SIT 799

Task 8.1: Guidelines for Human-AI Interaction

Introduction

In [1], the authors have proposed 18 design guidelines which can be used for developing AI based software systems. This report summarised these guiding principles and explained their importance.

AI system design guidelines

The proposed guidelines are categorised by various states of interaction between the AI system and the user. The following section interprets the guidelines as presented in [1].

Initial State

> **G1: Provide user a clear understanding of AI capabilities.** The paper provides the example of an AI health app that tracks metrics like number of steps, duration of exercise etc.

> **G2: Help the user understand the limitations of the system.** The paper uses a music app example using a suggestive message followed by displaying the potential song preferences.

Interaction State

> **G3**: **Real time analytics capabilities to the user.** The paper uses a navigation application example to underscore providing updates based on user's geographical location.

> **G4: Providing relevant information to the user.** The paper uses a web search example where timings for nearby shows are displayed when user searches for a movie.

> **G5: Mimic social interactions for smoother experience.** The paper uses an example of voice assistant using words which mimic a human conversation.

> **G6: Ensure that the AI system does not exhibit biases.** The paper shows the example of an autocomplete feature which suggests pronouns catering to both genders.

Error State

> **G7: Make the AI system easy to access.** The paper uses example of wake-up word to activate voice assistant application, this shows the ease of access.

> **G8: AI system should be easy to dismiss.** The guideline uses the example of e-commerce application's suggested products which can be ignored if the user does not want to engage.

> **G9: AI system should be open to corrective feedback.** AI system must have a corrective feedback mechanism or override pipeline to allow for the user to take over.

> **G10: Prioritize user experience over AI system's output.** This is explained using Autocomplete app, where the app suggests a collection of words instead of enforcing the output.

> **G11: Encourage transparency.** The paper uses navigation example to highlight the importance of making the user understand the reasons behind output of an AI system.

Temporal feedback state

> **G12: Remember user responses.** The paper uses web search example to highlight how contextual information can be extracted when user responses are preserved.

G13: Learn from user actions. The AI system should map the user responses to improve future behaviour and provide a personalised experience.

G14: Limit the update steps to provide smoother. The application should prioritize consistency. This is explained by the example of a music app, where the playlist is updated but the UX is preserved.

G15: Provide the option of user override. The AI system should prioritize user response to provide better UX, e.g.: the AI system of an email app might miss marking an email important, but the user should have the ability to override.

G16: Prompt user for changes in future behaviour of the system. The paper explains this by highlighting changes to future ad suggestions when user hides a particular ad.

G17: Allow customization capabilities. The paper uses location-based photo grouping feature to highlight how AI system can be customized based on user preference.

G18: Provide feature changes information to the user.

Conclusion

AI based software development has challenged traditional software development design paradigms. This report provides comprehensive interpretation of guidelines mapped to real world scenarios which can be used to design a user-friendly AI system.

References

1. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.