University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

9-22-2020

Surrogates and Artificial Intelligence: Why AI Trumps Family

Ryan Hubbard

Jake Greenblum
The University of Texas Rio Grande Valley

Follow this and additional works at: https://scholarworks.utrgv.edu/som_pub

Part of the Bioethics and Medical Ethics Commons

Recommended Citation

Surrogates and Artificial Intelligence: Why AI Trumps Family
PENULTIMATE DRAFT

Ryan Hubbard, PhD

Jake Greenblum, PhD

Abstract
The increasing accuracy of algorithms to predict values and preferences raises the possibility that artificial intelligence technology will be able to serve as a surrogate decision-maker for incapacitated patients. Following Camillo Lamanna and Lauren Byrne, we call this technology the *autonomy algorithm* (AA). Such an algorithm would mine medical research, health records, and social media data to predict patient treatment preferences. The possibility of developing the AA raises the ethical question of whether the AA or a relative ought to serve as surrogate decision-maker in cases where the patient has not issued a medical power of attorney. We argue that in such cases, and against the standard practice of vesting familial surrogates with decision making authority, the AA should have sole decision-making authority. This is because the AA will likely be better at predicting what treatment option the patient would have chosen. It would also be better at avoiding bias and, therefore, choosing in a more patient-centered manner. Furthermore, we argue that these considerations override any moral weight of the patient's special relationship with their relatives.

Introduction
Experts in artificial intelligence (AI) claim that the exponential growth and development of AI "promises to change the landscape of medical practice" (Yu et al. 2018) This change will come with its own set of ethical questions, so it is crucial for bioethicists to anticipate the future integration of AI into medical practices. The increasing accuracy of algorithms to predict personal values and preferences raises the possibility that artificial intelligence technology could, in the near future, be developed to serve as a possible surrogate decision-maker for incapacitated patients. Following Camillo Lamanna and Lauren Byrne, we call this technology the *autonomy algorithm* (AA) (2018). Such an algorithm would mine medical research, electronic health records (EHR), sociodemographic data, and social network information to make predictions regarding patient medical preferences.

This raises the following question: Should the AA or a relative serve as surrogate decision-maker? We argue that the common principles employed to make decisions for incapacitated patients recommend that all relevant stakeholders should defer to the AA's recommendation when the following conditions are met: (a) an adult patient without decision-making capacity has not issued a medical power of attorney, (b) relevant values and preferences can be inferred from their digital footprint and other data

sources, and (c) the patient lacks an advance directive specifying their treatment preferences. [1] Stakeholders should defer to the AA because the AA will likely be better at predicting what treatment option the patient would have chosen and avoiding bias, which would aid in identifying a medical option that is more patient-centered. Furthermore, we argue that these considerations override any moral weight given to the patient's special relationship to their family.

Some patients who meet the aforementioned conditions may have the preference that the AA not make recommendations on their behalf. In such cases we believe that the AA should recuse itself from identifying these patients' preferred treatments. It is worth noting, however, that at least one study suggests that most patients prefer that an algorithm be at least part of a shared decision-making process with surrogates and physicians (Wendler, et. al., 2016).

In what follows we begin with an overview of how medical decisions ought to be made on behalf of incapacitated patents. Next, we discuss the current status of AI in medicine and its ability to predict preferences, values, and personality traits. We then discuss how an autonomy algorithm would function to predict incapacitated patients' treatment preferences. We then present our main argument in support of using the AA to make decisions for incapacitated patients. Finally, we respond to a number of potential objections to our view.

Before proceeding, we acknowledge that our argument may depend on the kind of incapacitated patient that is involved, including children. [2] For our purposes, we will not consider whether the AA should make decisions for patients – like young children – who have never had capacity. It may be that parents have a special prerogative to direct the lives of their children that overrides the AA, but this issue is simply beyond the scope of our paper. Our proposal applies only to patients who meet conditions a – c above and who were previously capacitated.[3]

Surrogate Decision-Making

It is common practice that the substituted judgment principle (SJP) and the best interest principle (BIS) should guide surrogate decision-making for incapacitated patients.[4] According to the SJP, the surrogate ought to choose the option the patient would have chosen had she had capacity.[5] By contrast, according

---

[1] Although we believe that the AA should have sole decision-making authority, we also believe that this authority is contingent on regulatory oversight. Moreover, we believe that the actual process of implementing the practice of AA deferral should proceed gradually. We discuss these issues later in our article.

[2] Other types of incapacitated patients include the intoxicated, the mentally ill, seniors with dementia, and patients with little to no consciousness.

[3] For non-capacitated persons who never previously had decision-making capacity, the AA would be unable to apply the substituted judgment standard. The same, of course, is true of human surrogates.

[4] Many states designate a priority of surrogate decision-makers (typically with spouses given top priority, parents second, adult children third, etc.) in the absence of an advance directive and that such surrogates should make decisions based on these principles.   However, these principles apply even when the patient has established an advance directive, as advance directives cannot specify a course of action for all possible scenarios. As Allen Buchanan and Dan Brock state, "[s]ince instructional advance directives can neither cover every contingency nor be fully self-explanatory, someone must be identified as having principal responsibility for interpreting and applying the instructional advance directive as choices arise" (135, 1989).

[5] The conventional view is that the SJP ensures that the patient's autonomy is respected. Others argue that the principle is responding to the patient's authenticity or dignity (Brudney, 2009). It makes no difference for our argument what the moral basis of the SJP is. All that matters for our purposes is that the principle itself is correct.

to the BIS, the surrogate should choose what would be in the patient's best interest.[6] It is commonly held that the SJP should take priority over the BIS, that the latter only goes into effect if the patient's preferences are unknown (Beauchamp and Childress, 2001). We will only appeal to the SJP to make our case, since we argue that the AA would make recommendations based on the patient's known preferences.

The primary moral reason for appointing a surrogate decision-maker is that they are best able to fulfill the SJP. Indeed, most states have laws that identify a prioritized list of surrogate decision-makers that give priority to those closest to the patient such as their spouse, adult child, or parent.[7] There are at least two moral reasons in favor of these laws. First, the familial surrogate presumably knows the patient better than the healthcare professionals caring for them (Beauchamp and Brock, 136, 2001). This enables the surrogate to choose options that the patient themself would have chosen if capacitated. As Daniel Brudney writes, "We want the surrogate to make the decision because we believe that she has a relevant epistemological advantage over anyone else on the scene. If and when she has no such advantage…then there might not be sufficient reason to let her be the decision-maker" (38, 2009). Second, the intimate – such as a close relative or friend – is presumed to be motivated to act on the patient's behalf.[8]

This rationale for appointing intimates to make decisions for incapacitated patients suggest two criteria for identifying a surrogate: The Criteria of Epistemic Advantage and the Criteria of Fidelity. The former claims that whoever has the highest epistemic advantage – whoever knows the patient best – should be designated as surrogate decision-maker. The latter claims that the surrogate decision-maker should be motivated to act on behalf of the patient's preferences rather than, say, the surrogate's own preferences. We return to these criteria below.

The Current Status of Artificial Intelligence

Current research in artificial intelligence focuses on machine learning algorithms that identify patterns from large amounts of data (Yu et al., 2018). This kind of AI already plays important roles in decision-making. For example, in 2014 a venture capital firm appointed an algorithm named Vital to its board of directors.[9] It is also now common for algorithms to aid in investment decisions.

In their overview of artificial intelligence in healthcare, K. H. Yu and colleagues discuss how "AI systems have specialist-level performance in medical diagnostic tasks, can better predict patient prognosis than clinicians, and can assist in surgical interventions" (2018). For example, in a recent experiment, a deep-learning algorithm outperformed dermatologists in diagnosing skin malignancy (Esteva, et al. 2017). Additionally, an AI system has been developed that can diagnose diabetic retinopathy just as well as

---

[6] See Allen Buchanan and Dan Brock (1989, 31-34).

[7] We say primary because the BIS is the other, secondary, ground for appointing surrogates. Again, our focus is only on the SJP.

[8] Of course, the rationale for appointing intimates is also partly practical: intimates are usually more readily available to responsibility make choices for the patient.

[9] See "Algorithm Appointed Board Director," *BBC News*, May 16, 2014, www.bbc.com/news/technology-27426942; and Sophie Brown, "Could Computers Take Over the Boardroom?," *CNN Business*, October 1, 2014, www.cnn.com/2014/09/30/business/computers-ceo-boardroom-robot-boss/index.html.

physicians (Gulshan 2016). And it has been reported that IBM's AI 'Watson' has a success rate of 90% in diagnosing lung cancer, while physicians' diagnostic success rate is only 50%.[10]

In addition to diagnostic tasks, algorithms are also able to predict Big Five personality traits: openness, conscientiousness, extraversion, agreeableness and neuroticism. A meta-analysis showed "that predictive power of digital footprints over [Big Five] personality traits is in line with the standard "correlational upper-limit" for behavior to predict personality" (Azucar, et. al. 2018).[11] [12]

Furthermore, one study found that Facebook likes "can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender" (Kosinski et al., 2013). Considering that the data input in this study was only Facebook likes, adding additional factors of a digital footprint is highly likely to increase the accuracy and scope of AI's predictive power. Other studies attest to this power (Gou, et al. 2014; Yarkoni 2010; Chen et al. 2014; Golbeck et al. 2011; Warshaw et al. 2015; Marengo, D and Settanni, M, 2019).

So far, we have shown that AI is able to predict many aspects of one's identity and we suspect that prediction capacity will only increase over time. But what about AI's potential to predict medical preferences in particular? There is some evidence of this potential. One study suggests that Big Five "personality traits are associated with specific health care preferences" related to end-of-life care among men with prostate cancer (Lattie, et al). For instance, the study showed that "neuroticism was associated with a lower preference for palliative care, whereas agreeableness was linked to a lower preference for life support" (Lattie, et al). This provides some evidence to believe that AI's potential to predict medical preferences is real.

Given that AI is able to predict Big Five personality traits based on social media data, and that there are strong associations between these traits and certain medical preferences, future AI, with appropriate data, will likely be able to predict medical preferences for incapacitated persons. While speculative, we believe that AI's ability to identify additional aspects of a person's identity will increase its ability to accurately predict medical preferences.

How the Autonomy Algorithm Would Work

According to Camillo Lamanna and Lauren Byrne, the AA is a computational process that "takes data about patients as input and derives a confidence estimate for a particular patient's predicted health care-related decision as output" (2018). Specifically, the algorithm would draw on social network data to predict what the patient would choose if they had capacity, as well as sociodemographic data such as age, gender, and education. Annette Rid and David Wendler have documented how these kinds of sociodemographic factors can influence treatment preferences (2014).   In addition, the AA would mine the patient's EHR, population-wide EHRs, and the vast reservoir of medical research to predict which treatment options would align with the patient's substituted judgment.

Shahla Siddiqui and Voo Teck Chuan distinguish between two kinds of social networking data that could help inform the algorithm: "One type is a specific expression of a preference regarding a particular

---

[10] https://www.wired.co.uk/article/ibm-watson-medical-doctor (accessed 5/9/2019)
[11] We thank an anonymous reviewer for pointing out this study.
[12] Relatedly, another study showed that mining digital footprint data "can effectively predict consumers decision-making styles" (Chen Y. J. et. al.)

treatment or scenario. Another would be a general expression that discerns a person's life outlook and values" (2018). This data could be drawn from Facebook likes, comments on social media platforms, time spent watching videos on certain topics, link navigation, and other indicators of preference and personality traits.

We should note that it is already not uncommon to use social media data to inform important legal decisions. As Jessica Berg remarks, "Social media have been used in family law (e.g. divorce and child custody), employment law, and criminal law (to sow criminal behavior and juror misconduct). Courts are unlikely to reject the use of social media in surrogate decision-making cases" (2012). Additionally, it is not uncommon for clinicians to search their patient's social media data to aid in determining treatment (Jent et al. 2011). Incorporating social media data into the decision-making process – and so into the input of an algorithm – would therefore be extending precedent. Indeed, using information from social media to make decisions for incapacitated patients will likely become even more prevalent and can be "especially useful in understanding an incapacitated person's wishes" (Jent et al. 2011).

Some may worry that a person's online presence may not accurately portray their true self, given peoples' apparent propensity to present an overly flattering view of themselves. There is evidence, however, that this concern is misguided. One study supports the view that "people use OSNs [online social networks] to communicate their real personality;" the study also found "no evidence of self-idealization" (Back, et al. 2010). Other studies corroborate this claim (Turkle 1997; Vazire and Gosling 2004).

Argument for the Autonomy Algorithm

The discussion so far leads us to three main considerations that attest to the AA's ability to satisfy the Criterion of Epistemic Advantage. First, studies have shown that surrogates predict patients' preferences no better than the base-rate (Smucker et. al 2000; Houts 2002; Shalowitz et. al., 2007). If the AA can outperform the base-rate, then it will likely predict patient preferences more accurately than intimates. The studies we mentioned earlier, which show that algorithms are able to derive accurate personality traits and values from a person's digital footprint, suggest that the AA would likely meet this condition.

Second, one recent study suggests that an algorithm's predictive accuracy can exceed that of intimates. This study found that an algorithm, based on Facebook 'likes', predicted aspects of a person's personality better than that person's friends and family (Youyou et al. 2015). The algorithm required 300 likes to outperform a spouse's prediction, 150 likes to outperform family members, and seventy likes to outperform friends. Since preferences and personality are closely tied, these studies suggest that an algorithm is likely to generally predict preferences better than intimates.

Third, surrogate decision-makers are notoriously bad at choosing an option that the patient would choose. A meta-review of sixteen studies found that surrogates are able to predict patients' medical preferences with sixty-eight percent accuracy (Shalowitz et al. 2006). Another study showed this rate to be sixty-three percent (Barrio-Cantalejo et al. 2009). Indeed, many studies have suggested that surrogates generally fare poorly in accurately predicting the decisions the patient would make (Pope 2010). These studies also attest to the AA's likely epistemic advantage.

What about the Criterion of Fidelity, the criterion that states that surrogates should be motivated to act solely on the basis of patients' preferences? Would the AA satisfy this criterion? It is likely that the AA will generally satisfy the Criterion of Fidelity better than intimates. A number of considerations, advanced by Thaddeus Pope, support this claim. First, "surrogates frequently do not understand the clinical status of the patients whom they represent;" second, "many surrogates have clinically diagnosable conditions such as stress, depression, and anxiety;" third, surrogates often have difficulty distinguishing their preferences from the patient's preferences (2010; Marks and Arkes 2008). Since the AA is not vulnerable to these psychological shortcomings, it is likely it would more reliably decide in accordance with the patient's preferences.

We have shown that there are strong reasons to believe that the AA would better satisfy the Criteria of Epistemic Advantage and Fidelity than intimates. Thus, based on the SJP, decision-makers should defer to the AA's treatment recommendation.

In addition to satisfying the SJP better than intimates, the AA would also relieve family and other proxies – such as physicians – of the burden of identifying the patient's treatment preferences, which can often be stressful (Rid and Wendler 2014). It may even deliver economic benefits by preventing treatments that patients would prefer to avoid.

Objections

In this section we address a number of potential objections to our view. The first objection, inspired by Allen Buchannan and Dan Brock's influential theory of surrogate decision-making, is that a patient's family has the right to make medical decisions on the patient's behalf on account of their special relationship to the patient, and that this right overrides all non-familial claims to make decisions on the patient's behalf. Mainly two considerations support this objection: (1) families typically have greater knowledge of the patient's values and are more concerned with the patient's good than others. Furthermore, (2) because all of us have a strong interest in intimate relationships, overriding the familial surrogate's right to make medical decisions is impermissible (Buchanan and Brock 136-7, 1989).

Let us respond to these considerations. In response to (1): We agree that family members typically know more about their loved one's values and that they are usually more concerned with their best interests than others are. Thus, we agree with consideration (1) above. But we disagree with what it is meant to show. For in comparison with the AA, which has superior epistemic advantage compared to the patient's family, we believe that the family's claim to make decisions based on their epistemic position is undermined.

Consideration (2), again, is the idea that we should be reluctant to interfere with the family since we all have an interest in forming intimate relationships. In response: we believe that the loss in value of not deferring to a family member's decision would be substantially outweighed by the value gain in promoting the patient's autonomy via the AA. Additionally, if deferring to an AA undermines the value of special relationships, and this is a strong enough reason to not defer to the AA, then presumably it would also be a reason strong enough to override patients' autonomous choices when they undermine this value. This, however, is absurd and, so, by parity of reasoning we should reject this consideration. Note that the absurdity is evident regardless of whether the patient is capacitated or not; it is obvious insofar as overriding the patient's autonomy, based on familial considerations, is clearly unacceptable.

We suspect that to the extent that this defense of the family's right to make medical decisions seems plausible, it is due to the fact that we ignore the substantial epistemic gains the AA will offer. This is understandable, after all, simply because we are only beginning to understand how artificial intelligence will change the world. Nevertheless, the standard justification for the presumption in favor of familial autonomy that we discussed above is simply less persuasive in light of the AA's epistemic advantage.

Another point here is relevant: We are highly suspicious that familial independence would be significantly undermined by the AA. After all, many of the decisions previously reserved for families – about, for example, what kind of education children should receive or who should make reproductive choices – we now correctly believe should reside, at least partly if not completely, with others. For example, in regard to education, we now correctly believe that the state should require that children have at least basic mathematical knowledge and critical thinking skills to prepare them for a life in a pluralistic democratic society. We now also correctly believe that those who are capable of giving birth should at least have the primary say in whether or not to do so. And yet, in spite of these losses to familial independence, few of us think the family's autonomy has been substantially curtailed (or, if we do, most of think this curtailment is justified). Instead, we think the family has simply been rebalanced in favor of other values, such as women's equality and children's future autonomy.

The second objection to our view is that it would be dehumanizing. As Camillo Lamanna and Lauren Byrne write, "We submit that it is the process of making a decision that is humanizing and autonomy affirming. Therefore, it would be dehumanizing to automate this process and defer to algorithmic outputs as a matter of course" (Lamanna and Byrne, 908). In reply to this objection, a number of points are in order. First, this appeal to dehumanization needs to be articulated. A person is dehumanized when that individual is regarded in such a way that removes the qualities that ground their worth as a person, such as their agency. We do not think that any individual would be dehumanized in this manner by deferring to an AA's decision. While the deliberative role – which is part of agency – is removed from the family, our thesis does not entail that a family member's deliberative capacity would be undermined. Part of one's humanity is the capacity for agency and deliberation. Deferring to the AA would not undermine this capacity just as deferring to the next of kin does not undermine the deliberative capacity of a more distant family member who may want to be part of the deliberative process. Our thesis, therefore, does not threaten the humanity of the patient's intimates.

Perhaps the worry is not that any individual would be dehumanized, but rather that deferring to the AA would dehumanize the context of medical decision-making by removing the 'humanity' from it. The argument for this worry might be something like the following: In the medical decision-making context, we express our humanity through a process of interpersonal deliberation. But deferring to an algorithm would prevent interpersonal deliberation by replacing a human-oriented process – in this case decision-making – with a tech-oriented process. Therefore, designating the AA would fail to express our humanity in the medical decision-making context.

In response, even if deferring to the AA is dehumanizing in this sense, we believe that the benefit of the AA outweighs the cost. After all, modern medicine avails itself of every available technology to diagnose and treat patients. These technologies might remove some notion of humanity from the decision-making context, yet they are clearly warranted. Moreover, the loss of humanity is likely more than made up for by the gain of bringing the patient's values and preferences more accurately into the

decision-making process. This gain could also easily be characterized by the same notion of humanity at work in the objection.

The third objection alleges that the AA would be subject to algorithmic bias.[13] Indeed there have been many cases in which an algorithm's output is biased against certain groups. For example, bias against African Americans has been detected in algorithms designed to predict recidivism.[14] In response, we acknowledge that the AA would not be immune to bias. We believe, however, that this concern does not undermine our proposal. There is already much work being done on identifying and preventing algorithmic bias (Courtland, 2018). This work should be implemented when developing the AA and its team of developers should include an algorithmic bias specialist. Moreover, once the AA is developed, it should be periodically audited by an independent board to detect for any bias. It is worth emphasizing that the AA would analyze personal data from social media to detect preferences in addition to preference correlations in demographic data. This personalized approach may be able to help prevent algorithmic bias, by, for example, training the AA to identify the patient's preferences by examining their social media comments about end-of-life issues (Berg, 2012).

It is also worth noting that, in addition to the AA itself, the institutions that develop the AA might contribute to algorithmic bias. For example, for-profit institutions could nudge the algorithm to recommend more expensive treatment options, even if the patient would not prefer them. We believe that this is a serious concern and it explains why regulatory bodies should provide oversight of AA developments.[15]

Conclusion

In this paper we have argued that the autonomy algorithm should make decisions for incapacitated patients rather than their family members. This recommendation is at odds with the current and widely shared practice of deferring to family. In an effort to transition us away from this practice to our recommended one in favor of the AA, we believe it is wise that the AA be developed and tested in clinical trials in an effort to build public trust in the algorithm. A single study has already suggested that most people would want such an algorithm to be incorporated into the decision-making process if they were incapacitated (Wendler et al. 2016). Once public trust is built, there should be a policy in which, as a default, the AA act as the surrogate for patients who have not designated a medical power of attorney. Importantly, however, we believe that patients should have the right to opt out of this default policy, for instance in their advance directive.

In conclusion, although we have argued for vesting the AA with sole decision-making authority, we also believe that transitioning to our AA-centric proposal should be a gradual process. In an effort to arrive at

---

[13] We thank three anonymous reviewers for pressing us to address this issue.

[14] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[15] Relatedly, the algorithm should be as transparent as possible. For example, it would be reasonable for family members to inquire about how the AA made its recommendation. Families who disagree with the AA's recommendation would need to provide evidence for their view. Moreover, our proposal to use AA over families is consistent with employing the AA in this fashion only after an extensive period of public commentary and involvement. Transparency is also important for building public trust in the AA. While deep learning systems are notorious for being 'black boxes,' work is being done on making its processes explainable (Montavon, et. al.). We thank an anonymous reviewer for raising these issues.

our AA-centric proposal, the AA should initially be incorporated into the shared, medical decision-making process with familial surrogates.

Works Cited

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, *124*, 150-159.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological science*, *21*(3), 372-374.

Barrio-Cantalejo, I. M., Molina-Ruiz, A., Simón-Lorda, P., Cámara-Medina, C., Toral Lopez, I., del Mar Rodriguez del Aguila, M., & Bailón-Gómez, R. M. (2009). Advance directives and proxies' predictions about patients' treatment preferences. *Nursing Ethics*, *16*(1), 93-109.

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics*. New York: Oxford University Press, USA.

Berg, J. (2012). Surrogate decision making in the internet age. *The American Journal of Bioethics*, *12*(10), 28-33.

Brudney, D. (2009). Beyond autonomy and best interests. *Hastings Center Report*, *39*(2), 31-37.

Buchanan, A. E., & Brock, D. W. (1989). *Deciding for others: the ethics of surrogate decision making*. Cambridge: Cambridge University Press.

Chen, Y. J., Chen, Y. M., Hsu, Y. J., & Wu, J. H. (2019). Predicting Consumers' Decision-Making Styles by Analyzing Digital Footprints on Facebook. *International Journal of Information Technology & Decision Making*, *18*(02), 601-627.

Chen, J., Hsieh, G., Mahmud, J. U., & Nichols, J. (2014, February). Understanding individuals' personal values from social media word use. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 405-414).

Courtland, R. (2018). Bias detectives: the researchers striving to make algorithms fair. *Nature*, *558*(7710), 357-357.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118.

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 253-262.

Gou, L., Zhou, M. X., & Yang, H. (2014, April). KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 955-964.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, *316*(22), 2402-2410.

Houts, R. M., Smucker, W. D., Jacobson, J. A., Ditto, P. H., & Danks, J. H. (2002). Predicting elderly outpatients' life-sustaining treatment preferences over time: The majority rules. *Medical Decision Making*, *22*(1), 39-52.

Jent, J. F., Eaton, C. K., Merrick, M. T., Englebert, N. E., Dandes, S. K., Chapman, A. V., & Hershorin, E. R. (2011). The decision to access patient information from a social media site: what would you do?. *Journal of Adolescent Health*, *49*(4), 414-420.

Johansson, M., & Broström, L. (2011). Counterfactual reasoning in surrogate decision making–another look. *Bioethics*, *25*(5), 244-249.

Koppelman, E. R. (2002). Dementia and dignity: towards a new method of surrogate decision making. *The Journal of Medicine and Philosophy*, *27*(1), 65-85.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802-5805.

Lamanna, C., & Byrne, L. (2018). Should artificial intelligence augment medical decision making? The case for an autonomy algorithm. *AMA Journal of Ethics*, *20*(9), 902-910.

Lattie, E. G., Asvat, Y., Shivpuri, S., Gerhart, J., O'Mahony, S., Duberstein, P., & Hoerger, M. (2016). Associations between personality and end-of-life care preferences among men with prostate cancer: A clustering approach. *Journal of Pain and Symptom Management*, *51*(1), 52-59.

Marks, M. A., & Arkes, H. R. (2008). Patient and surrogate disagreement in end-of-life decisions: can surrogates accurately predict patients' preferences?. *Medical Decision Making*, *28*(4), 524-531.

Marengo, D., & Settanni, M. (2019). Mining facebook data for personality prediction: an overview. In H. Baumeister & C. Montag (Eds.), *Digital phenotyping and mobile sensing* (pp. 109-124). New York: Springer.

Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, *73*, 1-15.

Pope, T. M. (2009). Surrogate selection: an increasingly viable, but limited, solution to intractable futility disputes. . *Saint Louis University Journal of Health Law & Policy, 3*(2), 183-252.

Rid, A., & Wendler, D. (2014). Use of a patient preference predictor to help make medical decisions for incapacitated patients. *Journal of Medicine and Philosophy*, *39*(2), 104-129.

Shalowitz, D. I., Garrett-Mayer, E., & Wendler, D. (2006). The accuracy of surrogate decision makers: a systematic review. *Archives of Internal Medicine*, *166*(5), 493-497.

Siddiqui, S., & Chuan, V. T. (2018). In the patient's best interest: appraising social network site information for surrogate decision making. *Journal of Medical Ethics*, *44*(12), 851-856.

Shalowitz, D. I., Garrett-Mayer, E., & Wendler, D. (2007). How should treatment decisions be made for incapacitated patients, and why?. *PLOS Medicine*, *4*(3).https://doi.org/10.1371/journal.pmed.0040035

Smucker, W. D., Houts, R. M., Danks, J. H., Ditto, P. H., Fagerlin, A., & Coppola, K. M. (2000). Modal preferences predict elderly patients' life-sustaining treatment choices as well as patients' chosen surrogates do. *Medical Decision Making*, *20*(3), 271-280.

Smith, A., & Anderson, M. (2018). Social media use in 2018. Resource Document. Pew Research Center.

https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2018/02/PI_2018.03.01_Social-Media_FINAL.pdf

Turkle, S. (1997). Life on the screen: identity in the age of the internet. 1995. *New York: Touchstone*

Vazire, S., & Gosling, S. D. (2004). e-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology*, *87*(1), 123.

Warshaw, J., Matthews T., Whittaker S., Kau C., Bengualid M., & Smith B. (2015) *CHI'15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 797-806

Wendler, D., Wesley, B., Pavlick, M., & Rid, A. (2016). A new method for making treatment decisions for incapacitated patients: what do patients think about the use of a patient preference predictor?. *Journal of Medical Ethics*, *42*(4), 235-241.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363-373.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036-1040.

Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering* 2, 719–731.