Moore, P., Goodman, A., Xhafa, F., & Barolli, L. (2013). Nudging through technology: Choice architectures and the mobile information revolution. *2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, 255–261. https://doi.org/10.1109/3PGCIC.2013.44

Thomas, R. J., & Thomson, T. J. (2023). What does a journalist look like? Visualizing journalistic roles through AI. *Digital Journalism*, 1–23. https://doi.org/10.1080/21670811.2023.2229883

Turk, V. (2023, October 10). *How AI Reduces the World to Stereotypes*. Rest of World. https://restofworld.org/2023/ai-image-stereotypes/

Vaassen, B. (2022). AI, opacity, and personal autonomy. *Philosophy & Technology*, *35*(4), 88. https://doi.org/10.1007/s13347-022-00577-5

Van Den Eede, Y. (2011). In between us: On the transparency and opacity of technological mediation. *Foundations of Science*, *16*, 139–159. https://doi.org/10.1007/s10699-010-9190-y

von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, *34*(4), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

Wachter-Boettcher, S. (2017). *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*. New York: WW Norton & Company.

Wade, M. (2022). 'The giving layer of the internet': A critical history of GoFundMe's reputation management, platform governance, and communication strategies in capturing peer-to-peer and charitable giving markets. *Journal of Philanthropy and Marketing*, e1777. https://doi.org/10.1002/nvsm.1777

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301–1303. https://doi.org/10.1126/science.1121448

Weinmann, M., Schneider, C., & Brocke, J. V. (2016). Digital nudging. *Business & Information Systems Engineering*, *58*(6), 433–436. https://doi.org/10.1007/s12599-016-0453-1

Winfrey, J. C. (1981). Charity versus justice in Locke's theory of property. *Journal of the History of Ideas*, *42*(3), 423. https://doi.org/10.2307/2709185

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector – applications and challenges. *International Journal of Public Administration*, *42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, *20*(1), 118–136. https://doi.org/10.1080/1369118X.2016.1186713

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, *28*(1), 126. https://psycnet.apa.org/doi/10.1037/0012-1649.28.1.126

Zak, P. J., Stanton, A. A., & Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS One*, *2*(11), e1128. https://doi.org/10.1371/journal.pone.0001128

Zamani, H., Trippas, J. R., Dalton, J., & Radlinski, F. (2023). *Conversational Information Seeking* (arXiv: 2201.08808). arXiv. http://arxiv.org/abs/2201.08808