Challenges and Implications of Measuring User Experience for Wellbeing Research

Effie Lai-Chong Law
Durham University
The Palatine Centre, Durham University
Durham, DH1 3LE, United Kingdom
effielaw17@gmail.com

ABSTRACT
The notion of User Experience (UX) is increasingly applied to wellbeing-driven AI-infused technology. Despite the trend, approaches to measuring UX still draw heavily on the conceptual and methodological frameworks of emotions; a variety of self-reported scales and psychophysiological tools exist. Chatbots are used as an example to illustrate the challenges pertaining to human-AI UX; implications for addressing them are inferred.

Author Keywords
User Experience, Measurability, Emotions, Wellbeing, AI

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Measurement, Design, Evaluation

DOI: 10.37789/rochi.2021.1.1.2

INTRODUCTION

The notion of User Experience (UX) emerged more than two decades ago to broaden the scope of usability by expanding the focus on cognition to acknowledge the crucial role of emotion when interacting with digital technologies. Intriguingly, a universal definition of UX is still lacking, despite different attempts in the last decade (e.g. [1, 2, 3]). In fact, some HCI researchers and practitioners do not see the necessity or utility of having *the* definition of UX. Nevertheless, according to ISO 9241-210:2019, 3.15, UX is defined as "*user's perceptions and responses that result from the use and/or anticipated use of a system, product or service.*" This overly broad definition is qualified with two notes. Note 1 refines the phrase 'users' perceptions and responses' by referring to "*users'* ***emotions****, beliefs, preferences, perceptions, comfort, behaviours, and accomplishments*" (my emphasis). Note 2 refines it further with a comprehensive list of constructs, covering almost all aspects of human psychology (cf. [4).

Clearly, the definitional issue of UX is strongly related to the measurement issue. The conceptual and methodological concerns on the measurability of UX addressed a decade ago remain largely unresolved [5]. It entails further research, especially the scope of applying the notion of UX is ever diversifying. The trend can explicate the exaggerating claim that "UX is dead"[1], because the term UX has been eclipsed or replaced by the closely related ones with two prominent examples, namely Service Design (SD) (cf. the recent debate on the boundaries between UX Design and SD [6]) and Positive Computing [7]. This paper focuses on the latter.

The notion of Positive Computing emerged in 2012. It is referred to the study and development of technologies designed to enhance wellbeing, wisdom, and human potential [7]. Accordingly, computing technologies can be utilized to support people in pursuing happiness and sustaining mental and emotional health. Apart from the common focus on designing interactive digital technologies, Positive Computing and UX share the conceptual grounding in emotion and affect[2]. Challenges for understanding, defining and measuring emotions are thus relevant to both notions, and are explored subsequently.

CHALLENGES OF MEASURING EMOTIONS

Understanding human emotions is an age-old psychological research topic [9]. A number of theories and methodologies has been accumulated over decades, informing as well as inspiring the work in the other fields, including UX. Nonetheless, divergent views on the nature of emotions prevail. For instance, an ongoing debate in emotion research is whether to conceptualize and measure emotions as distinct states (e.g. Paul Ekman's six basic emotions) or relative points along certain dimensions (i.e. arousal, valence and dominance) (cf. for pros and cons of each approach, see [8]). These arguments are linked to multiple definitions of emotion; there exist at least 92 instances [10].

The main contention on the measurability of emotions is related to the fact that emotions are ephemeral, dynamic and compound (i.e. two or more emotions arise simultaneously). Nevertheless, one can argue that everything can be measured, but whether it is meaningful and valid is another question [5]. The meaningfulness and validity of measures,

[1] https://uxdesign.cc/ux-is-new-ux-is-dead-fe65e7ddf131

[2] We do not delve into the debate on the nuances of the three terms emotion, affect and mood (see [8]), but use them interchangeably.

be they subjective or objective, hinge on the extent to which humans are able to identify and describe accurately their own emotions, and the extent to which emotions are associated with other behavioral and physiological responses.

In HCI, measuring emotions with psychophysiological data (e.g. galvanic skin conductance, heart rate) has a relatively shorter history than measuring emotions with self-reported scales (e.g. Self-Assessment Manikin) [11]. While it is recommendable to collect both objective and subjective emotional data to triangulate empirical results and thus strengthen conclusions, it is not uncommon that the two types of data are not significantly correlated [12].

UX OF WELLBEING CHATBOTS

To put the above arguments in context, we look at a salient example of Positive Computing - the use of conversational agents or chatbots to enable users to improve their mental health conditions such as loneliness, anxiety, depression, stress and others [13]. Chatbot is a computing technology that enables users to access services and information by interacting with computers through natural language in a way that emulates human-human dialogue. Recent years have witnessed a surge of chatbot usage in a variety of sectors, especially health and e-commerce. Chatbots encompass services such as voice-based assistants (e.g. Google Home, Alexa, Siri), open domain agents for social chatter (e.g. ALICE, Cleverbot, Google Meena), agents for prolonged training, coaching, or companionship (e.g. Woebot, Wysa, Replika), and customer services.

The interest in chatbots is boosted by rapid advances in AI and Machine Learning (ML) technologies [15]. Other fields critical for the development of chatbots include HCI, design, linguistics, communication science, philosophy, psychology, and sociology. The power of AI/ML for chatbots rests in the recent development in natural language processing (NLP), understanding (NLU) and generation (NLG). Specifically, user intents and emotions are automatically detected and learned by smart algorithms of the chatbot, which can thus provide personalized and appropriate responses to satisfy user preferences, needs and goals. In the case of deploying a chatbot for mental health that supports the user to sustain positive emotions, the chatbot should be able to recognize changes in emotions accurately and in a timely manner, thereby proposing emotion regulation strategies (e.g. meditation) to address efficiently any potential threats (e.g. rumination) that may lead to negative emotions.

Challenges

Building such a chatbot involves several design decisions. The literature suggests that multisensory and embodied chatbots with customizable avatars could be more effective than their mono-modal and non-embodied counterparts (e.g. [14]). How to evaluate the user experience of the wellbeing-driven chatbot? Some approaches and challenges are identified:

- *Data fusion and triangulation*: Self-reported emotional responses and psychophysiological (voice, facial expression, gesture, heart rate) data when deploying the chatbot can be captured. It is necessary to calibrate each type of sensory data to indicate specific emotional states or points along the emotional dimensions. Nonetheless, instead of momentary/episodic evaluation that would interrupt the experiential flow, retrospective evaluation can be employed where users are asked to provide emotion labels when viewing the recorded interaction (i.e. the method of cued-recall debriefing [11]). Note that results derived from different sources of data can be divergent or even contradictory. The challenge is how to interpret and reconcile such inconsistent results to provide relevant and coherent feedback to users.
- *Accuracy and trust*: Speech Emotion Recognition has made notable advances in the last decade [16], but the accuracy is not yet at the level that justifies the full-fledged use of such a chatbot, considering the sensitivity and potential risk. Inaccurate responses of the chatbot can severely undermine UX and trust in it [17]. How to best explain to users its limitations and calibrate their trust in it is a challenge to tackle.

Implications

Measuring UX of digital technologies, be they wellbeing-driven or not, involves quantifying emotions, which can be controversial. Nonetheless, self-reported questionnaires and physiological instruments are still useful tools to allow the estimation of a user's emotional profile as a result of interacting with a computing technology. In case of inconsistent findings or insignificant correlations between different sources of data, it is difficult to determine which one should be used as a reference. As the calibration of objective data relies on human-in-the-loop, it seems to infer that subjective data could be used for benchmarking. However, when resources permitted, additional qualitative data via interviews should be collected to help disambiguate the quantitative data.

The accuracy of the AI/ML-infused system in terms of classifying and predicting the construct of interest depends much on quality and volume of data available for training the underlying algorithmic models. Emotions are particularly challenging, given that humans cannot interpret emotional expressions reliably as we experience in human-human communications. As errors seem inevitable, what needs to sustain positive UX and trust is to increase the transparency of the system, providing users with information at a preferred level of granularity to enable them to appraise the system's responses and thus decide an appropriate course of action.

CONCLUDING REMARKS

Two decades ago usability evolved to UX, which is in turn evolving to another notion. It could be labelled as Service Design Experience, Positive Computing Experience or Human-AI Experience. Whichever new term the research community would embrace, the basic understanding of emotions needs to be further substantiated. If AI/ML models of a wellbeing-driven technological device were built upon erroneous assumptions of emotions, the users would suffer from very bad user experience, and, even worse, the device would do harm rather than help.

REFERENCES

1. Gómez-López, P., Simarro, F. M., & Bonal, M. T. L. (2019, June). Analysing the UX scope through its definitions. In *Proceedings of the XX International Conference on Human Computer Interaction* (pp. 1-4).
2. Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009, April). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 719-728).
3. Mirnig, A. G., Meschtscherjakov, A., Wurhofer, D., Meneweger, T., & Tscheligi, M. (2015, April). A formal analysis of the ISO 9241-210 definition of user experience. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 437-450).
4. Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & Information Technology, 25*(2), 91-97.
5. Law, E. L. C., Van Schaik, P., & Roto, V. (2014). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies,* 72(6), 526-541.
6. Roto, V., Lee, J. J., Lai-Chong Law, E., & Zimmerman, J. (2021, June). The overlaps and boundaries between service design and user experience design. In *Designing Interactive Systems Conference 2021* (pp. 1915-1926).
7. Calvo, R. A., & Peters, D. (2014). *Positive computing: technology for wellbeing and human potential*. MIT Press.
8. Ekkekakis, P. (2013). *The measurement of affect, mood, and emotion: A guide for health-behavioral research*. Cambridge University Press.
9. Robinson, M. D., Watkins, E. R., & Harmon-Jones, E. (Eds.). (2013). *Handbook of cognition and emotion*. Guilford Press.
10. Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion,* 5(4), 345-379.
11. Bruun, A., Law, E. L. C., Nielsen, T. D., & Heintz, M. (2021). Do you feel the same? On the robustness of cued-recall debriefing for user experience evaluation. *ACM Transactions on Computer-Human Interaction (TOCHI), 28*(4), 1-45.
12. Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies,* 65(4), 275-291.
13. Sweeney, C., Potts, C., Ennis, E., Bond, R. R., Mulvenna, M., O'Neill, S., ... & McTear, M. (2021). Can Chatbots Help Support a Person's Mental Health?: Perceptions and Views from Mental Healthcare Professionals and Experts. *ACM Transactions on Computing for Healthcare, 2*(3), 1-16.
14. Følstad, A., Araujo, T., Papadopoulos, S., Law, E. L. C., Granmo, O. C., Luger, E., & Brandtzaeg, P. B. (2020). *Chatbot research and design*. Springer International Publishing.
15. Adamopoulou, E., & Moussiades, L. (2020, June). An overview of chatbot technology. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 373-383). Springer, Cham.
16. Schuller, B. W. (2018). Speech emotion recognition: two decades in a nutshell, benchmarks, and ongoing trends. *Communications ACM, 61(5), 90–99.*
17. Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627-660.