Uncovering User Affect Towards AI in Cancer Diagnostics

Stephanie Tom Tong$^{(\boxtimes)}$ and Pradeep Sopory

Wayne State University, Detroit, MI 48201, USA
{stephanie.tong,psopory}@wayne.edu

Abstract. Despite the rapid application of artificial intelligence (AI) to healthcare, we know comparatively little about how users perceive and evaluate these tools. Following "dual route" theories of information processing from decision science, we propose that because users lack the expertise to rationally understand AI through cognitive evaluation, they rely on their feelings or heuristic route processing to make judgments about AI systems and recommenders. Therefore, *affect* becomes an important component that influences people's willingness to adopt AI—and this may be especially true in a context like personal health, where affect is both explicit and heightened. Using the context of remote dermatological skin cancer screening, we examined people's affective perceptions of an autonomous AI algorithm capable of making recommendations about skin lesions (as either cancerous or benign). In a three-stage study ($n = 250$), we found that people do hold complex affective responses toward AI diagnostics, even without directly interacting with AI. Findings are relevant to designers of AI systems who might consider how users' *a priori* affect may make them more or less resistant to technological adoption. Additionally, the methodological approach validated in this study may be used by other scholars who wish to measure user affect in future research.

Keywords: Affect · Healthcare · Audience · Artificial intelligence

1 Introduction

> *Is it really possible to tell someone else what one feels?*
> *– Leo Tolstoy, Anna Karenina*

Although automated artificial intelligence (AI) technology has been applied to assist decision making across many domains of human life, one burgeoning area where AI systems have recently come into practice is modern-day healthcare. Automated algorithms facilitate everything from treatment decisions, communication among medical staff and their patients, and data retrieval and storage [25]. Such systems serve as an aid to patients and practitioners, with most systems requiring a human user to input commands and data.

However, recent gains in AI are dispensing with the need for human oversight: For example, in 2017, computer scientists at Stanford University [9] and the University of Heidelberg [11] developed a deep learning algorithm capable of detecting skin cancer by scanning thousands of images of patients' skin. The algorithm did so autonomously,

gaining accuracy with practice, and validation tests have shown that it either matched or outperformed board-certified dermatologists at diagnostic classification of patients' skin photographs without any human supervision [9, 11].

Other types of AI technology are now being offered directly to patients through mobile applications. A recent study counted 40 different mobile dermatology applications across Android and Apple systems that allow patients to "upload and receive dermatologist or algorithm-based feedback about the malignancy potential of lesions" [1]. Although several researchers and developers have worked to refine these autonomous AI technologies, few of them have paused to question whether people will entrust it with something as important as their personal health. There have been many popular press stories touting the promise of new mobile apps, recommender systems, and AI tools designed to improve people's personal health, yet so many of these systems fail to reach their potential, instead meeting their demise after users fail to fully trust them [13]. Users' reluctance to adopt AI technologies has also led some academic researchers to similarly conclude that AI will never fully enter into the realm of personal health if the user interface design does not meet users' needs [14].

One component currently being overlooked in the development of many of these automated tools for personal health is *audience affect* toward these AI technologies. This is a critical oversight because how people feel about these systems may influence their decision to adopt them. Therefore, this study poses the question, "How do people perceive and evaluate AI technology in the context of their personal health?"

2   Background

Following "dual route" theories of decision making, we propose that when evaluating any kind of unfamiliar or new object (in this case, AI diagnostic technologies), people often rely on cognitive routes of rational logic and affective routes of feeling. However, the context of personal health is a very unfamiliar one—and in such a case, we argue that people are more likely to rely on affective routes for decision making, as opposed to cognitive routes. Affect has been shown to be important form of information during decision making when evaluating unfamiliar objects [18], particularly with regard to healthcare decisions [10]. In this study, we define *affect* as the "good" or "bad" feelings we experience—both consciously and unconsciously—and posit that these positive or negative feelings color our evaluations, decisions, and judgments [18, 22]. These positive and negative feelings comprise our *integral affect,* which is the form of affect that occurs as a product of the decision or during the decision making process itself [18, 24].

Most existing studies have focused on people's direct thoughts and cognitions about AI, which are often derived after some experience with a system. Examining only users' cognitive route processing excludes an entire route of affective heuristics that may influence users' perceptions and attitudes toward AI recommender systems. Interestingly, those who have explored users' integral affect have found it to be an important component of their acceptance of algorithmic recommender systems [6, 15]: For example, recent work in human-computer interaction reveals that affect plays an important role in audiences' perceptions of the AI algorithms found in popular social media systems like Facebook and Twitter [8].

Elsewhere, Katz and colleagues [14] examined the nature of user acceptance of systems designed to aid in the management of Type 1 diabetes. Their findings suggested that one factor contributing to low adoption rates of these mobile applications was a failure to meet users' emotional needs; instead user interface designs lacked emotional sensitivity, causing a strong negative emotional response for users that led to rejection of the technology. As these results suggest, failure to accurately assess and account for user affect can have negative consequences for technological adoption.

This brief review shows that although a more specific focus on user affect would be relevant to AI research, yet much existing work has not explicitly examined the affective component in user evaluation of AI. Thus, this study's main contributions are to carefully examine the affective element of the human decision making process and to develop and validate a methodology to measure it, so affect can be easily assessed and accounted for in future AI systems research.

2.1 The Importance of Affect in AI Acceptance

To summarize, we propose that if users cannot rely on cognitive route processing when evaluating AI in a familiar context like social media, they are even less likely to do so in a less familiar context like healthcare. In forgoing cognitive routes and logical evaluation, people will turn to their feelings to make judgments about AI, and whether or not they should accept them [23]. Thus, knowing whether people will accept or reject AI for their personal health requires knowing more about their affective response.

2.2 The Present Study

Building on our review of studies on social recommenders, product recommenders, and AI mobile health applications, we hypothesize that users will have a complex affective response to autonomous AI in the context of cancer screening as well. Importantly, we conceptualize a priori integral affect as separate from but related to other key constructs in the recommender systems literature such as users' post facto trust or confidence, which are often derived after receiving detailed explanations of the algorithm's functions or after direct observation or contact with the algorithm itself [4, 9, 14, 16]. In this study, we assert that users' affect toward AI can be formed without ever interacting with it or seeing it make a recommendation. In this sense, people's "first impression" of AI produces an affective response. It is this initial affective response that we predict will impact people's acceptance of unfamiliar AI technologies. We examine this *a priori* affective response is what we examine in the current study.

To investigate this line of thinking, we rely on the *affect heuristic* framework from decision science [23] to predict that people's affect towards AI will be consistent with their evaluations of its potential risks or benefits. In essence, we theorize that a user's evaluation of technology follows his or her feelings. Specifically, the affect heuristic predicts that people develop an "affect pool" containing multiple "tags" of both good and bad feelings associated with a specific object. When people are asked to evaluate how risky or beneficial that object or technology is, they consult those affect pools for information. When affect pools contain positive feelings about technology, they judge

the technology's overall risks as low and potential benefits as high; but when they have negative affect towards technology, risks are evaluated as high and benefits as low.

The affect heuristic has been applied to understand people's affective response to many kinds of technologies—including nuclear power, pesticides, and food additives [see 23]. The current study extends its application into AI related affect. Therefore, the first step in understanding people's judgments regarding AI technology is to uncover people's affective response. Once audience affect is better understood, we can more accurately assess people's intent to accept or reject autonomous AI for personal health.

3   Method

This study investigated people's affective response to AI diagnostic technology using dermatology screening as the context. We created an illustrative scenario (Fig. 1) and then pretested it via a focus group of adults recruited from an urban university ($n = 12$, 4 male) that judged it as believable. After refining the scenario based on the focus group feedback, we executed the study in three stages: (1) affective item generation, (2) item refinement and scale creation, and (3) test of scale validation. We report on the results of all three stages of the study below.

> You have had a mole on your arm for years. It has been there as long as you can remember—since you were young. But over the last month, you have noticed that the mole has begun to change in color, shape, and size. You are now concerned enough about the mole that you decide to have it examined for signs of skin cancer.
>
> Recently, you have discovered a new form of skin cancer screening known as direct-to-patient dermatology. Direct-to-patient dermatology uses a computer algorithm to examine patients. The algorithm is a type of artificial intelligence technology. Instead of going to the dermatologist's office, patients can take photographs of their skin using their cell phones. The algorithm then evaluates patients' photographs looking for signs of skin cancer. Because the algorithm can screen for signs of skin cancer without any human help, it might make a doctor's appointment unnecessary.

Fig. 1. Experimental scenario text.

4   Results

4.1   Stage 1: Item Generation

In the first stage, the scenario in Fig. 1 describing AI dermatological screening with a deep learning algorithm was presented to another focus group of 25 participants (10 male). These focus group participants provided up to five words describing how they felt about AI algorithm described in the scenario. This procedure resulted in 43 unique affect-related words and phrases (e.g., surprising, worried, exciting, convenient, scary,

too new, etc.). These words and phrases were used to "seed" the survey items developed for Stage 2 of the study.

4.2   Stage 2: Scale Refinement

In the second stage, the scenario was presented to a sample of participants recruited from Amazon Mechanical Turk (Mturk) ($n = 85$ participants; 56 male). Due to the nature of the experimental context of skin cancer, we specifically sampled for individuals who have an elevated risk of developing skin cancer (e.g., people of Caucasian, non-Hispanic descent).

Mturk participants have been shown to successfully perform a range of experimental tasks [5], and often show great amounts of intrinsic motivation and demographic diversity [2, 3, 12]. As Mturk participants would also be somewhat familiar with computing technology, they were considered an ideal population for the current topic of investigation, AI.

After providing informed consent, participants answered basic demographic items (sex, age, race). They then indicated their familiarity with computer programming and computer algorithms using two items adapted from Lee and Baykal [17], "Which statement best describes your knowledge of computational programming (algorithms)?" followed by the response scale: 0 = "I have no knowledge at all", 1 = "a little knowledge", 2 = "some knowledge", 3 = "a lot of knowledge" ($r = .75$). The sample had a moderate familiarity with computing technology, $M = 1.87$, $SD = 0.79$.

Participants read the scenario and then provided their responses to the question "Please indicate how you feel about the AI technology described in the above situation" for all 43 affective items from Stage 1, using a 5-point scale with 1 = "not at all" to 5 = "a great deal". These responses were examined using exploratory factor analysis with promax rotation that, after dropping non-loading and cross-loading items, revealed 7 items reflecting *positive affect toward AI* ($M = 3.44$, $Mdn = 3.57$, $SD = 0.88$, $\alpha = .93$) and 7 items reflecting *negative affect toward AI* ($M = 2.04$, $Mdn = 1.85$, $SD = 0.95$, $\alpha = .82$) that together accounted for 53% of the variance (see Table 1).

Table 1. Item labels and factor loadings from Stage 2.

Item wording	Factor loadings
Positive affect	
Good	.71
Convenient	.70
Positive	.74
Hopeful	.90
Innovative	.76
Intriguing	.87
Interesting	.77
Negative affect	
Negative	.69
Nervous	.81
Unsure	.79

(*continued*)

Table 1. (*continued*)

Item wording	Factor loadings
Hesitant	.83
Concerning	.82
Scary	.86
Bad	.69

4.3   Stage 3: Scale Validation

In Stage 3, we assessed the final affect scales for construct validity by using a new sample of Mturk participants ($n = 140$, 82 male). We also assessed more specific demographics for this final validation study sample that consisted of participants with a range on *educational background* (19 = high school, 50 = some college, 58 = college degree, 13 = graduate degree) and *annual household income* (reported in US dollars; 22 = less   than $25,000,   52 = $25,000–$49,999,   35 = $50,000–$74,999, 15 = $75,000–$99,000, 16 = $100,000+). This sample also had a moderate level of familiarity with computer programming and algorithms, $M = 2.06$, $SD = 0.81$. As in Stage 2, we oversampled for individuals who have an elevated risk of developing skin cancer resulting in a final sample of: 116 = Caucasian/white,   7 = African-American/Black, 10 = Asian, 4 = Hispanic/Latino, 1 = Native American, 2 = other.

Table 2. Correlations of AI positive and negative affect with Technology Readiness Index.

Variable	1	2	3	4	5
1. Negative affect					
2. Positive affect	−.15*				
3. TRI innovation	−.27+	.15*			
4. TRI optimism	−.22+	.09	.50+		
5. TRI discomfort	.23+	.01	−.40+	−.51+	
6. TRI insecurity	.31+	−.19*	−.26+	−.32+	.36*

$N = 140$. *$p < .05$; +$p < .01$ (one-tailed).

After answering these demographic questions, participants answered the Technology Readiness Index (TRI), which assesses people's readiness to embrace new technologies across the four dimensions of *innovation, optimism, discomfort, and insecurity* [20]. Lastly, participants read the illustrative scenario in Fig. 1 and provided responses on the final set of affect scales.

Analyses revealed strong evidence of construct validity, with users' algorithmic affect scores correlating with the TRI across multiple dimensions in expected directions. Specifically, participants' overall positive AI affect toward the diagnostic algorithm was inversely associated with their scores on TRI insecurity and directly associated with feelings of innovation. Participants' negative AI affect scores were directly associated with their TRI insecurity and discomfort scores, and negatively associated with TRI innovation and optimism (see Table 2).

5   Discussion

Understanding user affective response to AI health systems is a necessary, but currently missing, piece of the HCI technology landscape. The results of this study suggest that people (a) have a complex affective response to algorithmic systems in the context of personal health and that (b) it is measureable.

These results shed new light on the role that integral affect plays in people's attitudes toward AI health technologies. Interestingly, this study demonstrates that people do have an affective response to AI technology and tools, even without ever interacting with it directly. This is similar to what some theorists have described as *dispositional trust* in machines, which—like the current study's measure of integral affect toward AI algorithms—is rooted in people's schema or heuristics regarding technology rather than in direct contact with it [19]. That people carry such affect into their decision making is an important consideration for scholars, medical professionals, and developers to consider.

Scholars who are working in the area of AI system adoption may consider how this pre-existing affect toward technology may be associated with users' likelihood to adopt those systems. Should other researchers wish to measure a priori user affect, this study provides a validated methodological approach and tool for assessing affect that can be easily applied to other contexts and forms of AI technology.

Medical professionals might consider that their patients' affective responses during decision making is driven not only by what they know, but also their feelings. Interestingly, medical professionals are often trained to balance patients' informational concerns regarding both "biomedical and psychosocial issues" [21] and ensuring that patients are well-informed by "communicating clinical evidence" and "presenting recommendations informed by clinical judgments" [7]. Common wisdom regarding patient decision making is that a well-informed patient will make wiser decisions [7]; however, the current results suggest that medical professionals ought to pay attention to their patients' emotion and affective responses as well as this is an equally important factor that may influence their decision making and behavior.

These findings are especially relevant to developers who are creating AI systems for application in contexts likely to be associated with high levels of affect, such as personal health. Designers might consider audiences' initial affective reactions to AI technologies, and the results of the present study can help them test audience response as they create the tools. Knowing ahead of time that they face high levels of *a priori* negative affect from their target audience may help designers address barriers to adoption up front, as opposed to during later stages of development or testing.

Though affective response to technology was shown to be multi-faceted, decisions about personal health often raises other strong feelings. When patients are asked to make decisions, they often weigh multiple factors such as medical technology, cost, technological efficacy, and side effects; each of these factors can create affect, thereby complicating the overall affective response. The current study focused specifically on medical decisions in the context of healthcare, but future studies might consider examining the affective response to AI technology in other decision making environment.

References

1. Brewer, A.C., et al.: Mobile applications in dermatology. JAMA Dermatol. **149**(11), 1300–1304 (2013). https://doi.org/10.1001/jamadermatol.2013.5517
2. Buhrmester, M., Kwang, T., Gosling, S.: Amazon's mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect. Psychol. Sci. **6**(1), 3–5 (2011)
3. Casler, K., Bickel, L., Hackett, E.: Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. Comput. Hum. Behav. **29**, 2156–2160 (2013). https://doi.org/10.1016/j.chb.2013.05.009
4. Chang, S.F., Harper, M., Terveen, L.G.: Crowd-based personalized natural language explanations for recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems, RecSys 2016, pp. 175–182. ACM, New York (2016). https://doi.org/10.1145/2959100.2959153
5. Crump, M.J., McDonnell, J.V., Gureckis, T.M.: Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One **8**(3), e57410 (2013)
6. Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P.: The role of trust in automation reliance. Int. J. Hum.-Comput. Stud. **58**(6), 697–718 (2003). https://doi.org/10.1177/1359105317693910
7. Epstein, R.M., Alper, B.S., Quill, T.: Communicating evidence for participatory decision making. JAMA **19**, 2359–2366 (2004). https://doi.org/10.1001/jama.291.19.2359
8. Eslami, M., et al.: First i "like" it, then i hide it: folk theories of social feeds. In: Proceedings SIGCHI Conference on Human Factors in Computing Systems (CHI 2016), pp. 2371–2382. ACM, New York (2016). https://doi.org/10.1145/2858036.2858494
9. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**, 115 (2017). https://doi.org/10.1038/nature21056
10. Ferrer, R.A., Green, P.A., Barret, L.F.: Affective science perspectives on cancer control: Strategically crafting a mutually beneficial research agenda. Perspect. Psychol. Sci. **10**(3), 328–345 (2015). https://doi.org/10.1016/j.copsyc.2015.03.012
11. Haenssle, H., et al.: Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann. Oncol. (2018). https://doi.org/10.1093/annonc/mdy166
12. Hauser, D.J., Schwarz, N.: Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behav. Res. Methods **48**, 400–407 (2016). https://doi.org/10.3758/s13428-015-0578z
13. Johnson, C.Y.: The tech industry thinks it's about to disrupt health care. Don't count on it. https://www.washingtonpost.com/news/wonk/wp/2018/02/09/health-care-the-industry-thats-both-begging-for-disruption-and-virtually-disruption-proof/?utm_term=.c7b4312afdd7. Accessed 12 Dec 2018
14. Katz, D., Price, B.A., Holland, S., Dalton, N.S.: Data, data everywhere, and still too hard to link: Insights from user interactions with diabetes apps. In: Proceedings of 2018 CHI Conference on Human Factors in Computing Systems. ACM, New York (2018). https://doi.org/10.1145/3173574.3174077

15. Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., Getoor, L.: User preferences for hybrid explanations. In: Proceedings of the 11th ACM Conference on Recommender Systems (RecSys 2017), pp. 84–88. ACM, New York (2017). https://doi.org/10.1145/3109859. 3109915

16. Lee, J., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors **46** (1), 50–80 (2004). https://doi.org/10.1518/hfes.46.1.50_30392

17. Lee, M., Baykal, S.: Algorithmic mediation in group decisions: fairness perceptions of algorithmically mediated vs. discussion-based social division. In: Proceedings of SIGCHI Conference on Computer Supported Cooperative Work (CSCW 2017), pp. 1035–1048. ACM, New York (2017). https://doi.org/10.1145/2998181.2998230

18. Loewenstein, G., Lerner, J.S.: The role of affect in decision making. In: Davidson, R.J., Scherer, K.R., Goldsmith, H.H. (eds.) Handbook of Affective Science, pp. 619–642. Oxford University Press, Oxford (2003)

19. Merritt, S.M., Ilgen, D.R.: Not all trust is created equal: dispositional and history-based trust in human-automation interactions. Hum. Factors **50**(2), 194–210 (2008). https://doi.org/10. 1518/001872008X288574

20. Parasuraman, A.: Technology readiness index (TRI) a multiple-item scale to measure readiness to embrace new technologies. J. Serv. Res. **2**(4), 307–320 (2000). https://doi.org/ 10.1177/109467050024001

21. Roter, D.L., Stewart, M., Putnam, S.M., Lipkin Jr., M., Stiles, W., Inui, T.S.: Communication patterns of primary care physicians. JAMA **277**, 350–356 (1997). https://doi.org/10. 1001/jama.1997.03540280088045

22. Schwarz, N.: Feelings-as-information theory. In: Van Lange, P., Kruglanski, A., Higgins, E. T. (eds.) Handbook of Theories of Social Psychology, pp. 289–308. Sage, Thousand Oaks (2011)

23. Slovic, P., Peters, E., Finucane, M.L., MacGregor, D.G.: Affect, risk, and decision making. Health Psychol. **24**(4S), S35–S40 (2005). https://doi.org/10.1037/0278-6133.24.4.S35

24. Västfjäll, D., et al.: The arithmetic of emotion: Integration of incidental and integral affect in judgments and decisions. Front. Psychol. **7**, 325 (2016). https://doi.org/10.3389/fpsyg.2016. 00325

25. Ventola, C.L.: Mobile devices and apps for health care professionals: uses and benefits. Pharm. Ther. **39**(5), 356–364 (2014)