Applied Psychology Seminar M1308.001100-001 Human-AI Interaction

Seoul National University, Fall 2022

Instructor: Dr. Sowon Hahn
Email: swhahn@snu.ac.kr, Office: 16-M315 Class Time: Wednesday 14:00-16:50

"We need to switch from a technology-centric view of the world to a people-centric one. We should start with people's abilities and create technology that enhances people's capabilities: Why are we doing it backwards? We have our priorities completely wrong." – Don Norman, *Why bad technology dominates our lives.*

Course Description:
Advances in artificial intelligence enable a variety of AI-infused systems. The Human-Computer Interaction (HCI) communities have developed principles of human-centered design for several decades. With increased automation and decision-making capabilities in AI-infused systems, it is crucial to understand human behavior interacting with intelligent systems. In this course, we will review topics in psychology, HCI, and AI technologies that are relevant to developing human-centered systems designs.

Class discussion Contribution to class will be worth 50% of your final grade. Students will be required to generate 2-3 discussion questions per article (it should be a full description of the issue instead of a simple question) and major issues from each article prior to each class. Students will also take responsibility for leading the discussion. Leading the discussion will entail the followings: 1) summarizing the key points to be gleaned from the articles, 2) using the discussion questions posted by other students to facilitate in-depth discussion. Leading the discussion (or we can call it presentation) will be worth 20% of your grade.

Final Paper (Proposal) Constituting 30% of the final grade, students will write a paper of their chosen topic within the field Human-AI Interaction. The paper should have a proposal format, evaluating current body of research and proposing a new study. Papers should be double-spaced with 1-inch margins and 11-pt standard font, and are recommended to be about 8-10 pages long.

Schedule:

September 7: Perspectives on Human-AI Interaction

- Horvitz, E. (1999, May). Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 159-166).
- Shneiderman, B., & Maes, P. (1997). Direct manipulation vs. interface agents. *interactions*, *4*(6), 42-61.
- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-13).

September 14: Human-Robot Interaction

- Thrun, S. (2004). Toward a framework for human-robot interaction. *Human–Computer Interaction*, *19*(1-2), 9-24.
- Fong, T., Nourbakhsh, I., Dautenhahn, K. (2002). A survey of socially interactive robots: concepts, design and applications. Technical Report CMU-RI-TR-02-29, Robotics Institute, Carnegie Mellon University.

- Bartneck C. and Forlizzi J. (2004). "A design-centred framework for social human-robot interaction," RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759), pp. 591-594.

September 21: Fairness in ML

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. In *Ethics of Data and Analytics* (pp. 254-264). Auerbach Publications.
- Madaio, M. A., Stark, L., Wortman Vaughan, J., & Wallach, H. (2020, April). Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1-35.

September 28: Interpretability

- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018, October). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)* (pp. 80-89). IEEE.

October 5: AI Ethics

- Deng, B. (2015). Machine ethics: The robot's dilemma. *Nature News*, *523*(7558), 24.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.
- Wang, R., Harper, F.M., Zhu, H. (2020). Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems April 2020, Pages 1–14.

October 12: Social Robots

- Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities?. *International Journal of Human-Computer Studies*, *43*(2), 223-239.
- Sarrica, M., Brondi, S., & Fortunati, L. (2020). How many facets does a "social robot" have? A review of scientific and popular definitions online. Information Technology & People, 33(1), 1–21.
- Edwards, A., Edwards, C., & Gambino, A. (2020). The Social Pragmatics of Communication with Social Robots: Effects of Robot Message Design Logic in a Regulative Context. International Journal of Social Robotics, 12, 945–957.

October 19: Human-Centered Design,

- Zhu, H., Yu, B., Halfaker, A., & Terveen, L. (2018). Value-sensitive algorithm design: Method, case study, and lessons. *Proceedings of the ACM on human-computer interaction*, *2*(CSCW), 1-23.
- Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017, May). UX design innovation: Challenges for working with machine learning as a design material. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 278-288).
- Rasmussen, J. (1999). Ecological interface design for reliable human-machine systems. *The International Journal of Aviation Psychology*, *9*(3), 203-223.

October 26: Learning Systems

- Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in cognitive sciences*, *20*(7), 512-534.
- Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., ... & Sardanelli, F. (2021). AI applications to medical images: From machine learning to deep learning. *Physica Medica*, *83*, 9-24.
- Lee, Y. C., Yamashita, N., Huang, Y., & Fu, W. (2020, April). " I Hear You, I Feel You": encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-12).

November 2: Artificial Intelligence and Decision Making

- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63-71.
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577-586.
- Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, *61*(4), 66-83.

November 9: Autonomous Vehicles

- Bagnell, J. A. (2015). *An invitation to imitation*. Carnegie-Mellon Univ Pittsburgh Pa Robotics Inst.
- Seppelt, B. D., & Lee, J. D. (2019). Keeping the driver in the loop: Dynamic feedback to support appropriate use of imperfect vehicle control automation. *International Journal of Human-Computer Studies*, *125*, 66-80.
- Park, S. Y., Moore, D. J., & Sirkin, D. (2020, April). What a driver wants: User preferences in semi-autonomous vehicle decision-making. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

November 16: Machine Mind

- Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, *12*(1),13-24.
- Yamamoto, T. (2004). From humanoid embodiment to theory of mind. In *Embodied artificial intelligence* (pp. 202-218). Springer, Berlin, Heidelberg.

- Cominelli, L., Mazzei, D., & De Rossi, D. E. (2018). SEAI: Social emotional artificial intelligence based on Damasio's theory of mind. *Frontiers in Robotics and AI*, *5*, 6.

November 23: Human AI Communication

- Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Cambridge, UK: Cambridge university press.
- Bartneck, C., & Forlizzi, J. (2004, September). A design-centered framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759)* (pp. 591-594). IEEE.
- Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74-88.

November 30: Presentation

December 7: Presentation