Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)
Edited by Maria Chiara Leva, Edoardo Patelli, Luca Podofillini, and Simon Wilson

Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5183-4_S33-01-310-cd

Neuro-Symbolic AI for Sensor-based Human Performance Prediction: System Architectures and Applications

Inês Filipa Fernandes Ramos

Department of Computer Science, Università degli Studi di Milano, Italy. E-mail: ines.fernandes@unimi.it

Gabriele Gianini

Department of Computer Science, Università degli Studi di Milano, Italy. E-mail: gabriele.gianini@unimi.it

Ernesto Damiani

Department of Computer Science, Università degli Studi di Milano, Italy. E-mail: ernesto.damiani@unimi.it

Recently, due to the rapid development of deep learning methods, there has been a growing interest in Neuro-symbolic Artificial Intelligence, which takes advantage of both explicit symbolic knowledge and statistical sub-symbolic neural knowledge representations. In sensor-based human performance prediction (HPP) for safety-critical applications, where maintaining optimal human and system performance is a major concern, neuro-symbolic AI systems can improve sensor-based HPP tasks in complex working settings. In this paper, we focus on the advantages of hybrid neuro-symbolic AI systems, present the outstanding challenges and propose possible solutions for HPP in the safety-critical application domain.

Keywords: Artificial Intelligence, Artificial Neural Networks, Neuro-Symbolic Systems, Human Performance Prediction, Safety-Critical Applications, Explicit Symbolic Knowledge.

1. Introduction

The assessment of human performance represents an important challenge in safety-critical industries such as aviation, medicine, military, space exploration, automotive, nuclear and infrastructure industries, where errors can result in loss of life, significant damage to property or environment. This topic has been heavily researched and translated into practice through the discipline of human factors and ergonomics (HFE). To anticipate and prevent critical scenarios, it is essential to identify what are the reliable predictors of human performance that can be monitored and analysed by the computer/machine and used to support the human operator. Many factors in a complex work setting can impact human performance. Human performance can be modelled by two macro factors: task complexity and human capability. In the field of physical, organizational and cognitive ergonomics, most of the external/extrinsic factors to the human that can impact the task complexity and human capability have been taken into account in the design of systems, procedures and policies for safety-critical systems. On the other hand, intrinsic factors (inherent to the individual) that affect human capability and performance, such as fatigue, boredom, anxiety, stress or internal distraction Krueger (1989); Staal (2004); Hanoch and Vitouch (2004), are harder to identify, predict and prevent. To mitigate the occurrence of errors originated by intrinsic factors, the available solutions often consist of preventative measures enacted before the task, or operator state assessment using behavioural indicators or subjective questionnaires

In safety-critical complex systems, the prediction of decreased human performance and adaptive support of the human operator, can benefit from the monitoring of relevant intrinsic factors and their impact on the cognitive (attention, memory, decision-making, processing speed), physical (dexterity, resilience) and perceptive capability (auditory, visual perception and processing). Sensor data from body signals have been extensively used as indicators of intrinsic human factors, such as mental, cognitive and emotional states, that affect human performance in a complex and uncertain way. (see for instance Zhou et al. (2020)).

Deep Learning (DL), in the meanwhile, has gained momentum due to the increased availability of large data sets, computational power and data processing options. DL methods allow efficient feature mining and learning from raw data, including from heterogeneous time-series data.

1.1. *Challenges in sensor-based human performance prediction*

The majority of current studies that use sensor-based deep learning methods focus on mental workload, a multidimensional construct to represent how the human mental and cognitive limited resources are affected by the task demands, in a specific environmental context Young et al. (2015). Despite its usefulness to determine situations of excessive workload, which can lead to serious accidents in safety-critical scenarios, there is no common formal definition of the concept, of how to measure it, and no clear relationship with human performance Dehais et al. (2020).

Moreover, most studies use very well-defined simple tasks that are made to elicit different levels of mental workload easily differentiated from each other, but human performance degradation depends on a large number of known and unknown factors in complex human-system interaction scenarios, and it can have different meanings in different contexts and tasks.

Human performance in real complex work environments is a heterogeneous multi-factor multi-context problem, hard to capture and model in a meaningful and general way. When attempting to leverage black-box deep learning methods, it is therefore essential to consider the quantity and quality of the data that is used, and to consider the task and environmental context to understand and verify the generated outputs.

On this basis, some general open challenges in the field can be identified:

- The application of such data-driven systems to real complex work environments and tasks, requires the recording, integration and labelling of very large amounts of training data under many conditions (that should include uncommon safety-related emergency conditions for safety-critical applications).
- Current sensor-based HPP models lack integration of context and task-related knowledge, that are essential to the able to generalize such models to different tasks and work environments.
- HPP, especially for safety-critical applications, requires AI outputs and learned knowledge to be interpretable, verifiable, and trustworthy.
- Sensor-based HPP models require ways to identify and correct biases related to data being gathered from a limited sample of the population and to the large number of factors that can affect performance.

2. Neuro-symbolic AI solutions

The study of Neuro-symbolic systems was initially motivated by attempts to model the human brain, the mind, and learning and reasoning abilities in the field of computational cognitive modelling.

Currently, neuro-symbolic AI research has evolved by encompassing a variety of AI approaches that make use, on one side, of explicit symbolic knowledge representations – such as logic rules, ontologies, knowledge bases and graphs that allow to formally represent expert or world knowledge – and, on the other side, of neural knowledge representations, or embeddings. The two kinds of representations can be used in a hybrid or integrated way, allowing for the incorporation of different types of knowledge, at different stages of the learning/reasoning workflow.

Neuro-symbolic AI can thus be defined as the study of AI systems that take advantage of both *data-driven* (also referred to as *sub-symbolic*) deep neural learning and *knowledge-based symbolic* reasoning, with the goal of mitigating the weaknesses while preserving the strengths of these two complementary methodologies. Experience shows (see Sarker et al. (2021)) that, when used together, the two methodologies can bring several benefits: they require less labelled data, can correct/identify biased data, increase transparency and explainability, handle out-of-distribution data points, even when not in training data, maintain scalability, ability to handle data uncertainties, high recognition rates, and globally increase the capabilities of the system using background knowledge.

The way in which neural and symbolic representations can interact, obviously, depends on the kind of prior knowledge that is being processed. The work by von Rueden et al. (2021) identified the most commonly used sources of prior knowledge: scientific knowledge (formal, scientifically validated), world knowledge (common-sense, intuitive, validated by reasoning about observed world relationships), or expert knowledge (informal, validated by an expert group). Such knowledge can be represented by different types of formalizations, often dependent on the source and on how it can be integrated into the AI pipeline. Some symbolic knowledge representation types follow.

- Logic Rules can formalize knowledge about facts and dependencies (e.g. IF A THEN B).
- Knowledge Graphs (KGs) consist of vertices describing concepts, and edges representing relationships (e.g. "Man wears hat"); weighted edges quantify strength/sign of a relationship.
- Ontologies are KGs defining a set of representational primitives with which to model a domain of knowledge or discourse; the primitives are typically classes (or sets), attributes (or properties), and relationships among them.

Kautz proposed a taxonomy (reported by Sarker et al. (2021)) with 5 categories based on the architecture and integration of neural and symbolic representations in the system pipeline :

- *[Symbolic Neuro Symbolic]* - both the input and output information are symbolic: reasoning or learning is carried out by the neural system. Examples can be found in Xie et al. (2019); Manhaeve et al. (2018); Yang et al. (2018).
- *[Symbolic [Neuro]]* - a neural system is used to learn a search strategy or detect patterns within a symbolic reasoner. A notorious example is AlphaGo by Silver et al. (2016), a Go playing system that combines Monte Carlo Tree search and DL to narrow the search space and obtain the best next move; another one is represented by self-driving cars systems, where neural systems are in charge of perception and a symbolic reasoner of decision making.
- *[Neuro ∪ compile(Symbolic)]* - symbolic rules are used as input and output and are compiled to be processed by neural system. Examples are in Hohenecker and Lukasiewicz (2017); de Penning et al. (2010); Xu et al. (2018).
- *[Neuro → Symbolic]* – pipeline with a neural system followed by a symbolic reasoner. Examples are in Dang-Nhu (2020); Mao et al. (2019).
- *[Neuro[Symbolic]]* - a symbolic reasoner is used to process learned neural representations within the neural system. The neural network will learn from the inputs/outputs of the symbolic reasoner Sarker et al. (2021).

The work by van Bekkum et al. (2021) went further in describing the interaction between the subsymbolic and the symbolic components of a hybrid neuro-symbolic system, and proposed more than 15 architectures based on simple base modules. The proposed architectures are used herein to describe the proposed hybrid systems and provide an idea (even though a high-level one) of how a system can be configured for specific AI tasks and applications.

3. Neuro-symbolic AI: possible approaches for sensor-based HPP

Neuro-symbolic AI has recently regained attention due to the rapid development of DL methods, which has been accompanied by the increasing need of more data and, as the models become more complex, the need for explanations of the decisions it makes Sarker et al. (2021); von Rueden et al. (2021); Tiddi and Schlobach (2022). Developments have been recently made in the fields of natural language, image recognition, recommender systems and rule-based machine learning Tiddi and Schlobach (2022). However, little work has been proposed to tackle sensor-based predictive tasks. This can be partially attributed to the prevalence of machine learning methods, that rely on manual feature selection, over deep learning methods, that require unrealistic large amounts of data to be able to detect less frequent complex events from multiple sensor data Xing et al. (2020). Moreover, the integration of prior symbolic knowledge into deep architectures remains a challenge, as in order to maintain its meaning, interpretability and logic, the reasoning process cannot be made differentiable Tiddi and Schlobach (2022); Xu et al. (2018). The integration is therefore most commonly performed at the level of the learning algorithm or at the output of the model, as to validate the predictions von Rueden et al. (2021). Out of the benefits of using hybrid AI systems, the approaches that are proposed next focus on the perspective of using world or expert context-related knowledge as a prior and on making learned knowledge more interpretable, verifiable, and trustworthy, essential characteristics of AI for safety-critical applications.

3.1. *A first example*

3.1.1. *An ideal Neuro-Symbolic Agent*

NSCA is a Neuro-symbolic Cognitive Agent developed in de Penning et al. (2010) which integrates symbolic and neural representations, by using a Recurrent Temporal Restricted Boltzmann Machine (RTRBM) to encode temporal logical rules/relations in terms of beliefs and previously applied rules. These rules can be constructed based on expert knowledge about a complex system/task. The agent was developed to model complex temporal relations between the input data by induction, training the neural network to reflect the rules that can be applied based on the inputs to the visible layer (beliefs) and the relationship with previously applied rules.[a]

The beliefs can be continuous or binary data representing probabilities of occurrence of an event, state of the environment (e.g. $raining = true$), real values (e.g. $age = 16$) or even equalities or inequalities to represent beliefs over continuous variables (e.g. $speed < 30$). This is possible due to the continuous stochastic visible layer that is used to represent the beliefs as fuzzy sets with a gaussian membership function, useful when using subjective knowledge. The rules should be defined about the beliefs on conditions, scenarios or contexts and related to the previous time step. Any rule can then be mapped to the RTRBM as a relation between the hidden unit (rule), the visible units (beliefs) and the previously activated hidden units (previously applied rules). The network then learns (by induction) from training with real examples, using the initially defined rules as prior knowledge, and can be used to perform deduction over new data or for extraction of the updated rules in symbolic form.

In de Penning et al. (2010), the NSCA was used to model existing complex expert knowledge of driving instructors, trained with data from driving simulators and expert evaluations, and used to infer driver assessment scores.

3.1.2. *Neuro-Symbolic Agent for safety-critical environemnts*

It is now clearer how the principles and architecture of this method can benefit human performance prediction/assessment, specifically in safety-critical activities, in which getting training data is not feasible for many dangerous scenarios (such as in the oil and gas industry, aviation industry or rail industry). Moreover, in these types of industries it is common to perform human reliability assessment studies, that can provide qualitative or quantitative information about the likelihood of human error, under different conditions and at different stages of a task. The already existing domain and expert knowledge can be used as prior knowledge when predicting or assessing the response of an operator in different types of tasks, and emergency situations with different levels of risk and/or priority. In addition, the encoding of the prior knowledge as temporal logic rules maintains the transparency of the model and explainability of the inferred performance scores.

A possible way to apply this method consists of building a hybrid system, where one or more deep neural model can be used to automatically extract relevant complex patterns from body sensor data,

[a] An RTRBM consists of an artificial neural network with two layers: a visible layer V and a hidden layer, connected by weights W. Recurrent connections link the hidden unit activations at the current time step and the previous time step. The hidden units are conditionally independent and can be treated as calculating the posterior probability that a certain rule R is applied given the observed beliefs b (input data at current time step t in the visible layer V) and the previously applied rule r_{t-1} (i.e. $P(R|B = b, R_{t-1} = r_{t-1})$. The likelihood of the beliefs and previously applied rules can be computed based on the most applicable rules. The network can then be trained by comparing the inferred values of the beliefs and previous rules and the observed values.

to classify the state of the operator (can be attention, mental workload, physical workload, task engagement or emotion), followed by a neuro-symbolic cognitive agent, such as a NSCA, to encode task-related and context-related symbolic rules, and predict the operator performance based on prior and learnt knowledge.

As a practical example, in an alarm management control room scenario, a prior task-related rule can be related to the maximum number of alarms that can be turned on for a period of time, that will lead to an assessment of decreased operator performance. Both the maximum number of alarms and the period of time variables can be set initially according to expert knowledge and updated during training, according to the predicted state of the operator, the state of the process being monitored, the severity or priority of the alarms, or even the complexity of the task.

The source of the prior knowledge in this application is mostly scientific and/or expert knowledge, represented as temporal symbolic logic rules, and is an hybrid system architecture combining two categories of the Kautz proposed taxonomy: Neuro – Neuro[compile(Symbolic)], a strictly neural system followed by another neural system that processes compiled symbolic logic rules about prior knowledge, and also outputs symbolic logic rules. A possible general architecture for the hybrid system is the one represented in Figure 1, using the elementary pattern proposed in van Bekkum et al. (2021).

3.2. *A second example*

Another possible approach can be applied for operator performance/emotion/state multi-class classification task, by considering the semantic meaning of the labels and use it to constrain the learning algorithm, by adding an additional semantic loss term. An example is when classifying multiple states, where some contradict each other and others do not, such as positive emotions and negative ones. In this case, the learning performance might be improved by maximizing the difference of the predicted probabilities between the sets of conflicting labels. The work Xu et al. (2018), proposed a method for using symbolic knowledge in deep learning systems, through the application of a semantic loss function that constrains the output. It further demonstrated that applying a semantic loss term, even for the simple exactly-one constraint (exactly only one output label must be true and the rest false), can improve semi-supervised classification problems by increasing the confidence of prediction for unlabelled data. The source of the symbolic knowledge in this application it is world knowledge, represented as semantic consistency constraints, and the hybrid system general architecture, not covered by the Kautz taxonomy, can be described by an architecture proposed in van Bekkum et al. (2021), specifically a pattern that uses symbolic rules to inform neural learning (Figure 2).

3.3. *A third example*

The final approach proposed here focuses on using end-to-end deep learning methods and integrating a posteriori domain ontologies or knowledge graphs to increase the interpretability of the predictions, and produce more understandable explanations. This type of reasoning systems for post-hoc construction of explanations has been proposed before to construct explanations for the prediction of stock trends, by using a large existing knowledge graphs with data about events and price values Deng et al. (2019). Similarly, in Confalonieri et al. (2020), ontologies modelling the domain knowledge are added to the process of explanation generation, with an algorithm that uses decision trees to explain artificial neural networks. A user study showed that the understandability of the generated decision trees increased with the integration of the domain knowledge. An example of an application is to explain the outcomes of a deep learning model, trained to predict the driver's reaction time to a stimulus while driving, using heterogeneous features, such as the driver's mental state, the driver's characteristics (age, experience), the weather, noise level, light level, traffic, etc. In this case it is possible to build a domain ontology relating these concepts, enriching them with semantic information and allowing to generate more specific or general explanations, according to the user preference. The source of the

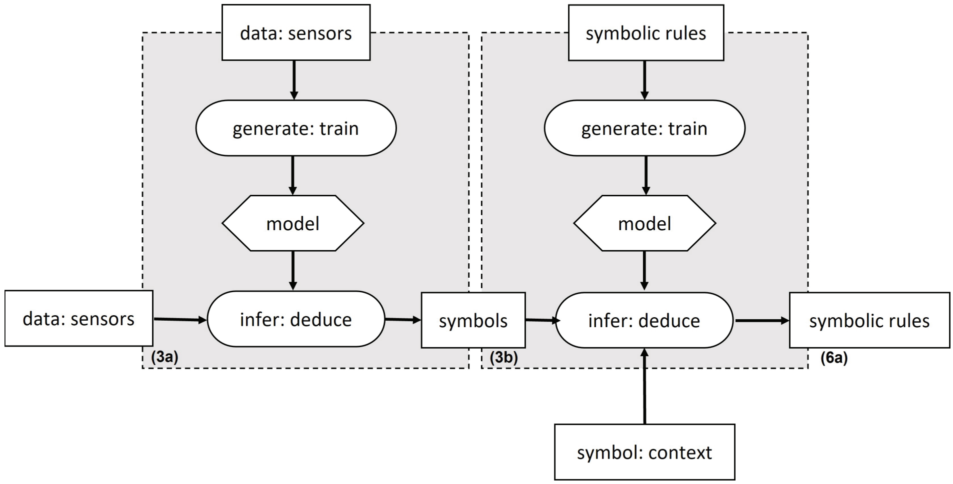

Fig. 1. Hybrid system architecture pattern modified from van Bekkum et al. (2021), to implement NSCA for HPP.

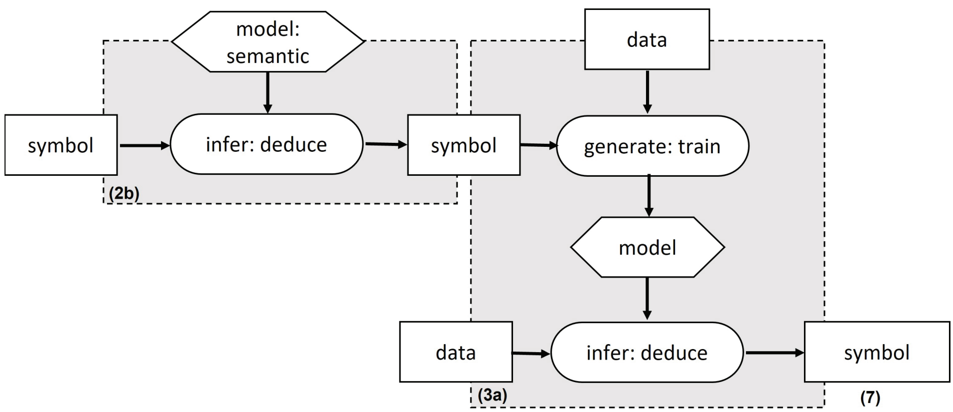

Fig. 2. Hybrid system architecture pattern proposed in van Bekkum et al. (2021), to implement informed learning with prior knowledge.

symbolic knowledge in this general application it is world knowledge, represented as a domain ontology, and the hybrid system general architecture can be described by both the Kautz category Neuro $\rightarrow$ Symbolic, and by an architecture proposed in van Bekkum et al. (2021) (Figure 3).

3.4. *A fourth example*

A more concrete application example that can benefit from all three above-mentioned hybrid neuro-symbolic approaches, can be in an industrial assembly line scenario, where the operator performs a sequence of complex activities. Assuming the factory is equipped with multiple sensors including body or operator focused sensors, this data can be used by a deep learning model to classify/detect simple activities from the raw sensor data (perception layer) and followed by a reasoning model to associate the detected sequences of simple activities with high-level more complex assembly line activities (reasoning layer), according to prior human knowledge about the assembly line tasks. This hybrid model can be used to detect human errors or deviations from the expected sequence of activities, for early detection of errors or prevention of future ones. The hybrid system

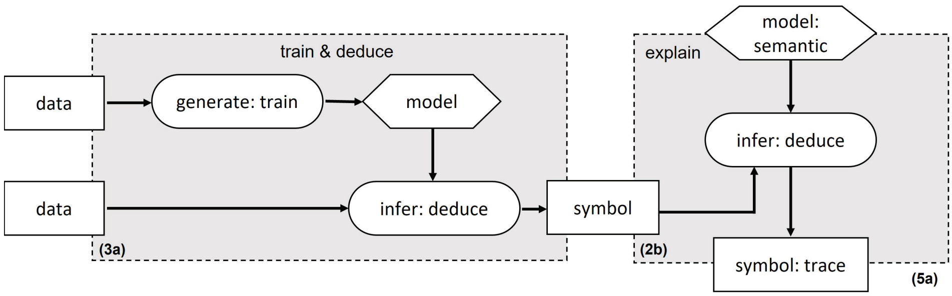

Fig. 3. Hybrid system architecture pattern proposed in van Bekkum et al. (2021), for integrating learning with domain knowledge graphs or ontologies for post-hoc explainability.

to be applied here is similar to the one in Figure 1, however the detected sequence of simple activities have a direct link to the final complex assembly activity that is detected, so for this application a hybrid system that allows for end-to-end training with only the raw sensor data and high-level assembly activity annotations given, would be ideal. Such a system, called Neuroplex, was proposed in Xing et al. (2020), in which neural networks were used for perception of simple events from sensor data and neurally reconstructed reasoning models were used to detect complex events with larger spatial and temporal dependencies (after being trained with human knowledge provided as logical rules about simple-to-complex event dependencies). In addition, a semantic loss was also applied on the intermediate symbolic layer, similarly to the second approach proposed here for human state multi-class classification, to constrain the symbolic output of the neural network in order to improve the training process. Another parallel AI system can be added to this application scenario, by training a deep learning model to predict these deviations, based on big sensor data from the operator and the system, with the goal to predict and prevent errors. To have a better understanding of the factors that contribute to errors and deviations, it is essential that the model predictions can be explained. Considering that in this scenario the operator and system factors, and the relationships between them, are complex and possibly hard to understand without knowledge about the processes and production activities of the factory, a dedicated domain ontology or knowledge graph can be built beforehand by experts to aid the construction of more understandable explanations.

4. Conclusions

The preliminary study performed in this work was able to establish the benefits that neuro-symbolic AI approaches can bring to the domain of sensor-based human performance prediction. It was also clear that, for this type of task, the integration of symbolic knowledge with deep neural architectures is difficult and there are few developed methodologies at the moment for this purpose. Further work is needed to be able to demonstrate the feasibility of hybrid AI systems for sensor-based HPP.

Acknowledgement

The authors acknowledge the support of the CISC project funded from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 955901.

References

Confalonieri, R., T. Weyde, T. R. Besold, and F. Moscoso del Prado Martín (2020). Trepan reloaded: A knowledge-driven approach to explaining artificial neural networks.

Dang-Nhu, R. (2020). Plans: Neuro-symbolic program learning from videos. *Advances in Neural Information Processing Systems 33*, 22445–22455.

de Penning, L., A. d. Garcez, L. C. Lamb, and J. Meyer (2010). An integrated neural-symbolic cognitive agent architecture for training and assessment in simulators. In *Proceedings of* *the 6th International Workshop on Neural-Symbolic Learning and Reasoning, NeSy10, held at AAAI-2010*.

Dehais, F., A. Lafont, R. Roy, and S. Fairclough (2020). A neuroergonomics approach to mental workload, engagement and human performance. *Frontiers in neuroscience*, 268.

Deng, S., N. Zhang, W. Zhang, J. Chen, J. Z. Pan, and H. Chen (2019). Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In *Companion Proceedings of The 2019 World Wide Web Conference*, pp. 678–685.

Hanoch, Y. and O. Vitouch (2004). When less is more: Information, emotional arousal and the ecological reframing of the yerkes-dodson law. *Theory & Psychology 14*(4), 427–452.

Hohenecker, P. and T. Lukasiewicz (2017). Deep learning for ontology reasoning. *arXiv preprint arXiv:1705.10342*.

Krueger, G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work & Stress 3*(2), 129–141.

Manhaeve, R., S. Dumancic, A. Kimmig, T. Demeester, and L. De Raedt (2018). Deepproblog: Neural probabilistic logic programming. *Advances in Neural Information Processing Systems 31*.

Mao, J., C. Gan, P. Kohli, J. B. Tenenbaum, and J. Wu (2019). The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. *arXiv preprint arXiv:1904.12584*.

Sarker, M. K., L. Zhou, A. Eberhart, and P. Hitzler (2021). Neuro-symbolic artificial intelligence: Current trends. *arXiv preprint arXiv:2105.05330*.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. (2016). Mastering the game of go with deep neural networks and tree search. *nature 529*(7587), 484–489.

Staal, M. A. (2004). Stress, cognition, and human performance: A literature review and conceptual framework.

Tiddi, I. and S. Schlobach (2022). Knowledge graphs as tools for explainable machine learning: A survey. *Artificial Intelligence 302*, 103627.

van Bekkum, M., M. de Boer, F. van Harmelen, A. Meyer-Vitali, and A. t. Teije (2021). Modular design patterns for hybrid learning and reasoning systems. *Applied Intelligence 51*(9), 6528–6546.

von Rueden, L., S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, M. Walczak, J. Pfrommer, A. Pick, R. Ramamurthy, J. Garcke, C. Bauckhage, and J. Schuecker (2021). Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*, 1–1.

Xie, Y., Z. Xu, M. S. Kankanhalli, K. S. Meel, and H. Soh (2019). Embedding symbolic knowledge into deep networks. *Advances in neural information processing systems 32*.

Xing, T., L. Garcia, M. R. Vilamala, F. Cerutti, L. Kaplan, A. Preece, and M. Srivastava (2020). Neuroplex: learning to detect complex events in sensor networks through knowledge injection. In *Proc.s of the 18th Conference on Embedded Networked Sensor Systems*, pp. 489–502.

Xu, J., Z. Zhang, T. Friedman, Y. Liang, and G. Broeck (2018). A semantic loss function for deep learning with symbolic knowledge. In *International conference on machine learning*, pp. 5502–5511. PMLR.

Yang, F., D. Lyu, B. Liu, and S. Gustafson (2018). Peorl: Integrating symbolic planning and hierarchical reinforcement learning for robust decision-making. *arXiv preprint arXiv:1804.07779*.

Young, M. S., K. A. Brookhuis, C. D. Wickens, and P. A. Hancock (2015). State of science: mental workload in ergonomics. *Ergonomics 58*(1), 1–17.

Zhou, T., J. S. Cha, G. Gonzalez, J. P. Wachs, C. P. Sundaram, and D. Yu (2020). Multimodal physiological signals for workload prediction in robot-assisted surgery. *ACM Transactions on Human-Robot Interaction (THRI) 9*(2), 1–26.