SIT799 Human Aligned Artificial Intelligence

Credit Task 8.1: Guidelines for Human-AI Interaction

Advances in artificial intelligence are enabling developers to integrate a variety of AI capabilities into user-facing systems. As automated inferences are typically performed under uncertainty, often producing false positives and false negatives, AI-infused systems may demonstrate unpredictable behaviours. The authors have proposed and evaluated 18 generally applicable design guidelines for human-AI interaction. Team members selected AI-infused products or features of their choice and looked for applications or violations of our initial set of design guidelines over a period. Participants found at least one application or violation of each of the proposed guidelines in each product category they tested, suggesting broad evidence of the guidelines' relevance.

The primary goal of the article is to evaluate the design guidelines rather than to evaluate an interface, later modified the heuristic evaluation by asking evaluators to attempt to identify both applications and violations of the proposed guidelines in an interface and to reflect on the guidelines themselves during the evaluation.

The phase 2 produced a set of 18 guidelines that closely match the guidelines in Table 1 mentioned in the article.

In the following sections a user study that tested these 18 guidelines and a subsequent expert validation of the guidelines that were slightly rephrased after the user study. A user study with 49 HCI practitioners was conducted to 1) understand the guidelines' applicability across a variety of products; and 2) get feedback about the guidelines' clarity. For each guideline, the participants were asked to first determine if the guideline "Does not apply" to their assigned feature and, if not, to explain why. After completing the evaluation, participants submitted their examples and ratings and were asked to fill in a final questionnaire which asked them to rate each guideline on a 5-point semantic differential scale from "Very confusing" to "Very clear" and provide any additional comments about the guidelines.

The questionnaires asked participants to rate the extent to which an example is illustrative of a guideline and the clarity of each guideline's wording on Likert scales.

Two different participants identified the same application of Guideline 1 for an activity tracker: "This guideline is applied in the activity summary view, where it shows a summary of my 'move', 'exercise' and 'stand' metrics." and "Displays all the metrics that it tracks and explains how. Metrics include movement metrics such as steps, distance travelled..." The 55 duplications were removed from the analysis. Instances where participants used "Does not apply" to indicate that they could not find examples of a guideline rather than to indicate that the guideline is not relevant for the product they were testing, as we intended by this designation.

The authors reviewed the guidelines to answer two important questions i.e. 1) Are the guidelines relevant? i.e. whether they could identify examples of each guideline across a variety of products and features? 2) Are the guidelines clear? i.e. whether participants understand and differentiate among them?

Participants found at least one application or violation of each of our guidelines in each product category they tested, suggesting broad evidence of the guidelines' relevance.

Examination of the instances revealed that in some cases participants firmly believed these guidelines were not relevant for the products they were testing while other participants reported either applications or violations of these guidelines in those same product categories.

The guidance’s further needed clarification when errors were determined to be systematic, which had four or more instances confused with another guideline or having multiple participants making similar comments about clarity.

The guidelines mentioned in the article only begin to touch on topics of fairness and broader ethical considerations. There is need of further exploration and continuous research in this area as the applications and integrations of AI driven structures is growing exponentially every day,