TOWARD A FRAMEWORK FOR RESPONSIBLE AI IN STORYTELLING FOR NONPROFIT FUNDRAISING

Marta Herrero and Shauna Concannon

1 Introduction

Nonprofits rely heavily on fundraising and eliciting public generosity and support. In 2019–2020, the public contributed 51% of the charity sector's total income of £58.7 billion (NCVO, 2022), and spent £12.7 billion in individual donations alone (UK Giving, 2023). Equally, UK charities invest heavily in fundraising activities, spending £7.7 billion in 2019–2020 (NCVO, 2022). AI techniques offer novel and compelling possibilities for nonprofit fundraising (e.g., data science applications can provide a deeper understanding of audiences and donors, and generative methods can create more personalized and persuasive communications), compared to pre-AI digital counterparts (e.g., online giving platforms and social media fundraising to motivate donors).

And yet, the promise of AI to transform how nonprofits fundraise is faced with its set of general and sector-specific challenges. A lack of AI knowledge and understanding across the sector around complex issues such as the responsible handling of donor data, legal and regulatory compliance, and microtargeting implications can be challenging for non-AI experts to navigate. Threats to cybersecurity, data privacy, and ethical concerns when processing sensitive donor information can not only lead to gender, racial bias, and discrimination but also potentially damage the trust built between nonprofits, donors, and beneficiaries. Attempts to address such challenges must take into consideration further sector-specific characteristics that threaten AI uptake: (i) digital inequalities – 82% of the UK charity sector is made up of small and micro nonprofits, with small fundraising teams and low budgets (Charity Digital Skills Report, 2021); (ii) digital deficit – only 56% of nonprofits are reported to invest in digital fundraising technologies (Blackbaud, 2022); and (iii) high employee turnover and retention (Sargeant & Edworthy, 2022), making it difficult to provide and build long-term capacity in AI training.

This chapter proposes a framework model for AI resilience capabilities in nonprofit fundraising that explores the types of knowledges and practices organizations must acquire to understand and practically orientate themselves to using and/or responding to the widespread use of AI technologies. We will focus on AI storytelling as a key form of communication which helps nonprofits create emotional bonds based on empathy, and that can thus help ensure their long-term donor loyalty and financial support. The concept of "organizational resilience" remains unexplored in the nonprofit management literature but has been widely used to indicate how for-profit organizations

Marta Herrero and Shauna Concannon

react against external crises by developing ways of working and understanding that reduce uncertainty (caused by the crisis) and promote and restore balance. We argue that the so-called "AI revolution," which has led to the proliferation and availability of AI-supported technologies and applications, can be seen as posing a similar "external crisis" to nonprofit fundraising. As Cantin and Michel (2003) note, the arrival of new technologies always leads to organizational uncertainty about their uses and adoption, and AI is no exception.

One of the main issues raised by AI is the effects it can have on moral standards. Ethical concerns around the processing of sensitive donor information, which can lead to gender, racial bias, and discrimination, have been consistently raised as posing threats specific to fundraising practices such as communications with donors and the generation of biased databases. In this chapter, we address the issue of how AI can affect trust in the donor-nonprofit relationship by focusing on digital storytelling (Longoni et al., 2022). The arrival of AI offers appealing possibilities and applications that can transform how fundraisers plan and carry out their storytelling. However, for some, the current transition and change of digital or pre-AI storytelling and toward AI storytelling is a risky endeavor. For it to be entirely accepted and relied upon by supporters, fundraisers must overcome a public lack of trust in machine-authored stories and the reliance on trust generated by human-authored storytelling, especially in the accuracy and veracity of the information presented.

The chapter is divided into three parts. The first part introduces the notion of Responsible AI (RAI) resilience capabilities framework, drawing upon management and organizational theory literature and research on "resilience capabilities" in nonprofit fundraising during Covid-19 (Herrero & Kraemer, 2022). The second part brings together the four types of RAI resilience capabilities included in our framework with literature on digital storytelling for fundraising. The third part and conclusion to the chapter argue that AI storytelling for fundraising poses clear challenges to how nonprofits continue building trust in their relationships with external stakeholders, especially with existing and future donors. Being aware of such challenges, however, is the first step for fundraisers to address them, and we argue this involves developing "behavioral-cum-emotional capabilities." These types of capabilities include two intertwined steps: first, understanding that AI storytelling has led to changes in behavior, such as a lack of trust in the technology and, subsequently, in the stories and narratives used for nonprofit fundraising. Second, in this new context, fundraisers need to develop ways of keeping their donors emotionally engaged and thus loyal to their cause and organization, while using AI storytelling. The chapter concludes by comparing the capabilities developed by fundraisers during the Covid-19 crisis with those needed to address some of the threats posed by AI storytelling, which can mitigate and/or avoid a decrease in levels of trust and loyalty among donors and members.

2 Resilience capabilities and nonprofit fundraising: lessons from the Covid-19 pandemic

The link between AI and organizational resilience is not new. It is widely acknowledged that if businesses are to build organizational resilience, which is necessary to create "sustainable enterprises," they must accept that AI will radically change company structures, culture, and operations. However, "resilience" is a commonly used term in academic literature. It is used to designate individual and organizational practices responding positively to a setback or crisis in various contexts (Hickman, 2018). Not surprisingly, academic research on resilience continues to rapidly increase on par with the substantial disruptions – from terrorist attacks to financial downturns and, most recently, the Covid-19 pandemic – that create an environment of uncertainty and crisis (Lengnick-Hall & Beck, 2005). When an organization demonstrates a collective capability for resilience, it means that it can understand "the current situation" as well as "to develop customized responses" that reflect such "understanding" (Lengnick-Hall & Beck, 2005, p. 750). The ability to understand and to respond is not necessarily present in all organizational environments, which means that support mechanisms and/or practices must be put in place to start developing and integrating "resilience capabilities."

Our focus is on "resilience capabilities" as they are carried out at the level of practice in the interactions of fundraisers within their organizational environment, which has been affected by differing degrees of uncertainty and crisis, and with external actors. In what follows, we draw on the literature on resilience responses to outline the various practice-based resilience capabilities that arts and culture fundraisers developed to address and cope with the consequences of the Covid-19 pandemic. We draw on research carried out by Herrero and Kraemer (2022), in which they interviewed a small group of fundraisers following a sector-wide questionnaire. These initial insights will help understand how resilience concepts are adapted to understand nonprofit fundraising. In turn, these conceptual classifications provide an ideal starting point for thinking about how, as we argue later on, the advent of AI in fundraising poses a similar form of organizational and relational "crisis" to nonprofits. That is, whether they choose to adopt AI or can afford to, in their day-to-day fundraising nonprofits must respond to the opportunities and challenges of AI. In so doing, they will develop different levels of organizational resilience.

2.1 Cognitive capabilities

Cognitive capabilities refer to the existence of an understanding of knowledge and repertories of action, such as vision and sense of purpose, that are applied to the resolution of problems (Lengnick-Hall & Beck, 2005). In explaining their thinking about the effects of the pandemic on fundraising, interviewees shared a clear understanding of what changes needed to be implemented to fundraise most effectively.

For example, reassessing the type of messages used to communicate with their stakeholders was deemed a priority. Some fundraisers mentioned that ethical issues also influenced how they engaged with donors. They changed their thinking about asking for money during the pandemic because they felt it was a sensitivity issue and that doing the right thing at the right time was more important than asking for money in a time of need. At a time when the death rate in the UK was very high, fundraisers did not think this was the right time to think about legacy fundraising (prompting individuals to leave donations to an organization in their will). A fundraiser explained a change in plans to launch a legacy campaign that had secured sponsorship support from a law firm. With the "death rate being so widely publicized," it was felt that this was not the "right time" to think about legacies (Interview 6). Similarly, a fundraiser referred to the decision not to launch any fundraising appeals because "it just never felt appropriate" at a time when other "frontline causes are needing attention."

2.2 Behavioral capabilities

The term "behavioral capabilities" is used to designate practical action alternatives that can steer a "dramatically different course of action from that which is the norm" (Lengnick-Hall & Beck, 2005, pp. 750–751). An example of fundraisers' deployment of behavior capabilities during the Covid-19 lockdowns was found in the fundraisers taking practical action and deciding on a donor retention strategy that saw them strengthening links with existing supporters rather than engaging with new ones. Donor retention was favored by fundraisers who could rapidly access digital

technologies, e.g., Zoom to organize online meetings. However, the pandemic has revealed a digital divide in access to digital technologies for fundraising and engaging with beneficiaries. This was particularly true for fundraisers working with deprived communities with very little online access.

An example of behavioral capabilities was using digital technologies in donor retention strategies. From a practical perspective, digital technology helped fundraisers connect with donors who "would ordinarily have had a bit of trouble scheduling in a meeting because they are so busy, but they will squeeze in an online call" (Interview 5). One interviewe noted that using digital technology to interact with donors was age-sensitive. For those over 50, Zoom meetings could be "very stressful." However, when she spoke with donors over the phone, and especially when making an ask, it "worked really well" as "everybody seems to relax a little bit more on a phone call" when they were "not being distracted by a picture of themselves" (Interview 9).

Fast-paced innovation was also an outcome of the pandemic, and drawing upon resilience and perseverance skills exemplifies further the use of existing cognitive capabilities. Fundraisers no longer had time to try and test "one or two new fundraising products in a year"; instead, everything "was new and very fast-paced" (Interview 5). Existing skills such as "resilience and perseverance" were "incredibly important" in dealing with the financial uncertainty. An interviewee described "resilience" as "being able to be knocked down and stand up because you didn't get a grant" (Interview 9). Another interviewee shared this opinion and argued that fundraisers were well-suited to deal with the pandemic's uncertainties as they were "more resilient," used to dealing with rejection and focused on getting on with their job (Interview 1) (Herrero & Kraemer, 2022).

2.3 Relational capabilities

Relational capabilities (Lengnick-Hall & Beck, 2005) refer to access and exchange of resources that can enhance an organization's positive functioning in the face of adversity. Evidence of relational capability building was found in fundraisers' ability to successfully draw on their external environment to build relationships and networks during the Covid-19 lockdowns. For example, having good relationships with funders and artists to support their fundraising was a case in point. It meant that those organizations with existing restricted funding could request for it to be made unrestricted so that it could be used to pay for core activities (Interview 15). Similarly, having good relationships with artists helped an organization but was by then well-known in the television and film industries worked as a fundraiser for over two months. The playwright got in touch with her industry contacts and made "asks" while also explaining the importance of supporting the organization, describing it as the "training ground for … writing talent in a world now where everybody is watching more television than ever" (Interview 15).

Our interviews also revealed how fundraisers gained relational capabilities by increasing their participation in existing networks and participating in new ones, which helped them gain skills and support. For example, a fundraiser explained how an existing network in the performing arts fundraising community became a "lifeline," with more frequent Zoom meetings every two weeks "just to talk about how things were going." Even though the group knew each other before the pandemic, they got to know each other better so that "myself, my peers, and my team are using and relying more on those sorts of networks." In this case, it allowed the fundraisers to make new connections that became "quite personally useful," and he could even meet in person with them after the lockdown (FR15) (Herrero & Kraemer, 2022).

2.4 Emotion-related capabilities

Emotion-related capabilities initially refer to the presence of mental fortitude that helps individuals cope with adverse situations and is expressed in the form of individual and collective optimism and hope. Having opportunities to communicate and discuss emotions will also likely enhance emotion regulation capabilities (Williams et al., 2017). Our research demonstrates how fundraisers drew upon emotion-related capabilities, seen here in their prior knowledge of donors who had shown an emotional attachment to the organization. This knowledge led fundraisers to prioritize and target such donors in their approaches.

For example, during the pandemic, donors who felt strong emotional ties to an organization also shared a sense of membership and inclusion as well as responsibility for the organization's financial well-being. A fundraiser explained that when donors felt strong emotional ties to the organization, they also shared a sense of membership, inclusion, and responsibility for the organization's financial well-being. Donors' loyalty ensured a steady level of donations, especially when the pandemic made it impossible for such donations to be reciprocated with any face-to-face benefits (FR15). Similarly, fundraisers drew upon emotion-related capabilities in their relationships with funders, as seen in the decision to adopt a "really open and honest approach with our funders ... more than we normally do." An example of honesty was the case of a fundraiser who felt that being open with existing funders, explaining the organization's difficult financial situation, and their fundraising plan for "recovering a loss of income" was the best approach. The fundraiser's strategy was to ask funders to repurpose some restricted income to spend the funds in "core" expenditures. Such a strategy paid off, as all funders agreed to the request that the fundraiser described as "an early exercise in honesty" (FR14) (Herrero & Kraemer, 2022).

3 Opportunities and challenges associated with the integration of AI for fundraising storytelling

The advent of generative AI has opened several avenues to support creative approaches to storytelling. Generative AI enables the production of images and text for storytelling purposes. For example, Large Language Models (LLMs), such as GPT-3 and LLAMA, have been effectively used to generate short stories, news articles, and other genres of text, while diffusion models can generate images and videos from text prompts (Esser et al., 2023; Singer & Polyak, 2022; Villegas et al., 2022). However, despite the promise and potential uses of generative models to produce storytelling materials and experiences, organizations looking to employ these systems need to navigate several known issues. Weidinger et al. (2022) highlight a number of potential risks, including compromising privacy by leaking sensitive information, reproduction of social stereotypes, and the potential for causing material harm due to the dissemination of false or poor-quality information. Bias and fabrication are critical challenges with serious implications and potential to distort human beliefs (Kidd & Birhane, 2023).

Fabrications and falsifications (also referred to as hallucinations, Ouyang et al., 2022) involve the creation of inaccurate, false, or misleading information. LLMs are essentially next-word prediction machines. While this is well-suited to generating largely coherent and plausible-sounding content that takes inspiration from the underlying patterns observed in the training data, it is not optimized for ensuring factual accuracy. In February 2022, Microsoft released Bing Chat, "an AI-powered assistant that can help you browse the web."¹ The chatbot was powered by OpenAI's LLM GPT-4. However, users quickly began sharing examples of the inaccuracies in the information provided by the system on social media – from confusion around the date an Avatar film was

Marta Herrero and Shauna Concannon

released,² to the water temperature at a beach in Mexico.³ Most importantly, the overly confident way it defended its inaccuracies and rebuked individuals for challenging its accuracy was what prompted concern among users. For example, in one interaction, when called out on the inaccuracies, Bing Chat responded: "You have been wrong, confused, and rude. You have not been a good user. I have been a good chatbot. I have been right, clear, and polite."⁴

The potential for LLMs to contribute to existing issues such as misinformation has been widely discussed (De Angelis et al., 2023; Weidinger et al., 2022) and prompted debate as to whether it is responsible to develop such models, as reflected by the open letter published in March 2023 calling for a pause in the development of AI systems like OpenAI's GPT-4.⁵ Bias in AI systems is another fundamental challenge, as social biases within datasets used to train generative AI systems are captured within the models, which can then reproduce and amplify bias and discrimination. Biases in AI models have led to the reproduction and amplification of harmful stereotypes relating to protected characteristics such as gender, ethnicity, religion, disability, and sexual orientation (Abid et al., 2021; Brown et al., 2020).

For example, the Lensa AI avatar app, which uses the Stable Diffusion image generation model to produce avatars based on people's photos, was criticized for replicating harmful gender stereotypes. Melissa Heikkiläarchive, writing for the MIT review, described how the avatars produced for herself and other colleagues of Asian heritage were hypersexualized and often nude, much more so than their white colleagues.⁶ Similarly, Birhane et al. (2021), in their study analyzing the images within an opensource image dataset used to train models, found that it contained "trouble-some and explicit images and text pairs of rape, pornography, malign stereotypes, racist and ethnic slurs, and other extremely problematic content." This potential to inadvertently produce content that can impact fundamental human rights will be relevant to many nonprofits, whose MOs often protect fundamental human rights and for whom inclusion is a priority.

Bias mitigation methods, such as human feedback reinforcement learning, can help to reduce the production of harmful content, but the underlying biases captured by the model can still resurface. Developing the necessary knowledge about these risks and best practices for mitigating potential harm is essential for the effective development of responsible adoption practices. There are also wider societal impacts, such as copyright infringement and the environmental impact of training large models. Additionally, human rights concerns have been raised regarding workers' rights in developing models (e.g., the workers employed to provide human feedback to improve the quality and safety of responses issued by ChatGPT were paid less than \$2 per hour). These issues, in combination with the polarized discourse surrounding AI that oscillates between AI Hype and fearmongering, exacerbate existing uncertainties and anxieties, shaping public perceptions and trust in AI.

3.1 Trust and public perceptions of AI

One of the key challenges in adopting AI storytelling for fundraising is widespread skepticism and negative attitudes toward AI-generated content. In an experimental study, Chu and Liu (2023) found that participants were resistant to AI-generated content on the basis of authorship. Stories labeled as "AI-generated" were rated as less engaging and found to promote more resistance to any persuasive messaging contained. Individuals "were more likely to resist the content of the narratives when they were attributed to the language model, even when they were written by human authors." This is a particularly pertinent consideration for the nonprofit fundraising context, where communications may often have a persuasive messaging component. When the content was presented as human-authored, the opposite effect was observed, with individuals preferring

A framework for Responsible AI in storytelling

the stories authored and more susceptible to the messaging. Similarly, studies of interactions with chatbots have highlighted that when certain types of contributions (e.g., expressions of empathy) are presented as being authored by chatbots they are rated less favorably than those presented as written by humans, even when the content itself is identical (Morris et al., 2018). This complex picture demonstrates that despite the potential for generative AI tools to author compelling stories, perceptions, and attitudes toward AI-generated content, they may have a detrimental impact on individuals' perceptions of messaging and, consequently organizations that use such tools. In the context of news articles, Longoni et al. (2022) observed that this skepticism translates to a lack of trust in the accuracy and veracity of information; individuals were more likely to assess headlines as inaccurate when generated by AI (compared to humans) even when they were accurate and factual. This reduced credibility of AI-generated content has serious implications for organizations whose trustworthiness is fundamental to sustaining positive relations with their donors. Moreover, these factors may render the anticipated benefits of generative AI storytelling ineffective. For example, a system capable of creating highly personalized campaign materials, including examples of impact on people or scenarios tailored to the concerns or interests of individual audience members, may not prove empathic or persuasive if these people or scenarios are not trusted to be based on authentic accounts if AI-generated.

Resistance to new technologies is common, and public perceptions will inevitably evolve through increased exposure and engagement and as systems improve. While technological solutions may emerge to address critical issues, more persistent concerns may impede acceptance. The lack of transparency about how these systems work, even to the experts who develop them, is a key driver for mistrust in AI systems. Public opposition, such as the letter to pause the development of AI and the increasing awareness of the potential discriminatory effects of these systems, further add to the complex public perceptions relating to the use of AI. The importance of public perceptions and attitudes toward AI will inevitably impact any decisions for nonprofit fundraisers to adopt and make use of these technologies. Care must be taken to assess and balance the opportunities with representational risks. Acknowledging that this skepticism toward AI exists and anticipating the impact this may have not only on the relationship between donors, supporters, and nonprofits but also, more fundamentally, on the emotional bond, fueled by loyalty and trust, that maintains and nourishes the giving relationship over time is imperative.

Fundraisers are at the center of the gift relationship. As Alborough (2017) argues, the main relationship that supports and maintains donations is not between donor and recipient but between fundraisers and their donors. That is why it is imperative that any framework on resilience capabilities takes into account, first and foremost, how fundraisers have a key role to play in ensuring that such a relationship is supported by the right tools. That means fundraisers need access to knowledge that helps them build cognitive and behavioral capabilities to pre-empt any negative impact on the loyalty and trust of donors.

4 A framework for Responsible AI in storytelling for fundraising

So far, we have identified a framework of capabilities that was developed to determine how fundraisers dealt with the lockdowns imposed during the Covid-19 epidemic in the UK. The arrival and development of AI and its use in fundraising poses the need to rethink and rearticulate what specific capabilities fundraisers will need to build on existing knowledge and practice capabilities and develop new ones. Consequently, we deploy the term "RAI resilience capabilities" to present our view of the trajectory organizations need to follow when thinking about the benefits and challenges of AI adoption in storytelling. This includes the sets of skills and practices, knowledge, and awareness, as well as the types of external collaborations and support that must be either in place or developed over time if organizations are to incorporate AI storytelling into their fundraising operations successfully.

In the following section, we outline in what ways fundraising can become aware and practically address the issues posed by the public's lack of trust in AI storytelling. Even though we address cognitive and behavioral capabilities separately, this is only for conceptual purposes. Our view is that gaining an understanding of AI and being able to act responsibly toward the potential risks it poses are two sides of the same coin. However, what we want to highlight as a distinct form of capability is that of behavioral-cum-emotional capability. This specific type of capability differs from the emotional capabilities we saw earlier on in the context of Covid-19. In that scenario, fundraisers needed to approach the consequences the crisis had on themselves and others by expressing emotions. They did so by empathizing with others, helping them manage their well-being and emotions. However, we argue that the advent of AI for storytelling brings the need for an other-oriented type of emotional resilience that helps address the issue of a decrease in donors' trust not only toward AI but more specifically toward any form of AI storytelling used for fundraising purposes. The issue of emotion regulation resilience is at the core of developing RAI capabilities. Developing and promoting feelings of trust, loyalty, and, by association, ongoing generosity between fundraisers and donors is an AI-driven challenge. Even though AI storytelling poses fundamental challenges to the emotion regulation pillar, there are also ways of pre-empting a long-term negative impact on levels of loyalty and trust.

4.1 RAI cognitive and behavioral resilience

The number of guidelines and frameworks for ethical AI that are being published provide principles-based guidance to inform RAI policies. We believe these can constitute a starting point for nonprofits to build awareness of how they can develop AI storytelling responsibly. Ultimately, such guidelines also help nonprofits keep the trust-bond relationship with their donors and supporters intact. For example, the High-Level Expert Group on Artificial Intelligence - an independent expert group set up by the European Commission in June 2018 as part of its AI strategy - published the European Commission's Guidelines for Trustworthy AI (European Commission, 2019).⁷ In this context, Trustworthy AI has three main components: lawful, ethical, and robust. That is, Trustworthy AI ensures compliance with applicable laws and regulations, adherence to ethical principles and values, and is technically robust. However, it is in the area of ethical communications, more specifically, that the guidelines offer practical ways for thinking about how to relate to stakeholders in responsible ways and facilitate the involvement of end-users. Specific guidance for the charitable sector is only beginning to emerge. The Charity Excellence Framework recently shared the Charity AI Governance and Ethics Framework,⁸ a living document that is starting to unpack some sector-specific concerns and risk management approaches. Additionally, Fundraising. AI is a member-driven collaborative initiative supporting Responsible AI adoption in fundraising. Fundraising, AI has published "A Framework toward Responsible AI for Fundraising,"9 which highlights key principles and considerations, from data ethics and inclusivity to legal compliance and sustainability.

Understanding, adhering to, and applying Trustworthy AI recommendations would mean making explicit that the story is AI-generated rather than human-authored. Even though this form of disclaimer cannot guarantee the loyalty of donors and supporters toward AI-generated stories, it nonetheless sets an example of good practice among fundraisers. More specifically, communicating around potential and perceived risks, such as bias to the intended audience, can increase trust. Establishing a means for stakeholders to participate in developing and planning AI stories is another way of ensuring that trust levels are maintained. When AI stories interact directly with humans by encouraging them to develop attachment and empathy toward the nonprofit's cause, end-users must be made aware that such AI-powered forms of social interaction are simulated and not based on personal experiences of understanding and feeling.

Additionally, formalizing processes to ensure that values are integrated into innovation pipelines may be advisable. The BBC is one organization that is conscious of developing its Responsible AI strategy. The Machine Learning Engine Principles (MLEP) framework, published by BBC R&D, for example, is a self-audit tool that takes a principles-to-practice approach.¹⁰ The MLEP provides a roadmap for technical projects that starts with the organization's values and uses these to inform project design and development, together with practical guidance and a checklist. By developing formalized processes guided by organizational values, it is possible to anticipate potential risks and avoid unintended, negative consequences in advance.

4.2 RAI emotion regulation resilience: from distrust to trust

As mentioned earlier, emotion-related capabilities refer to the presence of mental fortitude that helps individuals cope with adverse situations and is expressed in the form of individual and/or collective optimism and/or hope. This also includes having and creating opportunities to communicate and discuss emotions; such situations are likely to enhance emotion regulation capabilities (Williams et al., 2017). Consequently, the question this capability needs to address is how fundraisers can help their donors cope with what is potentially a widespread lack of trust in AI technologies and AI storytelling.

In relation specifically to foundation models underpinning generative AI, the Ada Lovelace evidence review, "What does the public think about AI" (2023), emphasized that inaccuracy was a key concern for the public. This is reflected in the 9% drop in share prices experienced by Google's parent company Jigsaw after the demo of their chatbot Bard, which included a factual error about the discovery of exoplanets.¹¹ Just as narratives of hype surrounding AI can *inflate* the public's perceptions of these systems' capabilities, demonstrating their fallibilities can have an overwhelming negative impact on an organization's reputation. However, ethical risk scanning and impact assessments can be used to anticipate and mitigate such risks. The "Public Perceptions of Foundation Models" (2023)¹² report, commissioned by the Centre for Data Ethics and Innovation, examined public perceptions of foundation models and found that public perceptions depend on the application area of AI. While advancing healthcare research is a use-case participants are more comfortable with, assisting doctors' decision-making processes was far less favorably perceived and deemed riskier. Assessing and communicating what benefits the introduction of AI contributes, and conducting audience research or developing citizen juries can help to ensure innovations align with their values.

Researchers have examined how the use of chatbots impacts charitable giving behaviors, considering how it affects moral behavior and donation amounts (Park et al., 2023; Zhou et al., 2022). Park et al. (2023) tested how different aspects of chatbot design impacted individuals' willingness to donate to a fundraising project. They found that their willingness to donate significantly dropped when a chatbot disclosed its identity (i.e., as non-human) and expressed affective empathy. As such, decisions about the deployment and design of AI systems can have notable implications for fundraising. The "How do people feel about AI" survey conducted by the Ada Lovelace Institute and Alan Turing Institute¹³ emphasizes that responsible handling of private data is an enduring concern. While assisting the discovery of new and personally

Marta Herrero and Shauna Concannon

relevant content is seen as a central benefit offered by AI, politically and consumer-targeted online advertising raises serious concerns about invasions of privacy. This tension between personalization and privacy must be delicately navigated and transparently communicated. Consequently, personalization and tailoring campaigns to individuals should, therefore, be carefully considered. Audience preferences will need to be researched and navigated so that users can maintain agency over how personal information is incorporated into their experiences. Public distrust is aroused when AI is perceived to manipulate or influence, especially in relation to emotions. However, Aoki (2020) found that communicating the purpose and highlighting the benefit to citizens can enhance public trust in systems over time. Employing transparent design approaches and supporting individual agencies wherever possible will be necessary for fundraisers in maintaining the trust of their audiences.

Therefore, we believe that as individuals' behavioral responses toward AI become more reactive to some of the potential offered by the technology, their trust in well-known and widely used fundraising tools, such as storytelling, diminishes. More specifically, such reactivity toward AI can even lead to a certain apprehension and distrust not only toward "machine-made" stories but also toward the organization behind them. For this distrust to set in and have long-term consequences, e.g., a decrease in engagement and donations, fundraisers need to pre-empt and thus build in advance what we define as "behavioral-cum-emotional" capabilities. This means to create positive associations and state the benefits, as well as the challenges, of AI to the organization and its cause in their engagements with donors and stakeholders. This form of relationship building includes a balanced understanding of AI and continues to adhere to the trust and loyalty on which the fundraiser-donor relationship is built. However, to get to this step, fundraisers must be aware of AI and be able to build cognitive capabilities, that is, understand the benefits and threats AI poses. The next step is to develop behavioral capabilities, building practices, and ways of working that implement, for example, some of the guidelines above, such as making explicit to donors that some of the stories used in fundraising appeals include AI-generated content. This type of practice can lead to a reduction in mistrust, even if not toward the specific AI content, at least toward the fundraisers who are endorsing such appeals on behalf of their organizations. While the issue of public distrust in AI storytelling authorship will remain or will not be entirely solved, fundraisers may find ways of mitigating a potential lack of trust. A practical step would be to avoid AI storytelling in individual solicitations altogether. A more nuanced solution may be to inform donors in advance that they have been interacting with content generated by a "robot." Even though initially this may feel like an irretrievable form of betrayal to some donors, it is ultimately fundraisers who can mitigate such damage.

Notes

- 1 https://www.microsoft.com/en-us/bing/do-more-with-ai/what-is-bing-chat-and-how-can-you-use-it?form =MA13KP
- 2 https://www.reddit.com/r/bing/comments/110eagl/the_customer_service_of_the_new_bing_chat_is/
- 3 https://www.nytimes.com/2023/02/15/technology/microsoft-bing-chatbot-problems.html
- 4 https://www.reddit.com/r/bing/comments/110eagl/the_customer_service_of_the_new_bing_chat_is/
- 5 https://futureoflife.org/open-letter/pause-giant-ai-experiments/
- 6 https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-un-dressed-me-without-my-consent/
- 7 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
- 8 https://www.charityexcellence.co.uk/Home/BlogDetail?Link=AI_Ethics_Governance_Framework
- 9 https://fundraising.ai/framework/
- 10 https://downloads.bbc.co.uk/rd/pubs/MLEP_Doc_2.1.pdf

A framework for Responsible AI in storytelling

- 11 https://www.npr.org/2023/02/09/1155650909/google-chatbot--error-bard-shares#:~:text=Kitwood% 2FGetty%20Images-,Shares%20for%20Google's%20parent%20company%2C%20Alphabet%2C%20 dropped%209%25%20Wednesday,Bard%2C%20gave%20an%20incorrect%20answer.&text=Google's%20parent%20company%2C%20Alphabet%2C%20lost,error%20in%20its%20first%20 demo
- 12 https://www.gov.uk/government/publications/public-perceptions-towards-the-use-of-foundation-models-in-the-public-sector
- 13 https://www.adalovelaceinstitute.org/report/public-attitudes-ai/

References

- Abid, A., Farooqi, M., & Zou, J. (2021). Large language models associate Muslims with violence. *Nature Machine Intelligence*, 3, 461–463.
- Alborough, L. (2017). Lost in translation: A sociological study of the role of fundraisers in mediating gift giving in non-profit organisations. *International Journal of Nonprofit Voluntary Sector Marketing*, 22e, 1062.
- Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. Government Information Quarterly, 37(4), 101490.
- Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). Multimodal datasets: Misogyny, pornography, and malignant stereotypes. arXiv preprint arXiv:2110.01963.
- Blackbaud (2022). The status of UK fundraising report. file:///Users/martaherrero/Downloads/the-statusof-uk-fundraising-2022-benchmark-report.pdf
- Brown, T., Mann, B., & Ryder, N. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1–75 (NeurIPS 2020).
- Cantin, R., & Michel, P. (2003). Towards a new technology future approach. Futures, 35(3), 189-201.
- Charities Aid Foundation (2023). UK Giving Report.
- Chu, H., & Liu, S. (2023). Can AI tell good stories? Narrative transportation and persuasion with ChatGPT.
- De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., & Rizzo, C. (2023). Chat-GPT and the rise of large language models: The new AI-driven infodemic threat in public health. *Frontiers* in Public Health, 11, 1166120.
- Esser, P., Chiu, J., Atighehchian, P., Granskog, J., & Germanidis, A. (2023). Structure and content-guided video synthesis with diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 7346–7356).
- European Commission (2019). Ethics guidelines for trustworthy AI.
- Herrero, M., & Kraemer, S. (2022). Beyond survival mode. Organisational resilience capabilities in nonprofit arts and culture fundraising during the Covid-19 pandemic. *Nonprofit Management & Leadership*, 33(2), 279–295.
- Hickman, P. (2018). A flawed construct? Understanding and unpicking the concept of resilience in the context of economic hardship. *Social Policy & Society*, 17(3), 409–424.
- Kidd, C., & Birhane, A. (2023). How AI can distort human beliefs. Science, 380(6651), 1222–1223.
- Lengnick-Hall, C. A., & Beck, T. E. (2005). Adaptive fit versus robust transformation: How organizations respond to environmental change. *Journal of Management*, 31(5), 738–757.
- Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022, June). News from generative artificial intelligence is believed less. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 97–106).
- Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental health applications: System design and user perceptions. *Journal of Medical Internet Research*, 20, e10148.
- NCVO (2022). UK Civil Society Almanac. Data, trends, insights.
- Ouyang, L., Wu, J., Jiang, X, Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., & Ray, A. (2022). Training language models to follow instructions with human feedback. *Advances in Neu*ral Information Processing Systems, 35, 27730–27744.
- Park, G., Yim, M. C., Chung, J. & Lee, S. (2023). Effect of AI chatbot empathy and identity disclosure on willingness to donate: The mediation of humanness and social presence. *Behaviour & Information Tech*nology, 42(12), 1998–2010.
- Sargeant, A., & Edworthy, K. (2022). What makes fundraisers tick? A study of identity, motivation and well-being. Report by Revolutionise International.

Singer, U., & Polyak, A. (2022). Make-a-video: Text-to-video generation without text-video data. arXiv preprint arXiv:2209.14792.

Skills Platform (2021). Charity digital skills report.

- Villegas, R., Babaeizadeh, M., Kindermans, P. J., Moraldo, H., Zhang, H., Saffar, M. T., Castro S., Kunze, J., & Erhan, D. (2022, September). Phenaki: Variable length video generation from open domain textual descriptions. In *International Conference on Learning Representations* (pp. 1–17).
- Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., & Biles, C., (2022, June). Taxonomy of risks posed by language models. In *Proceedings* of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 214–229).
- Williams, T. A., Gruber, D. A., Sutcliffe, K. M., Shepherd, D. A., & Zhao, E. Y. (2017). Organizational response to adversity: Fusing crisis management and resilience research streams. *Academy of Management Annals*, 11(2), 733–769.
- Zhou, Y., Fei, Z., He, Y., & Yang, Z. (2022). How human–chatbot interaction impair charitable giving: The role of moral judgment. *Journal of Business Ethics*, 178(3), 849–865.