Human-Computer Interaction approach with Empathic Conversational Agent and Computer Vision

First Author[1][0000−1111−2222−3333], Second Author[2,3][1111−2222−3333−4444], and Third Author[3][2222−−3333−4444−5555]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com
http://www.springer.com/gp/computer-science/lncs
[3] ABC Institute, Rupert-Karls-University Heidelberg, Heidelberg, Germany
{abc,lncs}@uni-heidelberg.de

Abstract. The abstract should briefly summarize the contents of the paper in 150–250 words.

Keywords: First keyword · Second keyword · Another keyword.

1 Introduction

Empathy, which entails comprehending and sharing others' emotions to forge emotional connections, is vital for human relationships. Similarly, in Human-Computer Interaction (HCI), empathy is crucial in ensuring more realistic, improved, convenient and meaningful interactions. However, the typical HCI, aimed at tailoring computer systems to meet the specific needs and preferences of individuals, still lacks the users' emotional state, therefore losing crucial information during these interactions [5]. Recent Artificial Intelligence (AI) techniques, such as Emotion Recognition (ER) and empathic Conversational Agents (CAs), when integrated with HCI allow for continuous understanding of the user's emotions throughout interactions and empathically providing responses, greatly contribute to an increase in the quality and deepness of interactions between humans and computers, improving the user's overall experience [8].

Artificial Intelligence (AI) encompasses various techniques and methodologies aimed at enabling machines to perform tasks that typically require human intelligence, whereas Deep Learning (DL) stands out as a specialized approach relying on Artificial Neural Networks (ANN) to process unstructured data (including images, voice, videos, and text, among others). ER, being a recent application of AI combined with DL, involves detecting human emotions through various modalities ranging from facial features, gestures, and poses, to speech and text captured through continuous interactions with the user [1]. CAs consist of computer programs designed to simulate human-like conversation and engage in interactions with users through natural language using various techniques,

including natural language processing (NLP), ML, and DL, to understand user input, interpret context, and generate appropriate responses.

Due to the immense potential of ER and CA, individually, and the numerous benefits provided when integrated with HCI, this study offers a comprehensive guide covering ER modalities and key design and functionality aspects of a CA, furthermore reviewing widely adopted datasets and methodologies. Lastly, proposing an innovative HCI approach to ensure more realistic and meaningful interactions by leveraging HCI in conjunction with ER techniques and an empathic CA.

The primary findings of this study can be summarized as follows:

- Detailed guide on how DL impacts HCI nowadays;
- Performed a literature review regarding ER and CAs;
- Explored the main methods and datasets used for ER and CAs;
- Proposal of a taxonomy encompassing HCI, DL, CA, and ER;
- Proposal of an architecture of an HCI approach aided with ER, through Computer Vision (CV) and Sentiment Analysis (SA), and emphatic CA.

This research is organized into six sections. Section 2 presents the main concepts behind DL, ER, and CA. Section 3 details the most used datasets to train, validate, and evaluate ER and CA algorithms. Section 4 introduces and discusses the core AI algorithms used nowadays to build ER systems and CAs, while Section 5 presents the architecture, features, and characteristics of our proposed solution for HCI aided with ER and an empathic CA. Lastly, Section 6 introduces the challenges and directions for future work and the conclusions.

2 Background

The evolution of technology has led to an increased demand for advanced HCI. This is no longer confined to basic hardware-based communication but now encompasses more sophisticated techniques that are gradually becoming a part of everyday life. These include voice recognition, face recognition, and gesture recognition, which are essential for facilitating more natural and intuitive interactions between humans and computers [2]. Furthermore, the ability of machines to perceive and interact with humans, whether in physical or virtual environments, needs an understanding of human motion. This understanding must account for physical constraints, such as muscle torque and gravity, as well as the intentions behind movements, making motion modeling a highly complex [7].

Deep learning, a subset of machine learning, has been crucial in advancing the capabilities of HCI. By utilizing computational models with multiple processing layers, deep learning makes easier the learning of data representations at several levels of abstraction. This has significantly enhanced performance in domains like speech recognition, visual object recognition, object detection, and even fields like drug discovery and genomics. The focus of this paper will be on supervised learning, a predominant form of deep learning. Supervised learning involves training a system with a labeled dataset, where the system learns to

map inputs to outputs based on example input-output pairs. During training, the system iteratively adjusts its parameters to minimize the difference between its outputs and the desired outputs. This process involves a high number of adjustable parameters, or weights, which define the system's input-output function [6, **?**].

The implementation of deep learning in supervised learning models follows a structured process. This encompasses phases like data collection, where the importance of data quality cannot be overstated, data preprocessing to enhance data quality, training the model, optimization based on validation, and testing [9]. In recent years, Facial Emotion Recognition (FER) systems have exemplified the efficacy of Artificial Neural Networks (ANNs) over traditional machine learning methods. ANNs have demonstrated superior performance in detecting and recognizing emotions in a subject-independent manner, analyzing training data from various individuals. This approach has opened new opportunities in fields like healthcare, security, business, education, and manufacturing [4, **?**,**?**].

In the context of computer vision, neural networks have been particularly successful in image classification tasks, including face identification and facial emotion recognition. These technologies are not only used in surveillance systems but also in medical diagnostics and user-interactive applications. Different neural network architectures have been employed to meet the specific requirements of these tasks, including the use of pre-trained networks for classification, feature extraction, and transfer learning. Transfer learning, in particular, involves adjusting and reusing layers of a neural network trained on one dataset to work with a new dataset, demonstrating the versatility and adaptability of neural networks in various applications [3, **?**].

3 Datasets

4 Methods

5 Proposed solution

6 Future work and conclusions

Acknowledgments. A bold run-in heading in small font size at the end of the paper is used for general acknowledgments, for example: This study was funded by X (grant number Y).

Disclosure of Interests. It is now necessary to declare any competing interests or to specifically state that the authors have no competing interests. Please place the statement with a bold run-in heading in small font size beneath the (optional) acknowledgments[4], for example: The authors have no competing interests to declare

[4] If EquinOCS, our proceedings submission system, is used, then the disclaimer can be provided directly in the system.

that are relevant to the content of this article. Or: Author A has received research grants from Company W. Author B has received a speaker honorarium from Company X and owns stock in Company Y. Author C is a member of committee Z.

References

1. Alrowais, F., Negm, N., Khalid, M., Almalki, N., Marzouk, R., Mohamed, A., Al Duhayyim, M., Alneil, A.A.: Modified Earthworm Optimization With Deep Learning Assisted Emotion Recognition for Human Computer Interface. IEEE Access **11**, 35089–35096 (2023). https://doi.org/10.1109/ACCESS.2023.3264260, https://ieeexplore.ieee.org/document/10091537/
2. Alrowais, F., Negm, N., Khalid, M., Almalki, N., Marzouk, R., Mohamed, A., Duhayyim, M.A., Alneil, A.A.: Modified earthworm optimization with deep learning assisted emotion recognition for human computer interface. IEEE Access **11**, 35089–35096 (2023). https://doi.org/10.1109/ACCESS.2023.3264260
3. Cîrneanu, A.L., Popescu, D., Iordache, D.: New trends in emotion recognition using image analysis by neural networks, a systematic review. Sensors 2023, Vol. 23, Page 7092 **23**, 7092 (8 2023). https://doi.org/10.3390/S23167092, https://www.mdpi.com/1424-8220/23/16/7092/htm https://www.mdpi.com/1424-8220/23/16/7092
4. Giannopoulos, P., Perikos, I., Hatzilygeroudis, I.: Deep learning approaches for facial emotion recognition: A case study on fer-2013. Smart Innovation, Systems and Technologies **85**, 1–16 (2018). https://doi.org/10.1007/978-3-319-66790-4_1/COVER, https://link.springer.com/chapter/10.1007/978-3-319-66790-4_1
5. Jaiswal, A., Krishnama Raju, A., Deb, S.: Facial Emotion Detection Using Deep Learning. In: 2020 International Conference for Emerging Technology (INCET). pp. 1–5 (Jun 2020). https://doi.org/10.1109/INCET49848.2020.9154121, https://ieeexplore.ieee.org/document/9154121
6. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 2015 521:7553 **521**, 436–444 (5 2015). https://doi.org/10.1038/nature14539, https://www.nature.com/articles/nature14539
7. Martinez, J., Black, M.J., Romero, J.: On human motion prediction using recurrent neural networks. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 **2017-January**, 4674–4683 (11 2017). https://doi.org/10.1109/CVPR.2017.497
8. Santos, B.S., Júnior, M.C., Nunes, M.A.S.N.: Approaches for Generating Empathy: A Systematic Mapping. In: Latifi, S. (ed.) Information Technology - New Generations. pp. 715–722. Advances in Intelligent Systems and Computing, Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-54978-1_89
9. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural Networks **61**, 85–117 (1 2015). https://doi.org/10.1016/J.NEUNET.2014.09.003