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

Amr Gomaa
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
Saarland Informatics Campus
Saarbrücken, Germany
amr.gomaa@dfki.de

Michael Feld
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
michael.feld@dfki.de

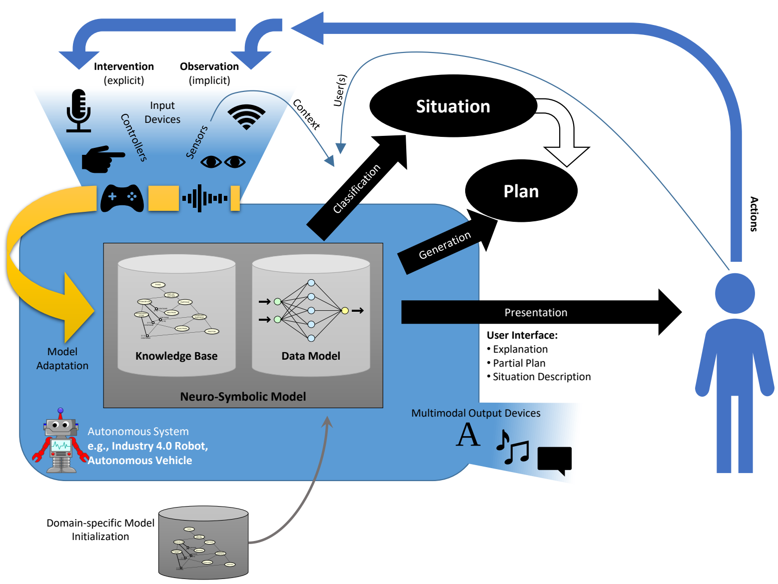


Figure 1: Overview of the envisioned user-centered neuro-symbolic human-in-the-loop learning system. The novice user and the learning agent (e.g., robot or autonomous vehicle) are in a continuous feedback loop, starting with user demonstrations, then judging the agent output and providing support through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

ABSTRACT

Recent advances in deep learning and data-driven approaches have facilitated the perception of objects and their environments in a perceptual subsymbolic manner. Thus, these autonomous systems can now perform object detection, sensor data fusion, and language understanding tasks. However, there is an increasing demand to further enhance these systems to attain a more conceptual and symbolic understanding of objects to 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., explaining how to act) and implicit teaching obtained through observing human behavior (e.g., through system 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 Blue Sky paper, 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.

CCS CONCEPTS

• **Human-centered computing** → **User centered design**; **HCI theory, concepts and models**; **HCI design and evaluation methods**; • **Computing methodologies** → **Neural networks**.

KEYWORDS

Human-Centered Artificial Intelligence; Multimodal Interaction; Data Fusion; Adaptive Models; Personalization

ACM Reference Format:

Amr Gomaa and Michael Feld. 2023. Towards Adaptive User-centered Neuro-symbolic Learning for Multimodal Interaction with Autonomous Systems.

ICMI '23, October 9–13, 2023, Paris, France

In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '23), October 9–13, 2023, Paris, France.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3577190.3616121

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) [7, 40, 51, 55]. 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 [51]. However, developing and implementing these concepts remains a challenging and complex task [40]. As a result, there is still much room for improvement and further research in this field [7]. When it comes to multimodal interaction, a variety of approaches have been explored using early and late data fusion techniques [12, 29]. For example, researchers have studied hand and gaze fusion techniques for interacting with screen-based indoor systems [31, 56]. 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 [14, 38, 45, 46, 49]. 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 [5, 6, 15, 24, 52], 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 [53, 54]. 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 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 [13, 18–21, 36]; 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 [17, 23, 25, 35, 39, 43, 56]. 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 [16, 53]. As an example, in the automotive domain, researchers attempted multimodal fusion approaches for in-vehicle object selection in multiple works [2, 44, 49]. 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. [37] studied the single task of outside-the-vehicle referencing from the passenger seat using pointing, head pose, and eye gaze. Similarly, Aftab et al. [3] 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 [50, 54]. More recent attempts have tried to combine deep learning with knowledge bases in joint models (e.g., for construction and population) [1, 42]. Some work has focused on integrating neural networks with classical planning by mapping subsymbolic input to symbolic one, which automatic planners can use [4]. 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 [11, 48]. 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 [30]. 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 [8]. 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) [22].

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

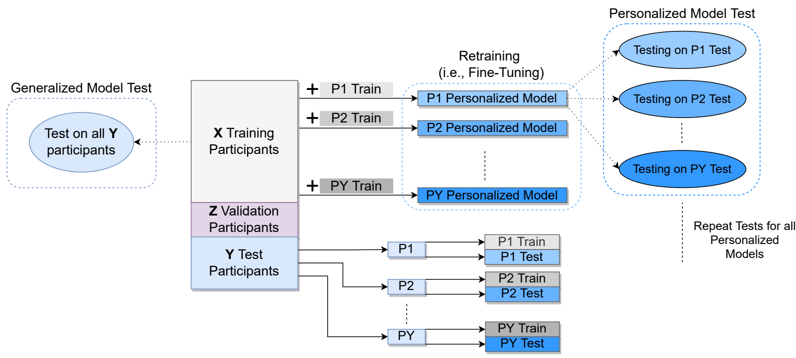


Figure 2: Proposed approach for model adaptation to generate personalized models through transfer and incremental learning techniques from [18].

used to detect and classify user interaction situations, and which devices are available to provide them efficiently (e.g., investigating user behavior as in [20])? Given the multitude of sensors available for an autonomous system that are 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 the sensors and world knowledge, combined with the implicit user input [9, 32], to characterize situations 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.

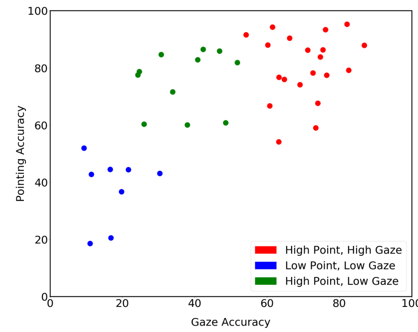
Multimodal Interaction (RQ2): What aspects of the system and interface design can be utilized of the given modalities in terms of fusion techniques, temporal dependencies, and learning models to achieve optimal performance (e.g., reference detection as in [21] and estimation of mental workload in [19, 36])? 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. Although well-established and 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- (across modalities) dependencies.

Neuro-symbolic Adaptation and Continual Learning (RQ3): How can the system adapt to the performance of user-specific tasks [21, 36]? How can the system be designed to continuously collect 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 [16]. 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 [41, 47, 54]. For continuous learning, there is a focus on increasing the number of classes that a neural network can predict, expanding datasets, and exploring the influence of update intervals and batch sizes used for adaptation [28, 53]. To adapt an initial model to a different domain, we find suitable methods in the domain of incremental learning [10, 26, 34].

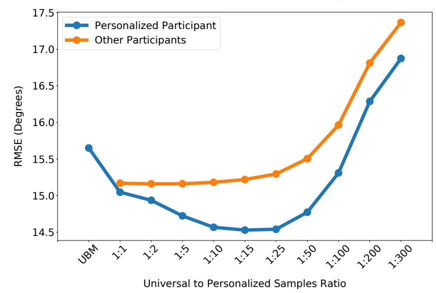
4 AUTOMOTIVE USE CASE

The following methodological example illustrates a multistage approach to achieve an adaptive neuro-symbolic autonomous system (from the automotive domain) with continuous user feedback Figure 1, according to the mentioned research questions.

The first stage is to understand the variances in driver behavior when performing the multimodal referencing task as in [19, 20, 46]. 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 [20] to maximize overall performance (i.e., turn gaze detection off for user accompanied by wandering behavior of the eye, 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 [19]. 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 [21] 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. Although 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., through 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 [27]), learning from the driver's behavior (e.g., Efficient Learning from Demonstrations [33]), and learning from the driver's feedback (e.g., Implicit Human Feedback Learner [9]). Since the main focus of this work is on adaptation and user-specific personalization, Figure 3a and Figure 3b show examples of related work results focusing on the adaptation aspect of [20]. Specifically, Figure 3a shows how driver reference 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 pointing- and gaze-specific accuracy clusters. This resembles the hybrid-fusion approach discussed earlier. Similarly, Figure 3b highlights the results of the incremental learning personalization approach previously discussed in [21]. 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.



(a) Clustering drivers' pointing and gaze behavior based on the system's perceived performance (i.e., referencing accuracy).



(b) RMSE results comparing different personalized models among themselves and against the UBM.

Figure 3: Some examples of different adaptation and personalization approaches from [20, 21] that are aligned with the suggested neuro-symbolic adaptation.

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.

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] Heike Adel. 2018. *Deep learning methods for knowledge base population.* Ph. D. Dissertation. LMU.
[2] Abdul Rafey Aftab, Michael von der Beeck, and Michael Feld. 2020. 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.* ACM, 595–603.
[3] Abdul Rafey Aftab, Michael Von Der Beeck, Steven Rohrhirsch, Benoit Diotte, and Michael Feld. 2021. Multimodal Fusion Using Deep Learning Applied to Driver's Referencing of Outside-Vehicle Objects. In *2021 IEEE Intelligent Vehicles Symposium (IV).* IEEE, 1108–1115.
[4] Masataro Asai and Alex Fukunaga. 2018. Classical planning in deep latent space: Bridging the subsymbolic-symbolic boundary. In *Proceedings of the Conference on Artificial Intelligence (AAAI'18).* AAAI Press, 6094–6101.
[5] Shaibal Barua, Mobyen Uddin Ahmed, and Shahina Begum. 2020. Towards intelligent data analytics: A case study in driver cognitive load classification. *Brain Sciences* 10, 8 (2020), 1–19.
[6] Eddie Brown, David R. Large, Hannah Limerick, and Gary Burnett. 2020. Ultrahapticons: "Haptifying" Drivers' Mental Models to Transform Automotive Mid-Air Haptic Gesture Infotainment Interfaces. 54–57.
[7] Joanna J. Bryson and Andreas Theodorou. 2019. *How Society Can Maintain Human-Centric Artificial Intelligence.* Springer Singapore, Singapore, 305–323.
[8] Ulrich Büker. 1998. Hybrid Object Models: Combining Symbolic and Subsymbolic Object Recognition Strategies. In *Proceedings of the International Conference on Information Systems, Analysis and Synthesis (ISAS'98).* IIIS, 444–451.
[9] Yuchen Cui, Qiping Zhang, Brad Knox, Alessandro Allievi, Peter Stone, and Scott Niekum. 2021. The EMPATHIC Framework for Task Learning from Implicit Human Feedback. In *Proceedings of the 2020 Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 155)*, Jens Kober, Fabio Ramos, and Claire Tomlin (Eds.). PMLR, 604–626. https://proceedings.mlr.press/v155/cui21a.html
[10] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. 2021. A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence* 44, 7 (2021), 3366–3385.
[11] Ivan Donadello, Luciano Serafini, and Artur d'Avila Garcez. 2017. Logic Tensor Networks for Semantic Image Interpretation. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'17).* IJCAI Organization, 1596–1602. https://doi.org/10.24963/ijcai.2017/221
[12] Jiang Dong, Dafang Zhuang, Yaohuan Huang, and Jingying Fu. 2009. Advances in multi-sensor data fusion: Algorithms and applications. *Sensors* 9, 10 (2009), 7771–7784.
[13] Michael Feld, Robert Neβelrath, and Tim Schwartz. 2019. 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.* 145–190.
[14] Kikuo Fujimura, Lijie Xu, Cuong Tran, Rishabh Bhandari, and Victor Ng-Thow-Hing. 2013. 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.* ACM, 56–62.
[15] Ray Fuller. 2005. Towards a general theory of driver behaviour. *Accident Analysis and Prevention* 37, 3 (5 2005), 461–472.
[16] Alexander Gepperth and Barbara Hammer. 2016. Incremental learning algorithms and applications. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'16).* ESSAN, 357–368.
[17] Milan Gnjatović, Jovica Tasevski, Milutin Nikolić, Dragiša Mišković, Branislav Borovac, and Vlado Delić. 2012. Adaptive multimodal interaction with industrial robot. In *2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics.* IEEE, 329–333.
[18] Amr Gomaa. 2022. Adaptive User-Centered Multimodal Interaction towards Reliable and Trusted Automotive Interfaces. In *Proceedings of the 2022 International Conference on Multimodal Interaction.* 690–695.
[19] Amr Gomaa, Alexandra Alles, Elena Meiser, Lydia Helene Rupp, Marco Molz, and Guillermo Reyes. 2022. 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.* 215–225.
[20] Amr Gomaa, Guillermo Reyes, Alexandra Alles, Lydia Rupp, and Michael Feld. 2020. 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, 501–509.
[21] Amr Gomaa, Guillermo Reyes, and Michael Feld. 2021. 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, 318–327.
[22] Alain Grumbach. 1995. Learning at subsymbolic and symbolic levels. In *Neural Networks: Artificial Intelligence and Industrial Applications – Proceedings of the Annual SNN Symposium on Neural Networks*, Bert Kappen and Stan Gielen (Eds.). Springer London, 91–94.
[23] Lisa Hassel and Eli Hagen. 2005. Adaptation of an automotive dialogue system to users' expertise. In *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue.* 222–226.
[24] Mir Riyanul Islam, Shaibal Barua, Mobyen Uddin Ahmed, Shahina Begum, Pietro Aricò, Gianluca Borghini, and Gianluca Di Flumeri. 2020. A novel mutual information based feature set for drivers' mental workload evaluation using machine learning. *Brain Sciences* 10, 8 (8 2020), 1–23.
[25] Srinivasan Janarthanam and Oliver Lemon. 2014. Adaptive generation in dialogue systems using dynamic user modeling. *Computational Linguistics* 40, 4 (2014), 883–920.
[26] Luo Jie, Tatiana Tommasi, and Barbara Caputo. 2011. Multiclass transfer learning from unconstrained priors. In *Proceedings of the International Conference on Computer Vision (ICCV'17).* IEEE, 1863–1870.
[27] Ya Jing, Junbo Wang, Wei Wang, Liang Wang, and Tieniu Tan. 2020. Relational graph neural network for situation recognition. *Pattern Recognition* 108 (2020), 107544.
[28] Christoph Käding, Erik Rodner, Alexander Freytag, and Joachim Denzler. 2016. Fine-tuning deep neural networks in continuous learning scenarios. In *Proceedings of the Asian Conference on Computer Vision (ACCV'16 Workshops).* Springer, 588–605.
[29] AA Karpov and RM Yusupov. 2018. Multimodal interfaces of human–computer interaction. *Herald of the Russian Academy of Sciences* 88, 1 (2018), 67–74.
[30] Troy Dale Kelley. 2006. 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 (Sept. 2006), 219–222. https://doi.org/10.5772/5736
[31] Hansol Kim, Kun Ha Suh, and Eui Chul Lee. 2017. Multi-modal user interface combining eye tracking and hand gesture recognition. *Journal on Multimodal User Interfaces* 11, 3 (2017), 241–250.
[32] W Bradley Knox and Peter Stone. 2009. Interactively shaping agents via human reinforcement: The TAMER framework. In *Proceedings of the fifth international conference on Knowledge capture.* 9–16.
[33] Quanyi Li, Zhenghao Peng, and Bolei Zhou. 2022. Efficient Learning of Safe Driving Policy via Human-AI Copilot Optimization. In *International Conference on Learning Representations.* https://openreview.net/forum?id=0cgU-BZp2ky
[34] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. 2017. Deep transfer learning with joint adaptation networks. In *Proceedings of the International Conference on Machine Learning (ICML'17) - Volume 70.* ACM, 2208–2217.
[35] Udara E Manawadu, Mitsuhiro Kamezaki, Masaaki Ishikawa, Takahiro Kawano, and Shigeki Sugano. 2017. A multimodal human-machine interface enabling situation-Adaptive control inputs for highly automated vehicles. In *2017 IEEE Intelligent Vehicles Symposium (IV).* IEEE, 1195–1200.
[36] Elena Meiser, Alexandra Alles, Samuel Selter, Marco Molz, Amr Gomaa, and Guillermo Reyes. 2022. 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.* 83–86.
[37] Mohammad Mehdi Moniri and Christian Müller. 2012. 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.* ACM, 241–248.
[38] Robert Neßelrath, Mohammad Mehdi Moniri, and Michael Feld. 2016. Combining speech, gaze, and micro-gestures for the multimodal control of in-car functions. In *Proceedings of the 12th International Conference on Intelligent Environments.* IEEE, 190–193.
[39] Natalia Neverova, Christian Wolf, Graham Taylor, and Florian Nebout. 2015. Moddrop: adaptive multi-modal gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 8 (2015), 1692–1706.
[40] Andrzej Nowak, Paul Lukowicz, and Pawel Horodecki. 2018. 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 (2018), 26–34.
[41] R. Polikar, L. Upda, S.S. Upda, and V. Honavar. 2001. Learn++: an incremental learning algorithm for supervised neural networks. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)* 31, 4 (Nov. 2001), 497–508. https://doi.org/10.1109/5326.983933
[42] Alex Ratner and Christopher Ré. 2018. Knowledge Base Construction in the Machine-learning Era. *Queue* 16, 3, Article 50 (June 2018), 12 pages. https://doi.org/10.1145/3236386.3243045
[43] Seth Rogers, C-N Fiechter, and Cynthia Thompson. 2000. Adaptive user interfaces for automotive environments. In *Proceedings of the IEEE Intelligent Vehicles*

Symposium 2000 (Cat. No. 00TH8511). IEEE, 662–667.

[44] Florian Roider and Tom Gross. 2018. 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*. ACM, 351–358.

[45] Florian Roider, Sonja Rümelin, Bastian Pfleging, and Tom Gross. 2017. 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*. ACM, 94–102.

[46] Sonja Rümelin, Chadly Marouane, and Andreas Butz. 2013. 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*. ACM, 40–47.

[47] Jeffrey C. Schlimmer and Richard H. Granger. 1986. Incremental learning from noisy data. *Machine Learning* 1, 3 (1986), 317–354. https://doi.org/10.1007/BF00116895

[48] Luciano Serafini and Artur d'Avila Garcez. 2016. Logic tensor networks: Deep learning and logical reasoning from data and knowledge. *arXiv preprint arXiv:1606.04422* (2016).

[49] Tevfik Metin Sezgin, Ian Davies, and Peter Robinson. 2009. Multimodal inference for driver-vehicle interaction. In *Proceedings of the 11th International Conference on Multimodal Interfaces*. ACM, 193–198.

[50] Jude W. Shavlik. 1994. Combining symbolic and neural learning. *Machine Learning* 14, 3 (March 1994), 321–331. https://doi.org/10.1007/BF00993982

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

[52] Erin T. Solovey, Marin Zec, Enrique Abdon Garcia Perez, Bryan Reimer, and Bruce Mehler. 2014. Classifying Driver Workload Using Physiological and Driving Performance Data: Two Field Studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 4057–4066.

[53] Gido M Van de Ven and Andreas S Tolias. 2019. Three scenarios for continual learning. *arXiv preprint arXiv:1904.07734* (2019).

[54] Laura Von Rueden, Sebastian Mayer, Jochen Garcke, Christian Bauckhage, and Jannis Schuecker. 2019. Informed machine learning–towards a taxonomy of explicit integration of knowledge into machine learning. *Learning* 18 (2019), 19–20.

[55] Wei Xu. 2019. Toward human-centered AI: a perspective from human-computer interaction. *interactions* 26, 4 (2019), 42–46.

[56] Yanxia Zhang, Sophie Stellmach, Abigail Sellen, and Andrew Blake. 2015. The costs and benefits of combining gaze and hand gestures for remote interaction. In *Human-Computer Interaction – INTERACT 2015*. Springer, 570–577.