UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Human control redressed: Comparing AI and human predictability in a real-effort task

Permalink

https://escholarship.org/uc/item/1bc0q4zw

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors

Kandul, Serhiy Micheli, Vincent Beck, Juliane <u>et al.</u>

Publication Date

2023

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

Human control redressed: Comparing AI and human predictability in a real-effort task

Serhiy Kandul University of Zurich, Zurich, Switzerland

Vincent Micheli University of Geneva, Geneva, Switzerland

Juliane Beck University of St. Gallen, St. Gallen, Switzerland

Thomas Burri University of St. Gallen, St. Gallen, Switzerland

Francois Fleuret University of Geneva, Geneva, Switzerland

Markus Kneer University of Zurich, Zurich, Switzerland

Markus Christen University of Zurich, Zurich, Switzerland

Abstract

Predictability is a prerequisite for effective human control of artificial intelligence (AI). For example, the inability to predict the malfunctioning of AI impedes timely human intervention. In this paper, we employ a computerized navigation task, a lunar lander game, to investigate empirically how AI's predictability compares to humans' predictability. To this end, we ask participants to guess whether the landings of a spaceship in this game performed by AI and humans will succeed. We show that humans are worse at predicting AI performance than at predicting human performance in this environment. Significantly, participants underestimate the differences in the relative predictability of AI and, at times, overestimate their prediction skills. We link the difference in predictability to differences in the approaches, i.e. different landing patterns, employed by AI and humans to succeed in the task. These results highlight important differences in perception of AI and human with implications for human-computer interaction.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.* ©2023 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).