

## IS AUTOMATED THERAPY DIGNIFIED?

Nancy Abigail Nuñez Hernández<sup>1</sup> , Björn Lundgren<sup>2,3</sup> \*

<sup>1</sup> Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Acatlán, State of Mexico, Mexico

<sup>2</sup> Friedrich-Alexander-Universität Erlangen-Nürnberg, Centre for Philosophy and AI Research, Erlangen, Germany

<sup>3</sup> Institute for Futures Studies, Stockholm, Sweden

\**bjorn.lundgren@fau.de*

**Abstract** With the advent of generative AI and large language models, there is great potential for AI-based therapy, also known as “automated therapy.” This raises substantive concerns that require careful ethical scrutiny. This paper contributes to this discussion by analyzing the ethical impact of automated therapy through the lens of its effects on human dignity. Although there are various methods of therapy, the focus here is on “human-based therapy,” which involves both the patient and medical personnel. The goal is to show that the concept of dignity can be used to analyze both the challenges and benefits of automated therapy. The paper ignores the contentious debate on the definition of dignity and instead focuses on what makes dignity fragile—that is, how it can be violated, diminished, frustrated, or extinguished. To that end, the paper draws on Suzy Killmister’s work to provide an overarching analysis of the effects of automated therapy on human dignity.

**Keywords** therapy; automated therapy; AI ethics; human dignity; informational dignity

### 1. Introduction

AI technology has shown tremendous results in medical areas such as cancer detection (Saba, 2020). With the advent of generative AI and large language models, there is great potential for AI-based therapy, also known as “automated therapy” (Burley et al., 2024). This raises substantive concerns that require careful ethical scrutiny. Here, we will contribute to part of that by analyzing the ethical impact of automated therapy through the lens of its effects on human dignity.<sup>1</sup>

Throughout this paper, the term “therapy” will refer to forms of treatment by a method other than surgery and medication to relieve emotional distress and mental health problems (Colman, 2015). There are various methods for therapy, and some involve sessions or meet-

---

1. There is some relevant literature related to the topics—for example, for discussion on carebots and dignity see, e.g., Pfeifer-Chomiczewska (2023), and for an ethical overview on automated therapy, see Burley et al. (2024). See also Freeman et al. (2018) and Fiske et al. (2019).

ings between a therapist and a patient. Here we focus on “human-based therapy,” which can be understood as therapy involving both a patient and medical personnel (of course, human-based therapy may also include exercises assigned by a human therapist for the patient to perform alone). The rapid development of AI technologies has seen applications in the early detection of mental health disorders, personalized treatment plans, and AI virtual therapists (Olawade et al., 2024). As such, “automated therapy” will be understood as situations in which an AI-based system has either (partially) replaced a human in human-based therapy, or has been added to complement human-based therapy, or to introduce new forms of possible therapy. While a detailed analysis could be done to compare different forms of human-based therapy with automated therapy, our goal here is somewhat different. We will aim for a more overarching analysis, and therefore, we will stay on a high level of abstraction.

Some philosophers have questioned the possibility of appealing to dignity in ethical analyses. Among them, Bernd Carsten Stahl, Doris Schroeder, and Rowena Rodrigues point to the lack of a standard definition of human dignity; Amanda Sharkey considers that the notion of dignity is too vague to be useful; and Ruth Macklin even claims that dignity “is a useless concept [...] [that] can be eliminated [from ethics debates] without any loss of content” (Stahl et al., 2023; Sharkey, 2019; Macklin, 2003, p. 1420). Nevertheless, human dignity is a persistent concern in medical ethics in general (Dresser, 2008; Pellegrino, 2008), and there is no *prima facie* reason why this would not apply to medical AI as well. However, those who take dignity seriously in healthcare realize that “so far little is known about how AI decision making in healthcare impacts patients’ perceptions of dignified interpersonal treatment” (Formosa et al., 2022, p. 1).

Based on these challenges, our goal is not to solve the conceptual or definitional issues of dignity; instead, we will show that the concept of dignity can be used to analyze the ethical challenges and benefits of automated therapy. To do so, instead of focusing on what dignity is, we are going to focus on what makes dignity fragile—that is, how it can be violated, diminished, frustrated, or extinguished. To that end, we will rely on the work of Suzy Killmister (2020), who provides the ingredients we need for our present purpose. This approach is not only due to the complexity of understanding the concept of human dignity (which is out of the scope of this work), but most importantly, this approach, based on Killmister’s work, allows us to recover important dimensions of the concept of dignity that are pervasive in the philosophical discussions on human dignity.

We will present Killmister’s account of the fragility of dignity in the next section (§2), after which we will turn to analyze different ethical challenges and benefits of automated therapy (§3). Finally, we end our article with some concluding comments (§4). Our focus here will be on the effects of technology on patients’ dignity. We believe this is a sensible focus, as the technology will be used in contexts where medical personnel are absent. This does not imply that it isn’t worthwhile to also consider the effects on medical personnel, and we will do so briefly at the end of §3.

## 2. The personal, social, and status strands of dignity

The concept of human dignity is pervasive in many contemporary debates, though, as with many other philosophical concepts, there is no consensus on a precise definition. This is illustrated by the variety of accounts of dignity within the philosophical literature, in particular the varied applied analysis such as those within medicine and bioethics (for the former, see, e.g., Sensen, 2011; Waldron, 2012; Rosen, 2018; Zylberman, 2016; and Etinson, 2020; for the latter, see, e.g., Barclay, 2016; Jones, 2015; Kelkar et al., 2024; and Boudierhem, 2024). Moreover, as Remy Debes recounts, until the 1830s–1850s, the term ‘dignity’ con-

noted social status, which can be linked to rank, hierarchy, or some social status; the term ‘dignity’ was not used to connote some fundamental, unearned, equally shared *moral* status among humans as it does nowadays (Debes, 2017). However, both connotations remain important aspects of the concept of dignity and of many of its uses (as can be seen in the works of Jeremy Waldron (2012) and Adam Etinson (2020), among others), and the precise relation between them has complicated the philosophical task of defining dignity. As Colyn Bird claims, “the concept of dignity is itself neither simple nor well defined, and so when speakers invoke it in moral contexts, exactly what they refer to is often unclear” (Bird, 2013, p. 1). In light of such difficulties, we want to avoid defining dignity in this paper. We take it that applied analyses can be performed without defining a concept. That is, as long as there is sufficiently steady agreement on a relevant subset of the extensions of a concept, the agreement need not be complete.<sup>2</sup> However, without relying on a definition, we have opted for an account that recovers the intricate complexity of the meaning of dignity, as can be found in Suzy Killmister’s work. It does so, insofar as it incorporates different dimensions of dignity (personal dignity, social dignity, and status dignity) recognized, in one way or another, by other philosophers.

Her theory offers a multidimensional approach to the concept of human dignity; it was developed based on key uses the concept seems to have in human moral lives, such as the inherent worth of human life, elevated status, an admirable quality, and respectful treatment. Considering the different roles dignity plays in our lives, Killmister proposes that the concept of dignity is like a tapestry comprising distinct yet complementary strands. The strands of this tapestry are the *personal*, *social*, and *status* dimensions of human dignity. Each one of these dimensions—or strands—of dignity can be violated or harmed in different ways, but can also be fostered in other ways (Killmister, 2020).

The *personal dignity strand* requires us to hold ourselves accountable to—and to succeed in meeting—personal and subjective dignitarian norms that we endorse (e.g., moral or religious norms). Having personal dignity means being subject to certain kinds of norms and acting in accordance with them, which has a positive impact on our self-respect. While meeting these personal norms is ennobling, violating them is disgraceful, even if it’s involuntary, as when someone lacks the capacity to hold herself to the standards or norms she previously set for herself. Hence, personal dignity’s fragility is thus exposed when someone is forced to do something she considers disgraceful. To illustrate this strand of dignity, we can think of norms of etiquette or any other normative demands that a person imposes on herself because she takes them to be ennobling. Hence, *personal dignity* is violated when a person transgresses these norms or is obliged to transgress them. For example, *personal dignity* can be difficult to uphold in a nursing home if the caregiver fails to help elderly people maintain basic norms of decorum (Killmister, 2020).

The *social strand* of human dignity, instead, demands that we uphold the dignitarian norms or standards of a community. Meeting these norms earns us respect from other members of the community, while transgressing causes rejection and a decrease in a person’s status within the community. Having *social dignity* means being subject to the norms held by members of a particular community and acting in accordance with them, which makes us socially dignified and commands the respect of our peers. While individuals’ failure to have *social dignity* is the result of not being included in a community that holds us to the relevant kind of standard, a violation of social dignity occurs when an individual is made to do or be something that is taken to be debasing or degrading in the eyes of the community. Since

---

2. Following, e.g., Etinson (2020); Schacter (1983) we take it that for the purpose this paper, we can identify properties of dignity without a definition. We see similar ideas in debates on other concepts, for example Lundgren (Lundgren, 2026).

this strand of dignity is tightly linked to the community, *social dignity* can differ from one culture and context to another. An example of how this strand of dignity would be violated is if someone's pictures of them being drunk and doing shameful things are exposed to the community. In such a situation, this person's standing would be lowered in the eyes of some members of the community.

Finally, the third strand of dignity is *status dignity*, which comes from membership in a specific social group that calls for respectful treatment; the relevant standards for this strand of dignity are a particular community's standards for how members, in virtue of their status, are to be treated (Killmister, 2020). Having *status dignity* is to be a member of a category that entitles recognition and respect; failing to have *status dignity* is not falling within any category, membership of which commands respectful treatment. *Status dignity* is violated by failing to treat a person in accordance with the recognition respect to which she is due. To illustrate this strand of dignity, we can think about a dignity that comes with nobility titles; for example, being the queen or king of a country commands respect, and such titles call on others to treat the queen or king in ways appropriate to the status that comes with the titles. If others fail to treat the queen or king in accordance with the appropriate form of respect, it follows that their status and *dignity* would be damaged.

In the next section, we will consider how these three strands of dignity relate to ways in which automated therapy changes the situation for patients and doctors. In doing so, we will not only apply but partially explicate some of the notions used by Killmister.

### 3. Applied ethical analysis

In this section, we use Killmister's dignity framework to analyze three categories of ethical considerations for patients—namely, privacy and data protection, access, and social aspects—after which we briefly address some considerations for doctors and psychologists. In each subsection, we will first give a brief overview of the challenge, using the broad conceptual toolbox available in ethics, after which we turn to address the three strands of dignity. Access and social aspects are discussed jointly, as their effects on dignity are strongly related.

Perhaps the most central aspect of all is omitted from the list above: broadly speaking, medical quality (i.e., the quality of mental healthcare). Although this concern is partially addressed under some of the other categories, it is essentially an empirical question that requires a different form of analysis than what we provide here. Needless to say, a non-functional medical treatment is not dignified, and this clearly holds for mental healthcare as well.

#### 3.1 Privacy and data protection

In this subsection, we start by addressing a category of effects that roughly falls under the heading of privacy and, more broadly and correctly, various reasons for data protection. Next, based on the brief analysis, we turn to address how such effects can be analyzed in Killmister's framework.

Starting with a brief analysis of the privacy and data protection effects, it is important to note that all information or data related concerns of therapy—whether conventional or automated—need not be considered privacy concerns; issues may also be related to concerns for individuals' *anonymity* or *autonomy* (and it remains contested precisely how these concepts relate to the concept of privacy). For example, in relation to anonymity, even *prima facie* insignificant data may be used to detrimentally affect patients' ability to be anonymous, which in turn can be abused for manipulative purposes (Lundgren, 2020b). In relation to

autonomy, there is a long-standing idea amongst some privacy scholars that privacy matters for autonomy.<sup>3</sup>

Nevertheless, medical information and medical interventions are, roughly speaking, fundamental privacy concerns and stand out as examples of what is privacy-sensitive in most, if not all, cultures and contexts. Amongst all medical care given, the information related to therapy is amongst the most privacy-sensitive, competing—perhaps only—with abortion, gynecology and urology.

Effects on users' privacy can be both positive and negative. Starting with the positive, it is not odd to think that *machines'* access or control of data has no privacy implications—that is, one may think that privacy is affected only by human access—then automated therapy has clear privacy benefits (cf. Macnish, 2018; Lundgren, 2020a). However, it is not unreasonable to think that a right to privacy also protects against risks to individuals' privacy (see, e.g., Lundgren, 2021b, 2021a), which raises concerns about how data is protected. Sadly, a study of 578 mental health apps revealed that 44% “shared personal health information with third parties” (Camacho et al., 2022), which illustrates not only that the security practices are improper but also a clear violation of the right to privacy.<sup>4</sup>

The dignity considerations for data protection are clear. Indeed, some think that privacy is grounded in dignity (see, e.g., Bloustein, 1964; Floridi, 2016). Although we fail to see how this presumption can be maintained—as there clearly are privacy concerns that go beyond dignity (i.e., we may want to limit access or control to information that has no dignity implications vice versa)—what matters for the present purpose is whether privacy concerns can also be about dignity. It would not be odd to speak of “informational dignity,” a concept that can be explained by the similar function of the concept of “informational privacy,” a part of privacy that focuses on the privacy of information, which—depending on one's theoretical foundations—is affected either through access or control of privacy-sensitive information by others. The concept of “informational dignity” is meant to function similarly to the concept of “information privacy,” but with a focus on dignity-sensitive information (i.e., information that affects one's dignity if accessed or controlled by others).<sup>5</sup> Based on this, one can easily see how an ability to be anonymous would be essential to safeguard an individual's informational dignity—and thereby dignity simpliciter—as dignity may require that certain acts be done in anonymity (cf. Matthews, 2010).

Based on the brief analysis above, let us now consider the effects on the three strands, starting with personal dignity. While the personal norms that make up our personal dignity are individual and vary from person to person, it is clear that machines may benefit some people, *vis-a-vis* their personal dignity, because sharing sensitive mental information with a human may violate personal dignity norms. (Here it is important to note that therapy, on the other hand, may require sharing such sensitive information—under a relation built on trust—but we will return to that when we address the social aspects of automated therapy.)

Thus, while sharing dignity-sensitive information with humans may be demeaning for some, sharing it only with machines can avoid any such demeaning information sharing, since machines do not (at least not at this time) have the social function that can impinge on our informational dignity (roughly speaking). While this may have negative implications for therapy (as noted above and as we will return to in section 3.2), the effects can also be

3. See, e.g., Westin (1967) and Fried (1968, 1970) for its link through social relations, an idea that is often associated with Rachels (1975).

4. One may think that if consent has been properly applied then this is not a problem, but it is questionable whether consent can be properly given for what is likely a purely commercial third-party sharing (Lundgren, 2025c).

5. As implied above, while the concept of privacy-sensitive and dignity-sensitive information overlaps, the overlap is merely partial.

highly positive, as it may allow individuals who would not engage in human therapy to feel more compelled to use automated therapy. Indeed, patients with low social trust may prefer to share data anonymously with an AI system.

However, these benefits require a situation in which the patient can trust that the dignity-sensitive information is protected. If patients have that trust, they may engage with automated therapy when they would not engage with human therapy. That trust may ensure that information sharing does not impinge on their personal dignity.

The situation for the social strand of dignity is similar but different in an important way.<sup>6</sup> The sharing of sensitive mental health information can clearly affect individuals' social dignity, as it sadly, in many societies, violates social norms that underpin our social dignity. Similarly, Killmister claims that sharing naked pictures of a cognitively impaired person violates their social dignity, because even if they do not care about it because of her cognitive condition, such pictures exhibit her transgressing the norms of her community and humiliate her (Killmister, 2020). This is not to say that such reactions are ethically justified—on the contrary, they are often wrong. However, notwithstanding their rightness, such reactions do affect an individual's social dignity negatively simply because they make it impossible for the individual to live up to the social norms required for maintaining one's social dignity. Protecting our social dignity against this risk does not require *trust* in data protection policies (as was sufficient and necessary for personal dignity), but it is necessary to have good policies in place to ensure that patient data is protected and that patient confidentiality is preserved. While automated therapy—at least in theory—can imply better data protection than traditional therapy (because humans are not involved), it is practically difficult to ensure security with private devices (Lundgren & Möller, 2019), and the standard for data protection—from the apps available—is highly problematic as we already exemplified; or, simply put, the current market of applications have an undignified handling of information simpliciter.

Finally, we turn to status dignity: the situation for status dignity is similar to social dignity, as it implies a requirement to protect your patient data. However, the reasons are somewhat different. It is in virtue of your status as a patient that the sharing—and imperfect protection of—your sensitive data is disrespectful, simply because it mistreats your status as a patient and the norms that ought to be upheld related to it, such as confidentiality.

In conclusion, automated therapy and related AI technologies have the potential to provide a more dignified handling of medical data, but this requires trustworthy, robust safeguards for user data—far from the current industrial norm—and policies that enable and require technology providers to address patients' legitimate data protection concerns. Of course, such considerations are worth little if the automated therapy is not otherwise dignified.

### 3.2 Access and social change

In this subsection, we start by briefly addressing how automated therapy changes patients' access to therapy and how it affects the social aspects of therapy. Next, based on the brief analysis, we turn to address how such effects can be analyzed in Killmister's framework.

Globally speaking, healthcare is a scarce resource; we have too little of it, and this is particularly true for mental healthcare.<sup>7</sup> Moreover, even in welfare states, access to healthcare can be cumbersome or include unreasonable waiting times, which can also be detri-

---

6. One may be concerned that the similarity of the applied analysis of Killmister's strands of dignity shows a limit to using her account in the given case. We do not think so. Instead, the fact that there is overlap between concerns for different strands of dignity showcases the pervasiveness of some of the impacts of automated therapy on dignity—that is, some uses of such technologies affect two or all three of the strands.

mental for mental care (Reichert & Jacobs, 2018). In this context, automated therapy can help reduce waiting times and increase access.<sup>8</sup> Moreover, as previously noted regarding care robots, AI systems “do not get tired, can provide care on a 24 h basis, have no worse or better days, are objective in their analysis, are not stressed or frustrated” (Pfeifer-Chomiczewska 2023, p. 796).<sup>9</sup> However, while stable functionality is desirable in any case, it may also hinder appropriate social training. That is, 24/7 access may be detrimental for patients who need to learn boundaries (Burley et al., 2024). One might think that this is a non-issue, since an AI-system can easily be programmed to adapt to whatever pattern is needed for patient-care, but whether this is realistic depends on whether these systems are a healthcare tool assigned by a doctor, or a system available on a more open market (perhaps under the broad category of lifestyle apps). The latter offers the most significant access benefit—because it can complement healthcare where it is the scarcest—but it also faces a dilemma: who would buy or use a software that is limited in use if a less limited device is available? Keep in mind that individuals who suffer from social problems might not realize precisely what they need; indeed, it would be odd to think that all people who need to practice social boundaries are also self-aware of this and have the willpower to engage with applications that frustrate their wishes (arguably, that seems less probable than the reverse).

The accessibility to these tools can also benefit people who might not need mental care, but suffer mentally from loneliness or struggle with human social connections. These types of problems are extremely common in modern society, especially among older people and across significant global disparities (Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults et al., 2020; Surkalim et al., 2022). While accessibility can create both benefits and risks, depending on individual needs, there is a further social aspect related to this: automated therapy changes the social aspects of therapy. As we discussed in the previous section, this may be both positive and negative. On the one hand, for some, it may be simpler to seek care if they can talk with a machine. As research has shown, stigma, concern for one’s reputation, and specific ideas about honor can contribute to individuals avoiding seeking mental healthcare (Foster et al., 2021). Moreover, for some patients, automated therapy might also be more helpful. For example, patients who, because of their peculiarities, test the patience of human healthcare providers may benefit from engaging with a system that cannot lose patience. That is, automated therapy might be precisely what some patients need. On the other hand, automated therapy lacks a social dimension, which may be detrimental for various reasons. Some forms of therapy might depend on developing a social bond with the therapist; indeed, some research suggests that “the therapeutic relationship predicts outcome across various psychiatric settings” (Priebe & McCabe, 2008), which current machines just cannot do. Their engagement with us is, by their very nature, fake; their “relations” with us are non-reciprocal (of course, this is not to say that healthcare professionals do not fake their engagement with us). This relates to another aspect, the idea that the care element of healthcare might be very important (see, e.g., Alvarado & Morar, 2024). However, “some research suggests that the appearance of caring might provide many of the benefits of actual caring; these results have not yet been tested over time, because the technology is too new” (Burley et al., 2024, p. 16).

---

7. See, e.g., Sexanar et al. (2007) for relevant, even if not very recent, discussion.

8. Of course, access to digital systems may not be easily accessible for everyone, which may have detrimental questions related to the fairness of unequal access. Such concerns arise when considering unequal access to AI, the Internet or other digital technologies, which might be due to economic or political circumstances, to a generational gap in the case of elderly people, digital literacy, etcetera.

9. See also Lundgren (2021c) for a similar argument about autonomous vehicles but without the problematic note of “objectivity” which ignores the important issue of algorithmic bias.

To address these concerns fully, we need to consider them within a broader ethical framework, which is precisely what we now turn to as we address the above issues using the three strands of dignity, starting with personal dignity. Personal dignity is affected in various ways by automated therapy. Improved access to mental healthcare can be essential for individuals' ability to maintain the dignity norms they have set for themselves. Simply put, waiting times for people suffering from mental concerns are disrespectful. However, in some special circumstances, the limitless accessibility of some systems may make it more difficult for some individuals to meet norms that require them, for example, to practice social boundaries. Relatedly, automated therapy can be precisely what one individual needs to maintain personal norms (e.g., because they need to practice social behavior in a structured way); or it may be contrary to what they need (e.g., because they need genuine empathy and social relations to succeed with therapy).

The effects on social dignity relate to many similar issues, as mental health may be essential for an individual's capacity to live up to the social norms set by the community. This means that the considerations for social dignity will be roughly similar to those for personal dignity, as the effects of this strand of dignity are mostly instrumental.

As we have applied—and perhaps even explicated—the concept of status dignity in the previous section, in terms of the dignity you deserve in virtue of your status as a patient, we can now see its impact for the consideration of how automated therapy changes access and the social aspects of therapy. Similarly, as in the above argument, we can conclude that, on the one hand, access is important for our dignity. Our status as patients—or, simply put, as humans (i.e., in reference to the foundational concept of human dignity)—demands access to healthcare in general. On the other hand, if that access is achieved by engaging with machines rather than humans, this may fundamentally imply an undignified approach (van de Rijt et al., 2026). That is, one may question whether the social response we are all owed—as patients and/or humans—is satisfied by a bot that cannot really empathize.<sup>10</sup> That is, being confronted with an AI system in a time of desperate need may be considered disrespectful. How do we balance these dual considerations? The implication is, *prima facie*, simple: Human contact is sometimes essential, so replacing humans with machines may be detrimental to patient dignity, as it is disrespectful. But that does not imply that *complementing humans with machines* would be detrimental to patient dignity.

However, this simple conclusion is complicated by the fact that *fake* human care is arguably more disrespectful than the absence of genuine care from an AI system. Consider a therapist making fun of their clients outside of the therapy session. In such a case, it is arguably fair to say that human fake care may be less dignified for patients than the fake care of a machine, because human fake care implies not only a lack of respectful treatment but a disrespectful treatment (by a moral agent). In contrast, the lack of empathy, care, and other emotional responses from a machine implies only a lack of respectful treatment. Of course, this is not an argument for automated therapy, but rather an argument that the evaluation thereof must be done in relation to a genuine rather than an idealized baseline and to the question of whether we can achieve improvements in human-based therapy.

In conclusion, automated therapy has the potential to improve our dignity by increasing access to mental healthcare that is essential for leading a dignified life. Yet, as we suggest, the value of that responsiveness is somewhat context-dependent and may not benefit all patients. Relatedly, the lack of genuine social relations between a machine and a patient may be both beneficial (e.g., making it more accessible for some—and thereby contributing to what we

---

10. We agree with Bender et al. (2021) that LLMs are mere stochastic parrots and we see no reason to think that other current AI models are able to empathize.

may call “mental dignity”) or it may lack the agential aspects that are essential for proper, respectful treatment. As we noted, this problem is not limited to AI systems; human disrespect is even worse. That is, no emotional response may be less harmful than a disrespectful emotional response. In this context, it may be important to note that what matters most is whether automated therapy contributes to patients’ ability to lead a dignified life. Just as it has been argued that care bots can provide patients with dignity—because the technology enables life choices that otherwise aren’t available—automated therapy might provide support for patients when human therapy is lacking.

Based on what we have said, there is also a further complicating factor: dependence on these systems can create a risk of an undignified life if one’s life becomes dependent on fake social relationships with machines. Simply put, extended reliance on fake social relationships may be undignified if it replaces or precludes genuine ones. There is a lot more to be said about the role of social machines and social AI, but that question goes beyond automated therapy and will therefore not be addressed here.<sup>11</sup>

### 3.3 The dignity of therapists

In this subsection, we will examine how automated therapy might affect therapists’ dignity. Although one of the most obvious effects may be due to replacement on the job market (i.e., AI systems performing human work tasks), we will not address that as such. That belongs to a much larger debate on the dignity of work (see, e.g., Yeoman, 2014). Unlike the other subsection, we will address this brief analysis directly from a dignity standpoint; we will not go through the other topics that may affect therapists as well. Instead, we will aim to introduce a few new issues (even if only very briefly).

If automated therapy is replacing rather than complementing therapists or if it is proving a better treatment than human-based therapy, that might affect therapists’ dignity in the following way. On the one hand, competition from automated therapy might make it difficult for the therapist to live up to their personal norms. For example, providing effective therapy may conflict with being a good therapist (if an AI system is performing better). It might also make them set norms for themselves that they would not consider dignified (e.g., by trying to provide competitive benefits on the market in a way that isn’t compatible with their other ethical norms). Furthermore, one may argue that all the ways we can update our personal norms are not necessarily dignified. On the other hand, it might turn out that automated therapy complements human therapy in ways that benefit human therapists. For example, they may take over some tasks that enable therapists to help more patients, which may be very fulfilling *vis-a-vis* their personal norms. Similarly, automated therapy can also affect the therapist’s ability to maintain social dignity in similar ways. Basically, the above arguments can be considered in lieu of social norms rather than personal ones.

Lastly, it could affect their status and dignity. On the one hand, as a therapist, one can be said to have some status, especially granted a successful skillset, which can be affected if it turns out that mere machines can perform their work. The status gained by their occupation might therefore not be as prestigious as before. On the other hand, if automated therapy improves therapy when used with or by human therapists, their status may increase—especially if there are types of therapy that can only be performed by humans. We can also note that even though AI systems are much stronger than humans in the game of chess, the popularity of the game is growing, and the valuation of the chess players seems to increase with it (Lundgren, 2024).

---

11. See, e.g., Lundgren (2025b) for a relevant discussion on the ethical challenge of social AI systems.

As such, the effects on the dignity of therapists will strongly depend on how the technology develops, which would provide a *pro tanto* reason in favour of a more collaborative method. However, the interesting question is not the dignity effect of therapists in isolation, but how we should value their dignity against that of the patients. Of course, some of the arguments we have presented for patients also favor a similar approach of what might be optimal for therapists. Nevertheless, in case of a conflict, we think that *pro tanto* reasons for healthcare providers are weak in comparison to the patients. Healthcare is, after all, for patients, not for providers, so their moral standing outweighs that of the providers. But this is a much more complicated question that we cannot deal with in full here.

#### 4. Concluding comments

In this paper, we have provided an analysis of automated therapy from a dignity perspective, which, in turn, has allowed us to provide some guidance for the ethical use of automated therapy. This is not all that ought to be said about automated therapy. For example, our discussion relating to medical personnel is limited. Nevertheless, some patient-related arguments also imply reasons against replacement, which illustrates that a dignified patient care may cohere with a dignified working life—something that is hardly surprising, even if more needs to be said on the issue.

Our main conclusion is that while automated therapy may provide benefits in relation to all aspects we considered, there are also many contextually sensitive considerations, as we have indicated, as well as risks and problems. Indeed, one may worry that a lot of the considerations relative to the dignity strands are of the form “on the one hand...on the other hand”, which may seem to set aside an all-things-considered evaluation. However, it actually provides guidance for design and use because the arguments spell out different choices that often imply either a benefit or a cost.

Lastly, when analyzing the ethics of the future of automated therapy, amongst all of the core values that need to be considered, we hope our brief discussion illustrates that dignity should be one of them.

#### Acknowledgement

We gratefully acknowledge the comments we received when presenting this at two workshops on AI and Human Dignity, which we organized in Germany and Mexico. Our gratitude is also extended to the helpful comments we received from two anonymous reviewers for this special issue. Finally, Nancy Abigail Nuñez Hernández’s contribution is part of the research project PAIDI 007/24 “La dignidad humana ante los retos de los usos de la Inteligencia Artificial” and the research project PAPIIT IA400426 “Sobre la posibilidad de automatizar las demostraciones matemáticas: retos en el uso de la Inteligencia Artificial en las demostraciones o pruebas matemáticas.” We appreciate the help of Marc Jiménez Rolland with adapting the references to the journal standard.

#### Conflict of interest

Lundgren is a member of the Ethics Advisory Board for the Estonian Centre of Excellence in AI.

## References

- Alvarado, R., & Morar, N. (2024). The incommensurability of caring: ML, clinical decision-making, and human reasoning in healthcare. *The American Journal of Bioethics*, 24(9), 113–115. <https://doi.org/10.1080/15265161.2024.2377146>
- Barclay, L. (2016). In sickness and in dignity: A philosophical account of the meaning of dignity in health care. *International Journal of Nursing Studies*, 61, 136–141. <https://doi.org/10.1016/j.ijnurstu.2016.06.010>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bird, C. (2013). Dignity as a moral concept. *Social Philosophy and Policy*, 30(1–2), 150–176. <https://doi.org/10.1017/S0265052513000071>
- Bloustein, E. J. (1964). Privacy as an aspect of human dignity: An answer to Dean Prosser. *NYU Law Review*, 39, 962–1007.
- Boudierhem, R. (2024). Shaping the future of AI in healthcare through ethics and governance. *Humanities and Social Sciences Communications*, 11(1), 416. <https://doi.org/10.1057/s41599-024-02894-w>
- Burley, J., Hughes, J., Stubbs, A., & Eisikovits, N. (2024). *The Ethics of Automating Therapy*. Institute for Ethics and Emerging Technologies. <https://philpapers.org/archive/BURTEO-71.pdf>
- Camacho, E., Cohen, A., & Torous, J. (2022). Assessment of mental health services available through smartphone apps. *JAMA Network Open*, 5(12), e2248784. <https://doi.org/10.1001/jamanetworkopen.2022.48784>
- Colman, A. M. (2015). *A Dictionary of Psychology* (4th ed.). Oxford University Press.
- Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults, Board on Health Sciences Policy, Board on Behavioral, Cognitive, and Sensory Sciences, Health and Medicine Division, Division of Behavioral and Social Sciences and Education, & National Academies of Sciences, Engineering, and Medicine. (2020). *Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System*. National Academies Press. <https://doi.org/10.17226/25663>
- Debes, R. (Ed.). (2017). *Dignity. A History* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199385997.001.0001>
- Dresser, R. (2008). Human dignity and the seriously ill patient. In Schulman, Adam (Ed.), *Human Dignity and Bioethics: Essays Commissioned by the President's Council on Bioethics*. Government Printing Office.
- Etinson, A. (2020). What's so special about human dignity? *Philosophy & Public Affairs*, 48(4), 353–381. <https://doi.org/10.1111/papa.12175>
- Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, 21(5), e13216. <https://doi.org/10.2196/13216>
- Floridi, L. (2016). On human dignity as a foundation for the right to privacy. *Philosophy & Technology*, 29(4), 307–312. <https://doi.org/10.1007/s13347-016-0220-8>

- Formosa, P., Rogers, W., Griep, Y., Bankins, S., & Richards, D. (2022). Medical AI and human dignity: Contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts. *Computers in Human Behavior*, 133, 107296. <https://doi.org/10.1016/j.chb.2022.107296>
- Foster, S., Carvallo, M., Lee, J., & Bernier, I. (2021). Honor and seeking mental health services: The roles of stigma and reputation concerns. *Journal of Cross-Cultural Psychology*, 52(2), 178–183. <https://doi.org/10.1177/0022022120982070>
- Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., Denne, M., Brown, P., Slater, M., & Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: A single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, 5(8), 625–632. [https://doi.org/10.1016/S2215-0366\(18\)30226-8](https://doi.org/10.1016/S2215-0366(18)30226-8)
- Fried, C. (1968). Privacy. *The Yale Law Journal*, 77(3), 475–493. <https://doi.org/10.2307/794941>
- Fried, C. (1970). *An Anatomy of Values: Problems of Personal and Social Choice*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674332485>
- Jones, D. A. (2015). Human dignity in healthcare: a virtue ethics approach. *The New Bioethics*, 21(1), 87–97. <https://doi.org/10.1179/2050287715Z.00000000059>
- Kelkar, A. H., Hantel, A., Koranteng, E., Cutler, C. S., Hammer, M. J., & Abel, G. A. (2024). Digital health to patient-facing artificial intelligence: Ethical implications and threats to dignity for patients with cancer. *JCO Oncology Practice*, 20(3), 314–317. <https://doi.org/10.1200/OP.23.00412>
- Killmister, S. (2020). *Contours of Dignity* (1st ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780198844365.001.0001>
- Lundgren, B. (2020a). A dilemma for privacy as control. *The Journal of Ethics*, 24(2), 165–175. <https://doi.org/10.1007/s10892-019-09316-z>
- Lundgren, B. (2020b). Beyond the concept of anonymity: What is really at stake? In K. Macnish & J. Galliot (Eds.), *Big Data and Democracy* (pp. 201–216). Edinburgh University Press. <https://doi.org/10.3366/edinburgh/9781474463522.003.0014>
- Lundgren, B. (2021a). Confusion and the role of intuitions in the debate on the conception of the right to privacy. *Res Publica*, 27(4), 669–674. <https://doi.org/10.1007/s11158-020-09495-9>
- Lundgren, B. (2021b). How we can make sense of control-based intuitions for limited access-conceptions of the right to privacy. *Journal of Ethics and Social Philosophy*, 20(3). <https://doi.org/10.26556/jesp.v20i3.1438>
- Lundgren, B. (2021c). Safety requirements vs. crashing ethically: What matters most for policies on autonomous vehicles. *AI & Society*, 36(2), 405–415. <https://doi.org/10.1007/s00146-020-00964-6>
- Lundgren, B. (2024). There is no scarcity problem. *Philosophy & Technology*, 37(4), 130. <https://doi.org/10.1007/s13347-024-00815-y>
- Lundgren, B. (2025a). How social should AI be? *Erkenntnis*. <https://doi.org/10.1007/s10670-025-01005-8>
- Lundgren, B. (2025b). On the limits of the data economy: The case of autonomous vehicles. *Science and Engineering Ethics*, 31(4), 16. <https://doi.org/10.1007/s11948-025-00540-5>
- Lundgren, B. (2026). Can deepfakes violate an individual’s moral right to privacy?. *Ethical Theory and Moral Practice*, 29, 125–139. <https://doi.org/10.1007/s10677-025-10514-y>
- Lundgren, B., & Möller, N. (2019). Defining information security. *Science and Engineering Ethics*, 25(2), 419–441. <https://doi.org/10.1007/s11948-017-9992-1>

- Macklin, R. (2003). Dignity is a useless concept. *BMJ*, 327(7429), 1419–1420. <https://doi.org/10.1136/bmj.327.7429.1419>
- Macnish, K. (2018). Government surveillance and why defining privacy matters in a post-Snowden world. *Journal of Applied Philosophy*, 35(2), 417–432. <https://doi.org/10.1111/japp.12219>
- Matthews, S. (2010). Anonymity and the social self. *American Philosophical Quarterly*, 47(4), 351–363.
- Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with artificial intelligence: Current trends and future prospects. *Journal of Medicine, Surgery, and Public Health*, 3, 100099. <https://doi.org/10.1016/j.glmedi.2024.100099>
- Pellegrino, E. D. (2008). The lived experience of human dignity. In A. Schulman (Ed.), *Human Dignity and Bioethics: Essays Commissioned by the President's Council on Bioethics*. Government Printing Office.
- Pfeifer-Chomiczewska, K. (2023). Intelligent service robots for elderly or disabled people and human dignity: Legal point of view. *AI & Society*, 38(2), 789–800. <https://doi.org/10.1007/s00146-022-01477-0>
- Priebe, S., & McCabe, R. (2008). Therapeutic relationships in psychiatry: The basis of therapy or therapy in itself? *International Review of Psychiatry*, 20(6), 521–526. <https://doi.org/10.1080/09540260802565257>
- Rachels, J. (1975). Why privacy is important. *Philosophy & Public Affairs*, 4(4), 323–333.
- Reichert, A., & Jacobs, R. (2018). The impact of waiting time on patient outcomes: Evidence from early intervention in psychosis services in England. *Health Economics*, 27(11), 1772–1787. <https://doi.org/10.1002/hec.3800>
- Rosen, M. (2018). *Dignity: its history and meaning* (First Harvard University Press paperback edition). Harvard University Press.
- Saba, T. (2020). Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. *Journal of Infection and Public Health*, 13(9), 1274–1289. <https://doi.org/10.1016/j.jiph.2020.06.033>
- Saxena, S., Thornicroft, G., Knapp, M., & Whiteford, H. (2007). Resources for mental health: Scarcity, inequity, and inefficiency. *The Lancet*, 370(9590), 878–889. [https://doi.org/10.1016/S0140-6736\(07\)61239-2](https://doi.org/10.1016/S0140-6736(07)61239-2)
- Schachter, O. (1983). Human dignity as a normative concept. *American Journal of International Law*, 77(4), 848–854. <https://doi.org/10.2307/2202536>
- Sensen, O. (2011). Human dignity in historical perspective: The contemporary and traditional paradigms. *European Journal of Political Theory*, 10(1), 71–91. <https://doi.org/10.1177/1474885110386006>
- Sharkey, A. (2019). Autonomous weapons systems, killer robots and human dignity. *Ethics and Information Technology*, 21(2), 75–87. <https://doi.org/10.1007/s10676-018-9494-0>
- Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). Dignity. In B. C. Stahl, D. Schroeder, & R. Rodrigues, *Ethics of Artificial Intelligence* (pp. 79–93). Springer International Publishing. [https://doi.org/10.1007/978-3-031-17040-9\\_7](https://doi.org/10.1007/978-3-031-17040-9_7)
- Surkalim, D. L., Luo, M., Eres, R., Gebel, K., Van Buskirk, J., Bauman, A., & Ding, D. (2022). The prevalence of loneliness across 113 countries: Systematic review and meta-analysis. *BMJ*, 376, e067068. <https://doi.org/10.1136/bmj-2021-067068>
- van de Rijt, J.-W., Coelho Mollo, D., & Vaassen, B. (2026). AI mimicry and human dignity: Chatbot use as a violation of self-respect. *Journal of Applied Philosophy*, 43(1), 95–111. <https://doi.org/10.1111/japp.70037>
- Waldron, J. (2012). *Dignity, Rank, and Rights* (M. Dan-Cohen, Ed.). Oxford University Press.

Westin, A. F. (1967). *Privacy and Freedom*. Atheneum.

Yeoman, R. (2014). Conceptualising meaningful work as a fundamental human need.

*Journal of Business Ethics*, 125(2), 235–251. <https://doi.org/10.1007/s10551-013-1894-9>

Zylberman, A. (2016). Human Dignity. *Philosophy Compass*, 11(4), 201–210. <https://doi.org/10.1111/phc3.12317>