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

Amr Gomaa [1 2]   **Michael Feld** [1]

Abstract

Recent advances in machine learning, particularly deep learning, have enabled autonomous systems to perceive and comprehend objects and their environments in a perceptual subsymbolic manner. These systems can now perform object detection, sensor data fusion, and language understanding tasks. However, there is a growing need to enhance these systems to understand objects and their environments more conceptually and symbolically. It is essential to consider both the explicit teaching provided by humans (e.g., describing a situation or explaining how to act) and the implicit teaching obtained by observing human behavior (e.g., through the system's sensors) to achieve this level of powerful artificial intelligence. Thus, the system must be designed with multimodal input and output capabilities to support implicit and explicit interaction models. In this position paper, we argue for considering both types of inputs, as well as human-in-the-loop and incremental learning techniques, for advancing the field of artificial intelligence and enabling autonomous systems to learn like humans. We propose several hypotheses and design guidelines and highlight a use case from related work to achieve this goal.

1. Introduction

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) (Xu, 2019; Nowak et al., 2018; Bryson & Theodorou, 2019; Shneiderman, 2020).

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 (Shneiderman, 2020). However, developing and implementing these concepts remains a challenging and complex task (Nowak et al., 2018). As a result, there is still much room for improvement and further research in this field (Bryson & Theodorou, 2019).

When it comes to multimodal interaction, a variety of approaches have been explored using early and late data fusion techniques (Dong et al., 2009; Karpov & Yusupov, 2018). For example, researchers have studied hand and gaze fusion techniques for interacting with screen-based indoor systems (Kim et al., 2017; Zhang et al., 2015). In the automotive domain, there has been a focus on controlling the vehicle and the infotainment system using touch-based approaches and multimodal combinations of hand gestures, gaze, and speech (Sezgin et al., 2009; Rümelin et al., 2013; Fujimura et al., 2013; Roider et al., 2017; Neßelrath et al., 2016). Furthermore, it is crucial to consider the impact of cognitive load on driver ability to perform when using these interfaces and how it affects driving performance (Fuller, 2005; Brown et al., 2020; Islam et al., 2020; Barua et al., 2020; Solovey et al., 2014), which emphasizes the significance of personalization. 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 (Van de Ven & Tolias, 2019; Von Rueden et al., 2019). 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. More specifically, we highlight multiple methods and techniques for learning-based adapted models utilizing transfer-of-learning and propose some new aspects
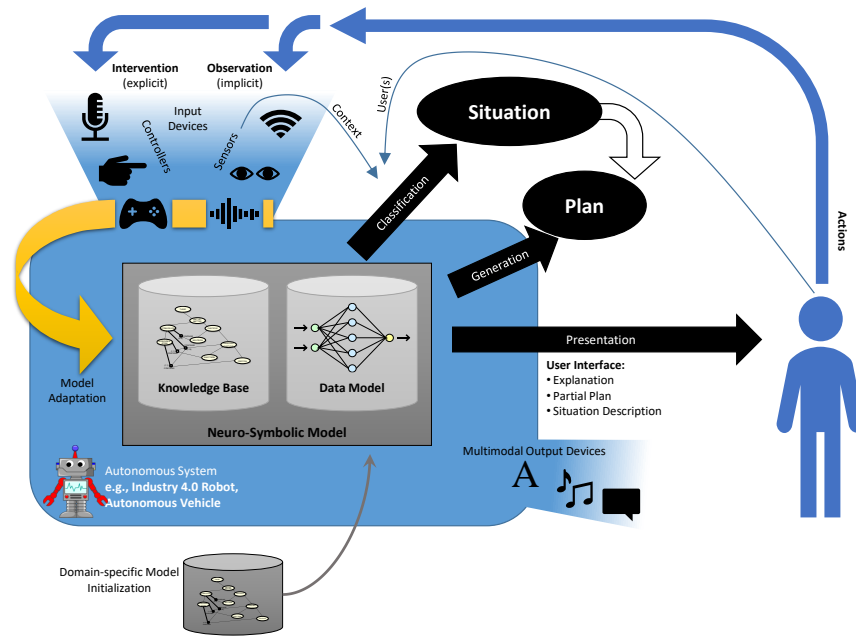
[1]German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany [2]Saarland Informatics Campus, Saarland University, Saarbrücken, Germany. Correspondence to: Amr Gomaa <amr.gomaa@dfki.de>.

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.

for continual learning for future work. Although these approaches apply to different domains, we focus on the automotive domain as an example of the rich work on driver personalization. More specifically, we demonstrate our suggestion on some of our previous work in the field of adaptive user interaction for the automotive domain (Gomaa et al., 2020; 2021; Gomaa, 2022; Gomaa et al., 2022; Meiser et al., 2022; Feld et al., 2019); however, the underlying learning techniques are valid for other domains as well.

2. Background and related work

Adaptive multimodal interaction combining speech, hand gestures, and gaze has been a topic of interest for the research community for the last 20 years in multiple domains, including robotics and automotive applications (Rogers et al., 2000; Hassel & Hagen, 2005; Janarthanam & Lemon, 2014; Manawadu et al., 2017; Zhang et al., 2015; Neverova et al., 2015; Gnjatović et al., 2012). Despite the previously discussed significant advances in the adaptation of multimodal interaction, a personalized user-centered approach is still lacking. Thus, an important goal and factor in the proposed research work is user-specific personalization through incremental learning techniques (Van de Ven & Tolias, 2019; Gepperth & Hammer, 2016). As an example, in the automotive domain, researchers attempted multimodal fusion approaches for in-vehicle object selection in multiple works (Roider & Gross, 2018; Aftab et al., 2020; Sezgin

et al., 2009). However, in-vehicle object referencing approaches do not generalize directly to outside-the-vehicle referencing, as the object's environment is static, limited, and in close proximity. Consequently, Moniri et al. (Moniri & Müller, 2012) studied the single task of outside-the-vehicle referencing from the passenger seat using pointing, head pose, and eye gaze. Similarly, Aftab et al. (Aftab et al., 2021) combined these modalities using a late fusion approach based on a neural network to reference objects from a stationary vehicle. While these approaches showed great promise, they still considered only a subsymbolic method for adaptation with a focus on data-driven approaches and did not consider user-specific behavior further.

Several approaches have proposed ways to insert human knowledge into neural networks as a way of initialization, to guide network refinement, and to extract symbolic information from the network (Shavlik, 1994; Von Rueden et al., 2019). More recent attempts have tried to combine deep learning with knowledge bases in joint models (e.g., for construction and population) (Ratner & Ré, 2018; Adel, 2018). Some work has focused on integrating neural networks with classical planning by mapping subsymbolic input to symbolic one, which automatic planners can use (Asai & Fukunaga, 2018). 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 (Serafini & Garcez, 2016; Donadello

et al., 2017). Other approaches include psychologically inspired cognitive architectures having a goal-directed organizational hierarchy with parallel subsymbolic algorithms running at the lower levels and symbolic ones running serially at the higher levels (Kelley, 2006). While subsymbolic learning methods, such as neural networks, have shown remarkable results in fields such as computer vision, NLP, and NLU, one problem they suffer from is a lack of explainability. On the other hand, while symbolic learning is "legible" by humans, it can lead to combinatorial growth that makes unfeasible solutions to complex problems (Büker, 1998). When combining both types of learning, it could be possible to obtain advantages while overcoming the disadvantages. For example, a teacher might teach a robot how to tidy up a table full of bottles in different stages. In the first stage, the teacher might guide the robot's arm, showing it how to clear one bottle from the table (subsymbolic learning by example). In the next stage, when the basic movements have been acquired, supervised learning can continue through verbal instructions (symbolic learning by instruction) (Grumbach, 1995).

3. Research Questions and Hypotheses

In line with the previous motivation and related work, the following research questions were developed to answer previous challenges from an abstract point of view while focusing 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)*. We envision these research questions as guidelines for future research on human-centered artificial intelligence.

- **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 (Gomaa et al., 2020))?

  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 (Knox & Stone, 2009; Cui et al., 2021), 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 (Gomaa et al., 2021) and estimation of mental workload in (Gomaa et al., 2022; Meiser et al., 2022))?

  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 (Gomaa et al., 2021; Meiser et al., 2022)? 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 (Gepperth & Hammer, 2016). 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). Some solutions have been proposed for each of these challenges (Schlimmer & Granger, 1986; Polikar et al., 2001; Von Rueden et al., 2019). 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 (Käding et al., 2016; Van de Ven & Tolias, 2019). To adapt an initial model to a different domain, we find suitable methods in the domain of incremental learning (Long et al., 2017; Jie et al., 2011; De Lange et al., 2021).
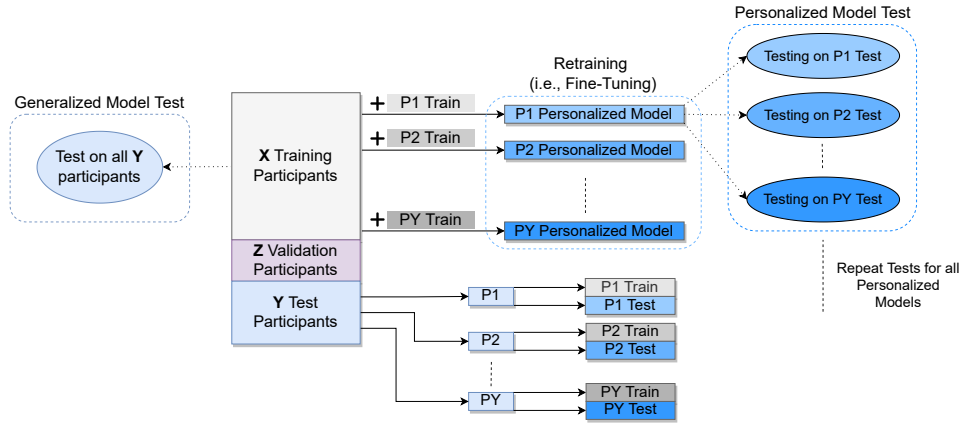
Figure 2. Proposed approach for model adaptation to generate personalized models through transfer and incremental learning techniques from (Gomaa, 2022).

4. Proposed Methodology

The proposed method follows the previously mentioned research questions to propose a multistage approach to reach an adaptive neuro-symbolic autonomous system with continuous user feedback (as seen in Figure 1).

The first stage of the proposed plan is to understand the variances in driver behavior when performing the multimodal referencing task as in (Gomaa et al., 2020; 2022; Rümelin et al., 2013). As an example, in the automotive domain, drivers perform different multimodal gestures to control the vehicle and query surrounding objects. These individual differences could be exploited by the system for personalization and adaptation through a user-centered design approach. Drivers could be clustered based on single-modality performance, and a switching mechanism could be applied within the overall system (Gomaa et al., 2020) to maximize overall performance (i.e., turn gaze detection off for user accompanied by wandering behavior of the eye and thus low accuracy of gaze detection). Furthermore, understanding the mental workload patterns of users could be exploited by the system and also to enhance its performance through model adaptation and personalization (Gomaa et al., 2022). The second stage would be creating an end-to-end learning-based multimodal fusion framework through constant and exhaustive monitoring of the users through system sensors. This is an initial step to automate the previously mentioned heuristics by the system (Gomaa et al., 2021) using hybrid learning where a pattern could be learned from intra- (within the modality) and inter- (among the modalities) dependencies. However, adaptation is an inherently continuous paradigm; thus, it is considered an ongoing process along the user observations and the multimodal fusion stages in drivers' categorization (i.e., clustering) and hybrid fusion approaches, respectively. Although model adaptation, in the previous context, is one alternative to the one-model-fits-all approach,

it still groups users in a particular model (i.e., cluster), constituting a many-models-fits-all approach. However, a more personalized approach would utilize transfer-of-learning and incremental learning techniques to eventually reach a single model or continuously adapting model per user. Figure 2 shows an approach to achieving these personalized models through incremental learning techniques. The data set is initially divided into training, validation, and test sets as in traditional learning approaches. The model is trained on **X** participants' data while the hyperparameters are chosen and validated on **Z** participants' data, and the final model is tested on **Y** participants' data. On the other hand, for the adaptation approach, each participant's data from the **Y** test set are further split (e.g., equally) into subtrain and subtest sets where the model is retrained and fine-tuned on the user-specific training data to produce personalized model weights that are optimized for this user. To assess the effect of this approach, the personalized model is tested on the same participant sub-test data and compared against other participants' sub-test data.

While the previous approaches optimize system performance based on current individual behavior, this behavior could change over time due to situational, emotional, or mental load variations and learning effects. Thus, a continuous learning approach is considered where the user can give feedback to the system implicitly (e.g., via dissatisfied looks or grunting as visual or auditory cues) or explicitly (e.g., repeating the given voice command). To achieve this goal, the study and data collection phase should include different variations in the situational and mental state for internal and external validity. Finally, situation-adapting learning techniques could be further utilized in this context, such as graph classification and node selection (e.g., Relational Graph Neural Networks (Jing et al., 2020)), learning from the driver's behavior (e.g., Efficient Learning from Demon-
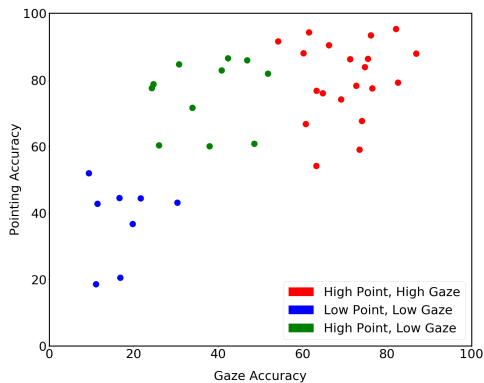
Figure 3. Clustering drivers' pointing and gaze behavior based on the system's perceived performance (i.e., referencing accuracy) from (Gomaa et al., 2020).
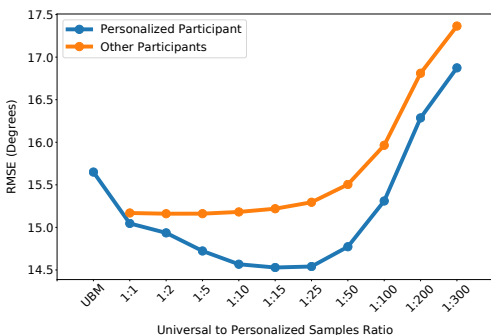

Figure 4. RMSE results comparing different personalized models among themselves (based on the sample weight of the user-specific data) and against the UBM (i.e., generalized model) from (Gomaa et al., 2021).

strations (Li et al., 2022)), and learning from the driver's feedback (e.g., Implicit Human Feedback Learner (Cui et al., 2021)).

Since the main focus of this work is on adaptation and user-specific personalization, Figure 3 and Figure 4 show examples of related work results focusing on the adaptation aspect of (Gomaa et al., 2020). Specifically, Figure 3 shows how drivers' referencing actions could be clustered based on pointing and gaze modality performance separately; then, each cluster is trained independently. Thus, each cluster model weight would be adapted to the cluster pointing- and gaze-specific accuracy. This resembles the hybrid-fusion approach discussed earlier. Similarly, Figure 4 highlights the results of the incremental learning personalization approach previously discussed in (Gomaa et al., 2021). It compares the personalized model subtest data against the average of the other non-personalized subtest data using the Root Mean Square Error (RMSE) metric. The figure also highlights further enhancement of this personalization approach; it was

noticed that adding the subtrain data of the personalized participant to the existing generalized model (also called Universal Background Model (UBM)) data with a 1:1 ratio is not the optimum solution due to its insignificant contribution size. Therefore, personalized participant subtrain data was emphasized (e.g., repeating the data multiple times), and its ratio increased for the training data **X** with a ratio of 1:2, 1:5, etc. until the optimum sample weight could be determined. While we focus precisely on these results for the referencing task, the methodology applies to any regression problem with a similar setup. Thus, it can be generalized to multiple sensors and multimodal platforms.

5. Conclusion

Although 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.

Acknowledgements

References

Adel, H. *Deep learning methods for knowledge base population.* PhD thesis, LMU, 2018.

Aftab, A. R., von der Beeck, M., and Feld, M. You have a point there: object selection inside an automobile using gaze, head pose and finger pointing. In *Proceedings of*

the 22nd International Conference on Multimodal Interaction, pp. 595–603. ACM, 2020. ISBN 9781450375818.

Aftab, A. R., Von Der Beeck, M., Rohrhirsch, S., Diotte, B., and Feld, M. Multimodal fusion using deep learning applied to driver's referencing of outside-vehicle objects. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1108–1115. IEEE, 2021.

Asai, M. and Fukunaga, A. Classical planning in deep latent space: Bridging the subsymbolic-symbolic boundary. In *Proceedings of the Conference on Artificial Intelligence (AAAI'18)*, pp. 6094–6101. AAAI Press, 2018.

Barua, S., Ahmed, M. U., and Begum, S. Towards intelligent data analytics: A case study in driver cognitive load classification. *Brain Sciences*, 10(8):1–19, 2020. ISSN 20763425.

Brown, E., R. Large, D., Limerick, H., and Burnett, G. Ultrahapticons: "Haptifying" Drivers' Mental Models to Transform Automotive Mid-Air Haptic Gesture Infotainment Interfaces. pp. 54–57, 2020. ISBN 9781450380669.

Bryson, J. J. and Theodorou, A. *How Society Can Maintain Human-Centric Artificial Intelligence*, pp. 305–323. Springer Singapore, Singapore, 2019. ISBN 978-981-13-7725-9.

Büker, U. Hybrid Object Models: Combining Symbolic and Subsymbolic Object Recognition Strategies. In *Proceedings of the International Conference on Information Systems, Analysis and Synthesis (ISAS'98)*, pp. 444–451. IIIS, 1998.

Cui, Y., Zhang, Q., Knox, B., Allievi, A., Stone, P., and Niekum, S. The empathic framework for task learning from implicit human feedback. In Kober, J., Ramos, F., and Tomlin, C. (eds.), *Proceedings of the 2020 Conference on Robot Learning*, volume 155 of *Proceedings of Machine Learning Research*, pp. 604–626. PMLR, 16–18 Nov 2021. URL https://proceedings.mlr.press/v155/cui21a.html.

De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G., and Tuytelaars, T. A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3366–3385, 2021.

Donadello, I., Serafini, L., and d'Avila Garcez, A. Logic Tensor Networks for Semantic Image Interpretation. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'17)*, pp. 1596–1602. IJCAI Organization, August 2017. ISBN 978-0-9992411-0-3. doi: 10.24963/ijcai.2017/221. URL https://www.ijcai.org/proceedings/2017/221.

Dong, J., Zhuang, D., Huang, Y., and Fu, J. Advances in multi-sensor data fusion: Algorithms and applications. *Sensors*, 9(10):7771–7784, 2009.

Feld, M., Neβelrath, R., and Schwartz, T. Software platforms and toolkits for building multimodal systems and applications. In *The Handbook of Multimodal-Multisensor Interfaces: Language Processing, Software, Commercialization, and Emerging Directions-Volume 3*, pp. 145–190. 2019.

Fujimura, K., Xu, L., Tran, C., Bhandari, R., and Ng-Thow-Hing, V. Driver queries using wheel-constrained finger pointing and 3-D head-up display visual feedback. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 56–62. ACM, 2013.

Fuller, R. Towards a general theory of driver behaviour. *Accident Analysis and Prevention*, 37(3):461–472, 5 2005. ISSN 00014575.

Gepperth, A. and Hammer, B. Incremental learning algorithms and applications. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'16)*, pp. 357–368. ESSAN, 2016.

Gnjatović, M., Tasevski, J., Nikolić, M., Mišković, D., Borovac, B., and Delić, V. Adaptive multimodal interaction with industrial robot. In *2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics*, pp. 329–333. IEEE, 2012.

Gomaa, A. Adaptive user-centered multimodal interaction towards reliable and trusted automotive interfaces. In *Proceedings of the 2022 International Conference on Multimodal Interaction*, pp. 690–695, 2022.

Gomaa, A., Reyes, G., Alles, A., Rupp, L., and Feld, M. 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*, pp. 501–509. ACM, 10 2020. ISBN 9781450375818.

Gomaa, A., Reyes, G., and Feld, M. 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*, pp. 318–327, New York, NY, USA, 2021. ACM. ISBN 9781450384810.

Gomaa, A., Alles, A., Meiser, E., Rupp, L. H., Molz, M., and Reyes, G. 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, pp. 215–225, 2022.

Grumbach, A. Learning at subsymbolic and symbolic levels. In Kappen, B. and Gielen, S. (eds.), *Neural Networks: Artificial Intelligence and Industrial Applications – Proceedings of the Annual SNN Symposium on Neural Networks*, pp. 91–94. Springer London, 1995. ISBN 978-1-4471-3087-1.

Hassel, L. and Hagen, E. Adaptation of an automotive dialogue system to users' expertise. In *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, pp. 222–226, 2005.

Islam, M. R., Barua, S., Ahmed, M. U., Begum, S., Aricò, P., Borghini, G., and Flumeri, G. D. A novel mutual information based feature set for drivers' mental workload evaluation using machine learning. *Brain Sciences*, 10 (8):1–23, 8 2020. ISSN 20763425.

Janarthanam, S. and Lemon, O. Adaptive generation in dialogue systems using dynamic user modeling. *Computational Linguistics*, 40(4):883–920, 2014.

Jie, L., Tommasi, T., and Caputo, B. Multiclass transfer learning from unconstrained priors. In *Proceedings of the International Conference on Computer Vision (ICCV'17)*, pp. 1863–1870. IEEE, 2011.

Jing, Y., Wang, J., Wang, W., Wang, L., and Tan, T. Relational graph neural network for situation recognition. *Pattern Recognition*, 108:107544, 2020. ISSN 0031-3203.

Käding, C., Rodner, E., Freytag, A., and Denzler, J. Fine-tuning deep neural networks in continuous learning scenarios. In *Proceedings of the Asian Conference on Computer Vision (ACCV'16 Workshops)*, pp. 588–605. Springer, 2016.

Karpov, A. and Yusupov, R. Multimodal interfaces of human–computer interaction. *Herald of the Russian Academy of Sciences*, 88(1):67–74, 2018.

Kelley, T. D. 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(3):219–222, September 2006. ISSN 1729-8814. doi: 10.5772/5736. URL https://doi.org/10.5772/5736.

Kim, H., Suh, K. H., and Lee, E. C. Multi-modal user interface combining eye tracking and hand gesture recognition. *Journal on Multimodal User Interfaces*, 11(3):241–250, 2017.

Knox, W. B. and Stone, P. Interactively shaping agents via human reinforcement: The tamer framework. In *Proceedings of the fifth international conference on Knowledge capture*, pp. 9–16, 2009.

Li, Q., Peng, Z., and Zhou, B. Efficient learning of safe driving policy via human-AI copilot optimization. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=0cgU-BZp2ky.

Long, M., Zhu, H., Wang, J., and Jordan, M. I. Deep transfer learning with joint adaptation networks. In *Proceedings of the International Conference on Machine Learning (ICML'17) - Volume 70*, pp. 2208–2217. ACM, 2017.

Manawadu, U. E., Kamezaki, M., Ishikawa, M., Kawano, T., and Sugano, S. A multimodal human-machine interface enabling situation-adaptive control inputs for highly automated vehicles. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1195–1200. IEEE, 2017.

Meiser, E., Alles, A., Selter, S., Molz, M., Gomaa, A., and Reyes, G. 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*, pp. 83–86, 2022.

Moniri, M. M. and Müller, C. Multimodal reference resolution for mobile spatial interaction in urban environments. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 241–248. ACM, 2012. ISBN 9781450317511.

Neßelrath, R., Moniri, M. M., and Feld, M. Combining speech, gaze, and micro-gestures for the multimodal control of in-car functions. In *Proceedings of the 12th International Conference on Intelligent Environments*, pp. 190–193. IEEE, 2016.

Neverova, N., Wolf, C., Taylor, G., and Nebout, F. Moddrop: adaptive multi-modal gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38 (8):1692–1706, 2015.

Nowak, A., Lukowicz, P., and Horodecki, P. Assessing artificial intelligence for humanity: Will ai be the our biggest ever advance? or the biggest threat [opinion]. *IEEE Technology and Society Magazine*, 37(4):26–34, 2018.

Polikar, R., Upda, L., Upda, S., and Honavar, V. Learn++: an incremental learning algorithm for supervised neural networks. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 31

(4):497–508, November 2001. ISSN 10946977. doi: 10.1109/5326.983933. URL http://ieeexplore.ieee.org/document/983933/.

Ratner, A. and Ré, C. Knowledge base construction in the machine-learning era. *Queue*, 16(3):50:79–50:90, June 2018. ISSN 1542-7730. doi: 10.1145/3236386.3243045. URL http://doi.acm.org/10.1145/3236386.3243045.

Rogers, S., Fiechter, C.-N., and Thompson, C. Adaptive user interfaces for automotive environments. In *Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No. 00TH8511)*, pp. 662–667. IEEE, 2000.

Roider, F. and Gross, T. I see your point: Integrating gaze to enhance pointing gesture accuracy while driving. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 351–358. ACM, 2018. ISBN 9781450359467.

Roider, F., Rümelin, S., Pfleging, B., and Gross, T. The effects of situational demands on gaze, speech and gesture input in the vehicle. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 94–102. ACM, 2017. ISBN 9781450351508.

Rümelin, S., Marouane, C., and Butz, A. Free-hand pointing for identification and interaction with distant objects. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 40–47. ACM, 2013.

Schlimmer, J. C. and Granger, R. H. Incremental learning from noisy data. *Machine Learning*, 1(3):317–354, 1986. ISSN 0885-6125, 1573-0565. doi: 10.1007/BF00116895. URL http://link.springer.com/10.1007/BF00116895.

Serafini, L. and Garcez, A. d. Logic tensor networks: Deep learning and logical reasoning from data and knowledge. *arXiv preprint arXiv:1606.04422*, 2016.

Sezgin, T. M., Davies, I., and Robinson, P. Multimodal inference for driver-vehicle interaction. In *Proceedings of the 11th International Conference on Multimodal Interfaces*, pp. 193–198. ACM, 2009. ISBN 9781605587721.

Shavlik, J. W. Combining symbolic and neural learning. *Machine Learning*, 14(3):321–331, March 1994. ISSN 0885-6125, 1573-0565. doi: 10.1007/BF00993982. URL http://link.springer.com/10.1007/BF00993982.

Shneiderman, B. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6):495–504, 2020.

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., and Mehler, B. Classifying driver workload using physiological and driving performance data: Two field studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 4057–4066, New York, NY, USA, 2014. ACM. ISBN 9781450324731.

Van de Ven, G. M. and Tolias, A. S. Three scenarios for continual learning. *arXiv preprint arXiv:1904.07734*, 2019.

Von Rueden, L., Mayer, S., Garcke, J., Bauckhage, C., and Schuecker, J. Informed machine learning–towards a taxonomy of explicit integration of knowledge into machine learning. *Learning*, 18:19–20, 2019.

Xu, W. Toward human-centered ai: a perspective from human-computer interaction. *interactions*, 26(4):42–46, 2019.

Zhang, Y., Stellmach, S., Sellen, A., and Blake, A. The costs and benefits of combining gaze and hand gestures for remote interaction. In *Human-Computer Interaction – INTERACT 2015*, pp. 570–577. Springer, 2015.