

Adaptive User-centered Neuro-symbolic Learning for Multimodal Interaction with Autonomous Systems

Amr Gomaa^{1,2}, Michael Feld¹

¹German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany

²Saarland Informatics Campus, Saarland University, Saarbrücken, Germany

Abstract

Recent advances in deep learning and data-driven approaches have facilitated the perception and comprehension of objects and their environments in a perceptual subsymbolic manner. Consequently, these autonomous systems can now perform tasks such as object detection, sensor data fusion, and language understanding. However, there is an increasing demand to further enhance these systems to attain a more conceptual and symbolic understanding of objects and their environments and acquire the underlying reasoning behind the learned tasks. Achieving this level of powerful artificial intelligence necessitates considering both explicit teachings provided by humans (e.g., describing a situation or explaining how to act) and implicit teaching obtained through observing human behavior (e.g., through the system's sensors). Hence, it is imperative to incorporate symbolic and subsymbolic learning approaches to support implicit and explicit interaction models. This integration enables the system to achieve multimodal input and output capabilities. In this extended abstract, we argue for considering these input types, along with human-in-the-loop and incremental learning techniques, to advance the field of artificial intelligence and enable autonomous systems to emulate human learning. We propose several hypotheses and design guidelines aimed at achieving this objective.

Keywords

Human-Centered Artificial Intelligence, Multimodal Interaction, Adaptive Models, Personalization

1. Introduction and Related Work

Human-centered artificial intelligence (HCAI) is an exciting new area of research that is attracting increasing attention from researchers of both artificial intelligence (AI) and human-computer interaction (HCI) [1, 2, 3, 4]. Despite the significant progress that has been made in developing autonomous systems, these systems still rely heavily on human operators, whether local or remote, to step in and assist or take control in situations where the system is unable to proceed. This highlights the need for HCAI techniques to promote trust, control, and reliability between users and machines [4]. However, developing and implementing these concepts remains a challenging and complex task [2]. As a result, there is still much room for improvement and further research in this field [3]. Several approaches have proposed ways to insert human knowledge into neural networks as a way of initialization, to guide the refinement of the network, and to extract symbolic information from the network [5, 6]. More recent attempts have

HITLAML'23: Human-in-the-Loop Applied Machine Learning, September 04–06, 2023, Belval, Luxembourg


✉ amr.gomaa@dfki.de (A. Gomaa); michael.feld@dfki.de (M. Feld)

🌐 <https://amrgomaaelhady.github.io/> (A. Gomaa)

🆔 0000-0003-0955-3181 (A. Gomaa); 0000-0001-6755-5287 (M. Feld)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

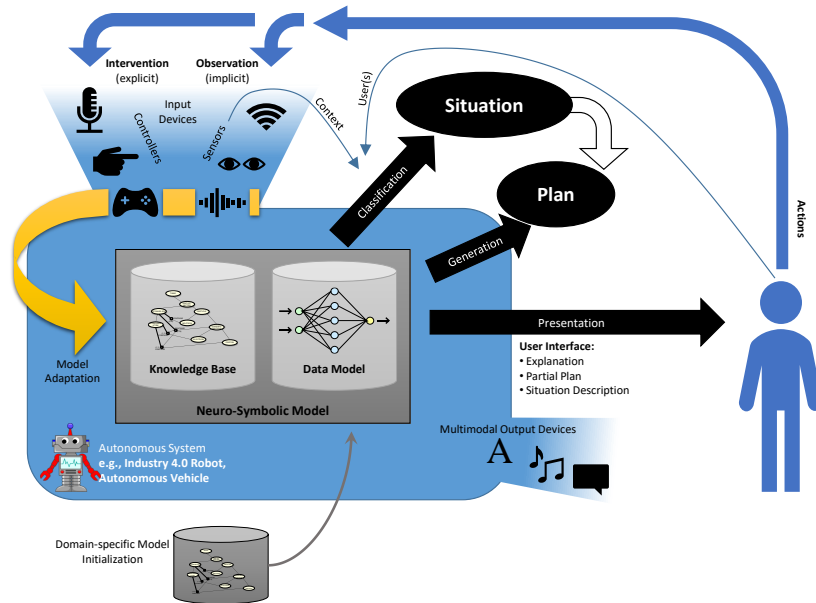


Figure 1: Overview of the envisioned user-centered neuro-symbolic human-in-the-loop learning system. The novice user and the learning agent (e.g., robot or autonomous vehicle) are in a continuous feedback loop, starting with the user’s demonstrations, then judging the agent’s output and providing support through thorough feedback.

tried to combine deep learning with knowledge bases in joint models (e.g., for construction and population) [7, 8]. Some work has focused on integrating neural networks with classical planning by mapping subsymbolic input to a symbolic one, which automatic planners can use [9]. Others have used Logic Tensor Networks to enable learning from noisy data in the presence of logical constraints by combining low-level features with high-level concepts [10, 11]. Other approaches include psychologically inspired cognitive architectures by having a goal-directed organizational hierarchy with parallel subsymbolic algorithms running in the lower levels and symbolic ones running serially in the higher levels [12]. Thus, we suggest that future work should focus on building autonomous systems that can learn and adapt to new situations, such as new classes, domains, or tasks [13, 6]. This will require shifting the focus from data-driven learning to interactive learning or human-in-the-loop learning, where the human plays a crucial role in supporting the system’s learning process. The proposed research concept focuses on developing adaptive and personalized approaches for human-in-the-loop learning that will enhance system performance and promote trust toward a reliable and controllable HCAI, as highlighted in Figure 1.

2. Approach

We propose the following research questions¹ as guidelines for future research on human-centered artificial intelligence. We focus on three factors: *Input features* (i.e., *Agent World View*), *Underlying design aspects* (i.e., *Multimodal interaction*), and *Learning method* (i.e., *Neuro-symbolic Adaptation and Continuous Learning*).

- **Agent World View (RQ1):** Which features of the agent (i.e., autonomous system) and the context (i.e., human behavior) can be used to detect and classify user interaction situations, and which devices are available to provide them efficiently (e.g., investigating user behavior as in [14])?

Given the multitude of sensors available for an autonomous system, possibly dynamic and not permanently available, a specific question will be to select the right level of granularity and fusion at which it can be combined with symbolic knowledge. This involves merging the available context information, both from sensors and world knowledge, combined with implicit user input [15, 16], to characterize the situation in a structured way. For example, in an industry scenario, a worker’s current task and the available robots would provide such input. In an autonomous vehicle scenario, knowledge about other passengers may help interpret the user’s goals and possible interaction. Based on available plans and solutions, a system has to estimate the success of a particular solution.

- **Multimodal Interaction (RQ2):** What aspects of system and interface design can be utilized of the given modalities in terms of fusion techniques, temporal dependencies, and learning models to achieve optimal performance (e.g., reference detection as in [17] and estimation of mental workload in [18, 19])?

To achieve an end-to-end multimodal fusion framework, it is vital to exhaustively investigate the interaction between the given modalities in terms of performance, timing, user behavior, and fusion techniques. While well-established, widely used data fusion approaches, such as late- and early-fusion approaches, are utilized here, more novel and empirical hybrid approaches should also be considered that combine heuristics with learning-based data fusion to achieve optimum performance. Additionally, there exists a timing dependency (e.g., modalities’ relative onset) between the modalities that the system can exploit. Thus, the time frames can be analyzed separately with no connection, or a pattern could be learned from intra- (within the modality) and inter- (among the modalities) dependencies.

- **Neuro-symbolic Adaptation and Continual Learning (RQ3):** How can the system adapt to the performance of user-specific tasks [17, 19]? How can the system be designed to continuously gather feedback from the user (both implicitly and explicitly) to guarantee constant development and enhancement of the underlying algorithms? How would that affect the system’s reliability and user trust?

Adaptation can be achieved at the architecture level using incremental learning [20]. Transfer learning (i.e., naive fine tuning) faces several challenges such as forgetting previously learned information (i.e., catastrophic forgetting), ever-changing features (i.e., concept shift), and how fast a model should be adapted (i.e., stability-plasticity dilemma).

¹Full paper presented at AI&HCI Workshop at ICML2023 and in-proceedings of ICMI2023 Blue Sky Papers.

Some solutions have been proposed for each of these challenges [21, 22, 6]. For continuous learning, there is a focus on increasing the number of classes a neural network can predict, expanding datasets, and exploring the influence of update intervals and batch sizes used for adaptation [23, 13]. To adapt an initial model to a different domain, we find suitable methods in the domain of incremental learning [24, 25, 26].

3. Conclusion

In conclusion, while designing user-specific interfaces is a complex and multifaceted process involving various considerations that this work cannot entirely describe, our position paper examines several essential aspects to facilitate this design process. Specifically, we discuss adapting learning models, including incremental and transfer learning, to enable personalized interaction with the system. This work also emphasizes the importance of system engineering considerations, such as real-time processing and system robustness, to ensure that user-specific interfaces are reliable and trustworthy. This paper highlights important considerations for future studies focused on human-centered artificial intelligence and trustworthy interfaces. In particular, we emphasize the importance of continuous learning and hybrid learning approaches to enable user-centered design that enhances the user experience. By following these guidelines, researchers can develop personalized and adaptive interfaces that respond to individual users' needs and behaviors, ultimately improving their satisfaction and engagement with the system. Furthermore, future research in this area should focus on developing frameworks and methodologies to assess the effectiveness of user-specific interfaces and explore the ethical and societal implications of these technologies.

Acknowledgments

This work is partially funded by the German Ministry of Education and Research (BMBF) under the TeachTAM project (Grant Number: 01IS17043) and the CAMELOT project (Grant Number: 01IW20008).

References

- [1] W. Xu, Toward human-centered ai: a perspective from human-computer interaction, *interactions* 26 (2019) 42–46.
- [2] A. Nowak, P. Lukowicz, P. Horodecki, Assessing artificial intelligence for humanity: Will ai be the our biggest ever advance? or the biggest threat [opinion], *IEEE Technology and Society Magazine* 37 (2018) 26–34.
- [3] J. J. Bryson, A. Theodorou, *How Society Can Maintain Human-Centric Artificial Intelligence*, Springer Singapore, Singapore, 2019, pp. 305–323.
- [4] B. Shneiderman, Human-centered artificial intelligence: Reliable, safe & trustworthy, *International Journal of Human-Computer Interaction* 36 (2020) 495–504.
- [5] J. W. Shavlik, Combining symbolic and neural learning, *Machine Learning* 14 (1994) 321–331. URL: <http://link.springer.com/10.1007/BF00993982>. doi:10.1007/BF00993982.

- [6] L. Von Rueden, S. Mayer, J. Garcke, C. Bauckhage, J. Schuecker, Informed machine learning—towards a taxonomy of explicit integration of knowledge into machine learning, *Learning* 18 (2019) 19–20.
- [7] A. Ratner, C. Ré, Knowledge base construction in the machine-learning era, *Queue* 16 (2018) 50:79–50:90. URL: <http://doi.acm.org/10.1145/3236386.3243045>. doi:10.1145/3236386.3243045.
- [8] H. Adel, Deep learning methods for knowledge base population, Ph.D. thesis, LMU, 2018.
- [9] M. Asai, A. Fukunaga, Classical planning in deep latent space: Bridging the subsymbolic-symbolic boundary, in: *Proceedings of the Conference on Artificial Intelligence (AAAI'18)*, AAAI Press, 2018, pp. 6094–6101.
- [10] L. Serafini, A. d. Garcez, Logic tensor networks: Deep learning and logical reasoning from data and knowledge, *arXiv preprint arXiv:1606.04422* (2016).
- [11] I. Donadello, L. Serafini, A. d'Avila Garcez, Logic Tensor Networks for Semantic Image Interpretation, in: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'17)*, IJCAI Organization, 2017, pp. 1596–1602. URL: <https://www.ijcai.org/proceedings/2017/221>. doi:10.24963/ijcai.2017/221.
- [12] T. D. Kelley, Developing a Psychologically Inspired Cognitive Architecture for Robotic Control: The Symbolic and Subsymbolic Robotic Intelligence Control System (SS-RICS), *International Journal of Advanced Robotic Systems* 3 (2006) 219–222. URL: <https://doi.org/10.5772/5736>. doi:10.5772/5736.
- [13] G. M. Van de Ven, A. S. Tolias, Three scenarios for continual learning, *arXiv preprint arXiv:1904.07734* (2019).
- [14] A. Gomaa, G. Reyes, A. Alles, L. Rupp, M. Feld, Studying person-specific pointing and gaze behavior for multimodal referencing of outside objects from a moving vehicle, in: *Proceedings of the 22nd International Conference on Multimodal Interaction*, ACM, 2020, pp. 501–509.
- [15] W. B. Knox, P. Stone, Interactively shaping agents via human reinforcement: The tamer framework, in: *Proceedings of the fifth international conference on Knowledge capture*, 2009, pp. 9–16.
- [16] Y. Cui, Q. Zhang, B. Knox, A. Allievi, P. Stone, S. Niekum, The empathic framework for task learning from implicit human feedback, in: J. Kober, F. Ramos, C. Tomlin (Eds.), *Proceedings of the 2020 Conference on Robot Learning*, volume 155 of *Proceedings of Machine Learning Research*, PMLR, 2021, pp. 604–626. URL: <https://proceedings.mlr.press/v155/cui21a.html>.
- [17] A. Gomaa, G. Reyes, M. Feld, Ml-persref: A machine learning-based personalized multimodal fusion approach for referencing outside objects from a moving vehicle, in: *Proceedings of the 23rd International Conference on Multimodal Interaction*, ACM, New York, NY, USA, 2021, p. 318–327.
- [18] A. Gomaa, A. Alles, E. Meiser, L. H. Rupp, M. Molz, G. Reyes, What's on your mind? a mental and perceptual load estimation framework towards adaptive in-vehicle interaction while driving, in: *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2022, pp. 215–225.
- [19] E. Meiser, A. Alles, S. Selter, M. Molz, A. Gomaa, G. Reyes, In-vehicle interface adaptation to environment-induced cognitive workload, in: *Adjunct Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*,

2022, pp. 83–86.

- [20] A. Gepperth, B. Hammer, Incremental learning algorithms and applications, in: Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'16), ESSAN, 2016, pp. 357–368.
- [21] J. C. Schlimmer, R. H. Granger, Incremental learning from noisy data, *Machine Learning* 1 (1986) 317–354. URL: <http://link.springer.com/10.1007/BF00116895>. doi:10.1007/BF00116895.
- [22] R. Polikar, L. Upda, S. Upda, V. Honavar, Learn++: an incremental learning algorithm for supervised neural networks, *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)* 31 (2001) 497–508. URL: <http://ieeexplore.ieee.org/document/983933/>. doi:10.1109/5326.983933.
- [23] C. Käding, E. Rodner, A. Freytag, J. Denzler, Fine-tuning deep neural networks in continuous learning scenarios, in: Proceedings of the Asian Conference on Computer Vision (ACCV'16 Workshops), Springer, 2016, pp. 588–605.
- [24] M. Long, H. Zhu, J. Wang, M. I. Jordan, Deep transfer learning with joint adaptation networks, in: Proceedings of the International Conference on Machine Learning (ICML'17) - Volume 70, ACM, 2017, pp. 2208–2217.
- [25] L. Jie, T. Tommasi, B. Caputo, Multiclass transfer learning from unconstrained priors, in: Proceedings of the International Conference on Computer Vision (ICCV'17), IEEE, 2011, pp. 1863–1870.
- [26] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, T. Tuytelaars, A continual learning survey: Defying forgetting in classification tasks, *IEEE transactions on pattern analysis and machine intelligence* 44 (2021) 3366–3385.