AI for real-world network operation

WP2– Fundamental AI building blocks

D2.1– Position paper on AI for the operation of critical energy and mobility network infrastructures

AI4REALNET has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101119527, and from the Swiss State Secretariat for Education, Research and Innovation (SERI).

DOCUMENT INFORMATION

DOCUMENT	D2.1– Position paper on AI for the operation of critical energy and mobility network infrastructures
TYPE	R
DISTRIBUTION LEVEL	PU
DUE DELIVERY DATE	30/06/2024
DATE OF DELIVERY	28/06/2024
VERSION	V1.0
DELIVERABLE RESPONSIBLE	POLIMI
AUTHOR (S)	Marco Mussi (POLIMI)
	Gianvito Losapio (POLIMI)
	Alberto Maria Metelli (POLIMI)
	Marcello Restelli (POLIMI)
OFFICIAL REVIEWER/s	Ricardo Bessa (INESC TEC)
	Antoine Marot (RTE)

DOCUMENT HISTORY

VERSION	AUTHORS	DATE	CONTENT AND CHANGES
Version 1.0	POLIMI	28/06/2024	First Delivery

ACKNOWLEDGEMENTS

NAME	PARTNER
Ricardo Bessa	INESC TEC
Daniel Boos	SBB
Clark Borst	TU Delft
Alberto Castagna	EnliteAI
Duarte Dias	INESC TEC
Adrian Egli	SBB
Andrina Eisenegger	FHNW
Yassine El Manyari	EnliteAI
Anton Fuxjäger	EnliteAI
Samira Hamouche	FHNW
Mohamed Hassouna	FGH & UKASSEL
Bruno Lemetayer	RTE
Roman Liessner	DB
Jonas Lundberg	LIU
Manuel Schneider	Flatland
Irene Sturm	DB
Julia Usher	ZHAW
Herke Van Hoof	UvA
Jan Viebahn	TenneT
Toni Waefler	FHNW

DISCLAIMER

This project is funded by the European Union and SERI. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union and SERI. Neither the European Union nor the granting authority can be held responsible for them.

SUMMARY

The rapid evolution of Artificial Intelligence (AI) is driving significant advancements across various sectors, enhancing efficiency, decision-making, and innovation. AI's widespread integration into daily life and industries demonstrates its transformative potential, creating a more interconnected and automated world. Beyond simple automation, AI tackles complex tasks and large-scale problems with a market value in the trillions of dollars. This highlights the growing significance of AI and the necessity for effective human-AI interaction. The AI4REALNET project seeks to harness AI to improve the efficiency, safety, and resilience of critical infrastructures such as power grids, railways, and air traffic networks. This position paper outlines AI4REALNET's approach to applying AI in network infrastructure operations, translating application needs into algorithmic proposals for effective human-AI collaboration in decision-making processes. The document includes a review of AI decision-making and human-AI interaction, use cases for each critical infrastructure, and digital environments for training AI agents. It also identifies promising research directions to enhance critical network infrastructures, such as advanced learning techniques and human-AI co-learning models. By combining advanced AI algorithms with human expertise, the AI4REALNET project aims to create resilient and efficient systems. This involves developing learning techniques, hybrid co-learning models, and autonomous AI systems operating under human supervision.

TABLE OF CONTENTS

LIST OF FIGURES **8**

1. INTRODUCTION **9**

2. SEQUENTIAL DECISION MAKING **11**
2.1. SINGLE-AGENT FRAMEWORKS 11
2.1.1. MARKOV DECISION PROCESSES 12
2.1.2. PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES 13
2.1.3. FACTORED MARKOV DECISION PROCESSES 13
2.1.4. SEMI-MARKOV DECISION PROCESSES 13
2.2. MULTI-AGENT FRAMEWORKS 14
2.2.1. MARKOV GAMES 14
2.2.2. PARTIALLY OBSERVABLE MARKOV GAMES 15
2.2.3. DECENTRALIZED PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES 15

3. AI APPLICATIONS IN CRITICAL INFRASTRUCTURES **16**
3.1. POWER GRIDS 16
3.2. RAILWAY NETWORKS 20
3.3. AIR TRAFFIC MANAGEMENT 23

4. AI-FRIENDLY DIGITAL ENVIRONMENTS **26**
4.1. POWER GRIDS: GRID2OP 26
4.2. RAILWAY NETWORKS: FLATLAND 27
4.3. AIR TRAFFIC MANAGEMENT: BLUESKY 28

5. RESEARCH PROPOSALS AND RELATED WORKS **28**
5.1. RL-BASED APPROACHES 29
5.1.1. DISTRIBUTED REINFORCEMENT LEARNING 29
5.1.2. HIERARCHICAL REINFORCEMENT LEARNING 32
5.1.3. INVERSE REINFORCEMENT LEARNING 33
5.2. EXPLAINABLE AI 35
5.2.1. COUNTERFACTUAL AND CONTRASTIVE EXPLANATIONS 36
5.2.2. FEATURE IMPORTANCE METHODS 37

5.2.3. VISUALIZATION-BASED EXPLANATIONS 39
5.3. AUTOMATION TRANSPARENCY 40
5.3.1. THE HUMAN SIDE OF HUMAN-AI COLLABORATION 40
5.3.2. KEY HUMAN PROCESSES 41
5.3.3. DESIGNING FOR TRANSPARENCY 50
5.3.4. AUTOMATION TRANSPARENCY RELATED RESEARCH PROPOSAL 51
5.4. KNOWLEDGE-ASSISTED AI 54
5.5. CO-LEARNING 58

6. CONCLUSIONS 61

REFERENCES 62

LIST OF FIGURES

FIGURE 1 - THE AGENT-ENVIRONMENT INTERACTION PROTOCOL. 12
FIGURE 2 - COMPARISON OF THE THREE DIGITAL ENVIRONMENTS FROM AI4REALNET. 26
FIGURE 3 - MODEL OF HUMAN DECISION-MAKING 54
FIGURE 4 - PRELIMINARY DESIGN FOR HUMAN-AI CO-LEARNING BASED ON Van den Bosch et al. (2019). 60

1. INTRODUCTION

Artificial intelligence (AI) has rapidly evolved, paving the way for an unprecedented enhancement of many sectors and industries. From healthcare to finance, from education to entertainment, AI technologies are being deployed to improve efficiency, improve decision-making, and offer innovative solutions to complex problems. The pervasiveness of AI is apparent in how it influences daily activities, such as personalized recommendations on streaming platforms, intelligent virtual assistants that streamline personal and professional tasks, and sophisticated algorithms that drive the success of social media and e-commerce platforms. This widespread adoption underscores AI's transformative potential and role in shaping a more interconnected and automated world.

Moreover, AI's capabilities extend far beyond simple automation and data analysis. It is increasingly being adopted to tackle complex tasks and manage large-scale physical problems. According to international reports (Chui et al., 2018), AI is predicted to unlock a market value in the order of trillions of US dollars. This indubitably witnesses the relevance that AI has acquired and will acquire in the next few years. Additionally, the widespread adoption of AI is creating the conditions for a new and inevitable interaction between humans and AI systems. In such a scenario, creating an ecosystem in which humans and AI systems interact in a healthy way, where the roles and positions of both actors have to be clearly identified, is a critical challenge for research and industry in the next few years.

In this scenario, the AI4REALNET project sets the ambitious goal to adopt AI-based solutions to enhance critical infrastructures' efficiency, safety, and resilience. The project's motivation and focus are directed to large-scale physical systems such as power grids, railways, and air traffic networks, where AI itself and, even more, the collaboration between AI and humans can potentially unlock new appealing opportunities. All these applicative scenarios share the complexity of the decision process and have a structure based on networks in common. In power grids, AI can enhance load and renewable energy forecasting and support operations with a high share of renewable energy with optimized preventive and remedial actions, facilitating the integration of renewable energy sources and improving grid stability. In railway networks, AI can optimize scheduling, predictive maintenance, and real-time traffic management, reducing delays and operational costs. For air traffic control, AI can improve flight path optimization, detect potential hazards, and manage airspace more effectively, increasing safety and reliability. Overall, integrating AI into these critical infrastructures, explicitly accounting for and leveraging human presence, can lead to more resilient, sustainable, and efficient systems, significantly benefiting operators and users. By developing innovative decision-making algorithms, AI4REALNET aims to augment human capabilities, ensuring that AI systems not only support and learn from human operators. The project also seeks to create trustworthy AI-assisted human control and develop novel AI algorithms within open-source environments that emulate real-world scenarios.

This deliverable sets the position of AI4REALNET consortium for the application of AI to the operations in network infrastructure. It aims to transform the needs of the described application scenarios into a roadmap with precise algorithmic proposals for the effective and actual adoption of AI systems that are able to collaborate with humans to solve the corresponding decision processes.

Following rigorous scientific principles, we start identifying the mathematical frameworks that match the peculiar characteristics of the applicative scenarios described above. To this end, in Section 2, we conduct a comprehensive review of the state-of-the-art in AI, discussing both the foundational and technical aspects of decision-making and ones related to human-AI interaction and co-learning. This review is grounded in the specific use cases identified by AI4REALNET, which reflect the current challenges and requirements associated with this kind of network-based environments. By examining existing solutions and technologies, we can identify gaps and opportunities where the AI4REALNET project impacts. This analysis is crucial for understanding the project's landscape and ensuring our efforts align with real-world needs.

In parallel, to effectively understand the needs and peculiarities of the application scenarios of network-based critical infrastructure, the AI4REALNET designed several use cases and digital environments for each critical infrastructure (power grid, railway, and air traffic management). In Section 3, for each application, a description of the main open challenges, relevant related works, and use cases are provided. Furthermore, in Section 4, the digital environments are described. These contributions will allow us to match the application scenarios with the mathematical frameworks described in the previous sections. Furthermore, the availability of digital environments will unlock important opportunities for training AI agents in simulation.

Once the knowledge of the available frameworks (Section 2) and the knowledge of the application scenarios are acquired (Section 4), we move to delineate the novel research direction that will be investigated in the context of the project. They reflect the view of the AI4REALNET consortium regarding which directions are believed to be promising and should be considered for making AI systems enhance the functioning of critical network infrastructures, guaranteeing the requested reliability and safety constraints. In Section 5, we identify several classes of approaches that are particularly relevant for addressing the open problems in our use cases. These approaches include advanced supervised and reinforcement learning techniques, hybrid human-AI co-learning models, and autonomous AI systems designed to work under human supervision. Each approach brings strengths and challenges, and their effective integration is key to achieving the project's objectives. For each of them, we revise the methodologies, the available approaches in the literature, and the challenges feeding the new research directions with particular reference to the needs and peculiarities of the application scenarios of interest for the project. By defining these research areas, we establish a clear direction for the AI4REALNET efforts and ensure that all project activities are aligned with the overarching goals of enhancing critical infrastructure through AI. Even beyond the project's scope, this leads to the identification of a position of the consortium, which concertizes into a structured, multi-faceted approach to developing AI systems that complement and enhance human decision-making capabilities in critical system operations. This concept will guide the research and development efforts within AI4REALNET, ensuring that we achieve our ambitious goals coherently and effectively, and has the ambition of bringing inspiration to academia and industry, contributing to the development of the next-generation AI systems.

2. SEQUENTIAL DECISION MAKING

Reinforcement Learning (RL, Sutton and Barto, 2018) is a branch of Machine Learning (ML, Bishop, 2006) that studies *sequential decision-making* problems. In RL, we consider the scenario in which a learning agent sequentially interacts with an environment. In this interaction framework, at each time step t, the agent receives an observation of the state of the environment s_t and selects action a_t. The action a_t causes an evolution of the environment to a new state s_{t+1} and generate a reward r_t, the function of the state s_t and action a_t. RL agents aim to learn how to optimize the expected sum of the rewards in an unknown environment by learning from data.

In this section, we survey the possible RL frameworks we are interested in to solve the tasks presented in Section 1. First, in Section 2.1, we discuss the commonly adopted frameworks for *single-agent* RL, then, in Section 2.2, we survey the relevant frameworks for *multi-agent* RL.

2.1. SINGLE-AGENT FRAMEWORKS

In this section, we revise the frameworks of interest for what concerns single-agent reinforcement learning. We start in Section 2.1.1 by presenting finite-horizon Markov Decision Processes (MDPs). Then, in Section 2.1.2, we generalize MDPs to Partially Observable MDPs (POMDPs), to model the case where the state cannot be completely observed. Subsequently, in Section 2.1.3, we introduce Factored Markov Decision Processes (FMDPs), a particularization of MDPs with a factorized structure. Finally, in Section 2.1.4, we present Semi-Markov Decision Processes (SMDPs), a generalization of MDPs considering more general temporal dynamics.

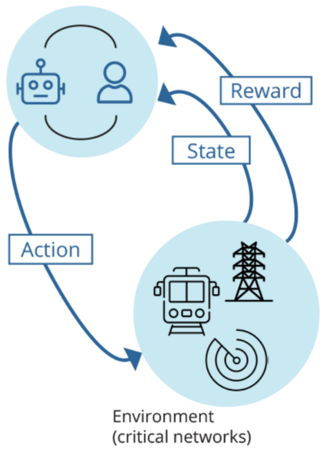

FIGURE 1 - THE AGENT-ENVIRONMENT INTERACTION PROTOCOL.

2.1.1. MARKOV DECISION PROCESSES

Markov Decision Processes (MDPs, Puterman, 1990) offer a well-studied mathematical framework to model sequential decision-making problems. An MDP is formalized as a tuple $\mathcal{M} := \langle \mathcal{S}, \mathcal{A}, P, R, H \rangle$, where $\mathcal{S}$ is the set of environment states the environment can assume, $\mathcal{A}$ is the set of actions the agent can execute, $P : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0,1]$ is the stochastic transition function, being $P(s'|s,a)$ is the probability of moving to next state s' when performing action a in state s, $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the reward function, being $R(s,a)$ is the reward the agent gets when performing action a in state s, and $H \in \mathbb{N}$ is the planning horizon.[1] In MDPs, both the transition probability and the reward function depend only on the current state and action, as the state and action histories are irrelevant (a.k.a. *Markov property*). The agent chooses actions according to a *policy* $\boldsymbol{\pi} = (\pi_t(a|s))_{t\in[H]}$ that maps, at each $t \in [H]$, each state to a (possibly stochastic) action, where $\pi_t(a|s)$ is the probability that the agent chooses a when the environment is in state s at time t, formally $\pi_t : \mathcal{S} \times \mathcal{A} \to [0,1]$. Finding an optimal solution to an MDP means searching for a policy $\boldsymbol{\pi}^*$ maximizing the expected sum of the rewards achieved during the H steps we interact with the environment, formally $\boldsymbol{\pi}^* \in \arg\max_{\boldsymbol{\pi}\in\prod} \mathbb{E}_{\boldsymbol{\pi}}[J(\boldsymbol{\pi})]$, where $\mathbb{E}_{\boldsymbol{\pi}}[J(\boldsymbol{\pi})] = \mathbb{E}_{\boldsymbol{\pi}}\left[\sum_{t=1}^{H} R(s_t,a_t)\right]$ is the expected cumulative reward of policy $\boldsymbol{\pi}$.

All the sequential decision-making problems we present in Section 3 can be effectively modeled using Markov Decision Processes (MDPs). Indeed, MDPs provide a framework for these problems by encapsulating the stochastic nature of decision-making and the optimization of long-term rewards, thereby facilitating the development of policies that can guide decision-makers. We point out, however, that when addressing the aforementioned scenarios with particular classes of algorithms (e.g., distributed and hierarchical RL), the resulting interactions give rise to other frameworks, which will be discussed

[1] In this work, we focus on finite-horizon scenarios due to their affinity with the real-world use cases we are considering. All the settings we introduce in this section can also be seen in a discounted infinite-horizon fashion by introducing a discount factor $\gamma \in [0,1)$.

in the remainder of this section.

2.1.2. PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES

The problem of partial observability is ubiquitous when facing sequential decision-making problems in the real world. In these problems, the agent can partially observe the environment's state.
The framework of MDPs can be generalized to take partial observability into account. Partially Observable Markov Decision Processes (POMDPs, Åström, 1965; Smallwood and Sondik, 1973; Monahan, 1982; Littman, 2009) are a generalization of MDPs in which the environment is assumed to have a well-defined latent state s that underlies and produces the environment's observations o.
A POMDP can be formally described as a tuple $\mathcal{M}_P := \langle \mathcal{S}, \mathcal{A}, P, R, H, \Omega, O \rangle$ where $\langle \mathcal{S}, \mathcal{A}, P, R, H \rangle$ are defined as before for the underlying MDPs, Ω is a set of possible observations, O is the conditional observation probabilities $O : \mathcal{S} \times \mathcal{A} \times \Omega \to [0, 1]$, i.e., $O(o|s', a)$ is the probability of observing $o \in \Omega$ when performing action a and transitioning to new state s'.

2.1.3. FACTORED MARKOV DECISION PROCESSES

Factored Markov Decision Processes (FMDPs, Boutilier et al., 1995, 1999; Degris and Sigaud, 2013; Osband and Van Roy, 2014) are a particularization of MDPs that makes it possible to represent the transition and the reward functions of some problems compactly (compared to an explicit enumeration of state-action pairs). Such formulation is useful when the action space is large. Formally, an FMDP is formalized as a tuple $\mathcal{M}_F := \langle \mathcal{S}, \mathcal{A}, P, R, H \rangle$, where the state space factors are in the form of $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_m$ where one state describes one combination of *substates* $\mathbf{s} = (s_1, \ldots, s_m) \in \mathcal{S}$, the action space $\mathcal{A}$ is defined as for MDPs, the transition probability P satisfies the factored Markovian form $P(\mathbf{s}'|\mathbf{s}, a) = \prod_{i \in [m]} P(s'_i|\mathbf{s}, a)$, the reward function satisfies the separable structure $R(\mathbf{s}, a) = \sum_{i \in [m]} R_i(s_i, a)$, and H is the horizon. This factorization structure means that each action has an independent transition when conditioned on the current state and action. The policy can be independently factored across the components of the action vector given a state, formally $\pi_t(a|\mathbf{s}) = \prod_{i \in [m]} \pi_{t,i}(a|s_i)$.

2.1.4. SEMI-MARKOV DECISION PROCESSES

Semi-Markov Decision Processes (SMDPs, Howard, 1963; Ross, 1970; Drappo et al., 2023) are a generalization of MDPs that admit temporally extended actions, i.e., actions that can execute for a certain time during which the agent has no control over the decision process. Formally, an SMDP is formalized as a tuple $\mathcal{M}_S := \langle \mathcal{S}, \mathcal{A}, P, R, H \rangle$, where $\mathcal{S}$ is the set of possible states, $\mathcal{A}$ is the temporally extended action space, $P(s', t'|s, a, t)$ is the probability of ending to state s' after $(t' - t)$ steps, by playing the temporally extended action $a \in \mathcal{A}$ in state s at stage $t \in [H]$ (H is the length of the episode), $R(s, a, t)$ is the reward accumulated until the termination of the temporally extended action a played in state s at stage $t \in [H]$ of the episode. This modeling approach allows for the precise representation of state transitions and rewards over extended and non-uniform time intervals, providing a more accurate and flexible framework for optimizing decisions in hierarchical structures, as we will see in Section 5.1.2. By leveraging SMDPs, one can address the complexities inherent in these problems, ensuring that strategies account for both the timing and sequence of decisions across different levels of the hierarchy.

2.2. MULTI-AGENT FRAMEWORKS

In this section, we summarize the basic frameworks of *Multi-Agent Reinforcement Learning* (MARL). In particular, we first focus (Section 2.2.1) on *Markov Games* (MG) and *Partially Observable Markov Games* (POMG), an extension of MG to the scenario of partial observability (Section 2.2.2). Finally, in Section 2.2.3, we move our attention to the Decentralized Markov Decision Processes (Dec-POMDPs), a specification of cooperative POMGs. All the Distributed Reinforcement Learning (DRL, Section 5.1.1) problems can be effectively modeled using Markov Games. Markov Games, also known as stochastic games, extend the Markov Decision Process framework to multi-agent environments where multiple agents interact, each with distinct goals and strategies.

2.2.1. MARKOV GAMES

Markov Games (MGs, Shapley, 1953a; Littman, 1994; Filar and Vrieze, 1997) formalize the interactions and decision-making processes of multiple agents in a dynamic environment. Formally, we define a Markov game as a tuple $\mathcal{G} := \langle n, \mathcal{S}, (\mathcal{A}_i)_{i\in[n]}, P, (R_i)_{i\in[n]}, H\rangle$, where n is the number of agents, $\mathcal{S}$ is the set of possible states of the environment, $\mathcal{A}_i$ is the set of possible actions available to agent i, for $i \in [n]$ (where n is the number of agents), $P : \mathcal{S} \times \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \times \mathcal{S} \to [0, 1]$ is the state transition function, which defines the probability of transitioning from one state to another given the current state and the joint actions of all agents, $R_i : \mathcal{S} \times \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to \mathbb{R}$ is the reward function for agent i, which specifies the immediate reward obtained by agent i starting from a given state and given the actions of all the agents, $H \in \mathbb{N}$ is the horizon.[2] The dynamics of MGs follow the principles of MDPs described before, where transitions between states are governed by stochastic processes and satisfy the Markov property. Specifically, the next state of the environment depends only on the current state and the joint actions of *all* the n agents, regardless of the history of previous states and actions. Agents in MGs aim to maximize their expected cumulative rewards over time by selecting actions that

[2] For $n \in \mathbb{N}$, we define $[n] := \{1, 2, \ldots, n\}$.

lead to favorable outcomes. This often involves learning optimal strategies through exploration and exploitation of the environment. Given the relation between the agents' rewards, we can divide multi-agent RL problems as follows. If all the rewards of the different agents behave according to each other in all the state and action spaces, the agents need to coordinate and collaborate, and we talk about *cooperative MARL*. On the other hand, if increasing the reward of an agent, the ones of the other agents decrease, we talk about *competitive MARL*. In this work, we focus only on the cooperative scenario.

The most general theoretical objective of MGs is represented by a Nash equilibrium (Nash, 1950). Given the set of n agents whose action spaces are represented by $(\mathcal{A}_i)_{i\in[n]}$, each agent i tries to find a policy $\pi^i : \mathcal{S} \to \Delta(\mathcal{A}^i)$ which maximizes its own value function. In this case, the cumulative reward of each agent i is controlled not only by its own policy π^i, but also by the policies of all other agents π^{-i}, thus the value function depends on the same time on the stochasticity of π^i and π^{-i}.

A Nash equilibrium for the Markov Game $\mathcal{G}$ is a joint policy $\pi^* = \left(\pi^{1,*}, \pi^{2,*}, \ldots, \pi^{n,*}\right)$ such that each value function of $\pi^{i,*}$ is larger than any other possible value function when the combined with the other policies $\pi^{-i,*}$. It represents an equilibrium point in which the value function of each agent i is maximized concerning its own policy and the other agents' policies. More than one equilibrium point may exist.

2.2.2. PARTIALLY OBSERVABLE MARKOV GAMES

Partially Observable Markov Games (POMGs, Hansen et al., 2004; Liu et al., 2022) extend the concept of Markov games to scenarios where agents have incomplete information about the true state of the environment. Formally, a POMG is defined by a tuple $\mathcal{G}_P := \langle n, \mathcal{S}, (\mathcal{A}_i)_{i\in[n]}, P, (R_i)_{i\in[n]}, H, (\Omega_i)_{i\in[n]}, O\rangle$, where $\langle n, \mathcal{S}, (\mathcal{A}_i)_{i\in[n]}, P, (R_i)_{i\in[n]}, H\rangle$ are defined as before for the underlying MG and Ω_i is the set of possible observations for agent $i \in [n]$, and $O : \mathcal{S} \times \Omega_i \times \cdots \times \Omega_n \to [0, 1]$ is the observation probability (i.e., $O(o_1, \ldots, o_n|s)$ the probability of observing $(o_1, \ldots, o_n) \in \Omega_1 \times \cdots \times \Omega_n$ in state $s \in \mathcal{S}$).

2.2.3. DECENTRALIZED PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES

Decentralized Partially Observable Markov Decision Processes (Dec-POMDPs, Bernstein et al., 2002; Amato et al., 2013; Oliehoek and Amato, 2016) generalize POMDPs to the multi-agent, cooperative, decentralized setting. Dec-POMDPs model a team of agents that must cooperate to solve some task by receiving local observations and individually selecting and executing actions over a sequence of time steps. The agents share a *single reward function* that specifies their objective but is not typically observed during execution. Execution is decentralized because each agent must select its action at each time step without knowing the actions chosen or observations received by the other agents.

Formally, a Dec-POMDP is defined by a tuple $\mathcal{M}_{PD} := \langle n, \mathcal{S}, (\mathcal{A}_i)_{i\in[n]}, P, R, H, (\Omega_i)_{i\in[n]}, O\rangle$ where n is the number of agents, $\langle \mathcal{S}, P, R, H\rangle$ are defined as before for POMDPs (Section 2.1.2) considering the action space $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ with $\mathcal{A}_i$ representing the finite set of actions for an agent $i \in [n]$. Moreover, Ω_i is the set of possible observations for agent $i \in [n]$, and O is the set of the conditional observation probabilities $O : \mathcal{S}\times\mathcal{A}\times\Omega_1\times\cdots\times\Omega_n \to [0, 1]$, i.e, $O(o_1, \ldots o_n|s', a)$ is the probability of observing $(o_1, \ldots, o_n) \in \Omega_1\times\cdots\times\Omega_n$ when performing action a and transitioning to new state s'. The essential difference w.r.t. POMG is that we have a single reward function $R : \mathcal{S} \times \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to \mathbb{R}$ shared over all the agents (i.e., $R_i = R$ for every $i \in [n]$).

3. AI APPLICATIONS IN CRITICAL INFRASTRUCTURES

The AI4REALNET project is focused on three critical infrastructures whose cyber-physical assets, systems, and networks are considered vital in Europe, and their disruption would have a debilitating effect on society. These infrastructures are from the energy (power grid) and mobility (railway and air traffic management) sectors, two of the five priority sectors identified in the European national AI strategies (Roy et al., 2021). This section discusses the challenges, project use cases defined by the industrial partners, related work, and AI-friendly digital environments used in the project for the development and validation of novel AI technologies.

3.1. POWER GRIDS

The energy sector has undergone a significant transformation in the last two decades driven by decarbonization, decentralization, and digitalization (Silvestre et al., 2018). Decarbonization efforts have led to the integration of Renewable Energy Sources (RES), replacing carbon-intensive technologies, and introducing new energy vectors such as green hydrogen.

CHALLENGES

This transition poses technical challenges across the entire power system due to the variability and uncertainty of RES in terms of balancing, voltage management, and line loadability. Decentralization is realized through distributed generation technologies and the emergence of prosumers, empowering local communities to produce and consume electricity. This leads to a need for new tools in long-term planning, operational (short-term) planning, and real-time network operation. Digitalization, driven initially by smart meters, is expanding to include grid users and service providers, fostering a more connected and intelligent energy landscape through concepts like Digital Twins and the Internet of Energy.

Aspects such as assets aging, the need to postpone large financial investments to increase power network capacity and meet ambitious RES targets, and climate change are creating pressure on the grid infrastructure, which needs to keep the same reliability standards that modern society is used to. Therefore, an important challenge is how to operate power grids with high-RES integration, minimizing the curtailment of carbon-free generation and, at the same time, ensuring high resilience to adversarial natural and man-made events (Panteli et al., 2017).

Furthermore, current power network supervision practices involving multiple screens and applications impose a heavy cognitive load on users, requiring them to prioritize, organize, and correlate disparate information and alarms before making decisions or taking action (Andrade et al., 2022). This fragmented ecosystem is increasingly challenging for operators to manage, especially as the number of applications grows and they operate in non-integrated formats. The result is information overload and a lack of contextual understanding of system problems (Marot et al., 2022a).

In summary, the energy transition is triggering significant changes in the operational landscape for power system operators, requiring reimagining the architecture of control centers and the role of human operators, emphasizing the need for more evolutionary, standardized, and modular integration. Additionally, there is a growing recognition of the importance of human-centric design in control center environments, acknowledging the central human role in decision-making. Thus, the focus should also be on creating interfaces and intelligent assistants that facilitate efficient decision-making processes and enhance operator effectiveness (Marot et al., 2022b; Viebahn et al., 2022; Fuxjäger et al., 2023), in managing complex operating conditions during the evolving energy landscape. Modern AI technology can bring value in fast decision-making on operating and planning power systems with high shares of RES (Heymann et al., 2024), where the full use of flexibility from various sources (generation, consumers, or grid assets) is fundamental. This is especially crucial under challenging scenarios, such as extreme weather events and cyber-attacks, where the system's adaptability becomes instrumental in maintaining infrastructure/system integrity and resilience.

RELATED WORKS

As a high-risk sector, power networks have largely used expert systems as a core AI technology (Viebahn et al., 2024). Expert systems are favored for their structured representation and storage of expert knowledge, enabling consistent decision-making and facilitating the documentation and transfer of expertise. One of the first state-of-the-art reviews was published in 1989, framing AI under the name "expert systems" (Zhang et al., 1989), and even today, expert systems remain prevalent in commercial products and grid automation. Notable examples include SPARSE, an online assistant for operators of substation control centers (Vale and e Moura, 1993), and an online transient stability analysis system utilized by B.C. Hydro control center (Demaree et al., 1994).

However, the increasing complexity of power systems, coupled with the integration of RES, has driven a growing demand for adaptable solutions capable of learning from data. This has led to significant research in Artificial Neural Networks (ANN) and other Machine Learning (ML) methodologies, such as decision trees and fuzzy inference systems, mainly focused on power system operation. Industry success stories include the use of decision trees and ANN for dynamic security assessment in the Hydro-Québec power system (Huang et al., 2022), various ML models for short-term RES forecasting, predicting distribution network faults and repair duration based on historical data (Kezunovic et al., 2020). These examples demonstrate the increasing importance and effectiveness of AI/ML techniques in addressing complex challenges within the energy sector.

Industry-driven challenges, such as Learning to Run a Power Network (L2RPN), have boosted collaboration among AI scientists and power system specialists (Marot et al., 2021), to solve congestion problems. After congestion occurs or potential congestion has been identified, power network operators need to apply remedial actions to contain the situation and prevent larger (cascading) events that may lead to a blackout. These collaborative efforts motivated different groups to develop a new Reinforcement-Learning-based assistant to aid human operators in operating electrical grids during normal operations and when the system is under stress due to overloads or disturbances. Although traditional tools like optimal power flow (Capitanescu, 2016) can be used to aid human operators in making decisions to solve congestion problems, they may have large computational times to provide very fast advice in the presence of multiple combinatorial solutions (e.g., re-dispatch, network topology change). In this scenario, the rapid inference capability of AI-based approaches can deliver good, fast solutions to humans, affording them additional time, if necessary, to conduct a thorough analysis and use supplementary study tools.

Dehnavi et al. (2022) proposed a power system partitioning and an associated congestion index to assign priorities to congested lines to assist operators in alleviating the congestion in more critical lines first. The assigned priorities are then assessed on their effectiveness using power transfer distribution factor metrics. Then, a zonal congestion management model is used to reschedule generation based on the zones where the congestion was alleviated most. Another approach to mitigate congestion is topology reconfiguration. Modifying the network's topology to mitigate or prevent congestion is a highly effective approach to avoid more costly approaches like generation re-dispatch or, in worst cases, load shedding. Many different approaches have been taken, with Marot et al. (2020) being one of the first to explore topology optimization using AI, which proposes a framework to develop topology controllers using RL. This framework has been used and led to various AI-based approaches to modify the topology of the network, such as a) Subramanian et al. (2021) that uses the above-described framework to develop a simple baseline approach, which is a modified cross-entropy method to take topological switching actions, b) Chauhan et al. (2023) that develops a deep-learning approach combined with heuristics that helps take in a reduced action space that facilitates faster learning, c) Lehna et al. (2023) that develops a curriculum learning strategy with increasingly complex environments for more and more capable agents, from greedy exploration search to behavior cloning to population-based RL-agent training, d) Zhou et al. (2021) that combines deep-learning with evolutionary training to account for non-differentiable functions such as grid simulation, e) Dorfer et al. (2022) that successfully applied AlphaZero to the power grid congestion problem enhancing the agent planning capability without any prior action space reduction. Huang et al. (2023) combined various interdisciplinary methods within this framework, including a scalable grid simulation environment and a highly scalable physics-informed three-stage deep RL agent training process (including a two-stage curriculum learning method). Gholizadeh et al. (2023) made a comparative study between five different RL algorithms for topology optimization such as Deep Q-Networks (DQN, Mnih et al., 2015), policy gradient (Silver et al., 2014) and actor-critic (Konda and Tsitsiklis, 1999) methods, where soft actor-critic (Haarnoja et al., 2018) demonstrated the fastest convergence and deep Q-learning exhibited higher stability and optimality.

Since this is a high-risk sector, aspects such as explainability and interpretability of AI-based systems are becoming fundamental requirements for AI adoption by industry (Heymann et al., 2023). In terms of challenges, this means that AI research should cover inherently interpretable AI models where humans can understand the mechanism that transforms input to outputs and modify it when the system behavior is distant from the expected one. When not possible, explainability (e.g., leveraging from the Shapley values formalism) should be available to understand the model better and support the model designer in improving its performance. Moreover, in an infrastructure traditionally operated by humans, research in human-centric AI should produce solutions that enhance human-machine collaboration and user experience. For instance, the seminal work "Ironies of artificial intelligence" (Endsley, 2023a), identified the need to develop AI systems with "self-awareness" where the AI system can detect and inform situations that are outside of its boundaries of operations. One L2RPN competition (Marot et al., 2022c) integrated an additional term in the score function that measures the capacity of the AI agent to send alarms when it is self-aware of the "incapacity" to solve a specific problem and informs the human operator.

The dynamic nature of energy systems also requires Adaptive AI systems that can adapt (online) to changing conditions, uncertainty (e.g., from RES), new data, and, if possible, human feedback. Finally, certification and formal verification of AI models that operate autonomously or provide recommendations to humans is essential to guarantee trust, but also require standardized methodologies, such as ISO/IEC 24029-2 "Artificial intelligence (AI) – Assessment of the robustness of neural networks – Part

2: Methodology for the use of formal methods''.

AI4REALNET USE CASES

The AI4REALNET industrial partners defined two use cases for the power grid infrastructure (a more detailed description can be found in Bessa et al. 2024):

- **AI assistant supporting human operators' decision-making in managing power grid congestion.** AI-based assistant that oversees the transmission grid, using SCADA data and energy management system tools to identify issues and categorize them for human intervention. It monitors power flows, adhering to defined operational conditions. Anticipating problems, it sends alerts to the operator with confidence levels, avoiding excessive alerts to maintain operator focus. Action recommendations include topological changes, re-dispatching, and renewable energy curtailment.
- **Sim2Real, transfer AI-assistant from simulation to real-world operation.** Considers the AI-based assistant in scenarios characterized by a) uncertainty from noisy and partially missing data, and b) when data limitations prevent full autonomy, the human operator can provide missing information to aid the AI in such situations.

3.2. RAILWAY NETWORKS

Traffic density on the European railway networks is constantly increasing. For example, the densest mixed-traffic railway network in the world is operated by the Swiss Federal Railways (SBB), with more than 12,000 switches and 32,000 signals. In 2023, about 1.32 million passengers were transported on the Swiss railway network every day (SBB, 2024), and demand is expected to increase significantly in the coming years. SBB is planning to increase its capacity by 30% by the year 2040 (CH-FOT, 2024). In Germany, the situation is similarly complex. Every day, over 40,000 regional, long-distance, and freight train journeys take place on the German rail network. The Deutsche Bahn network alone covers around 33,000 kilometers of track, supporting trains and stations with a wide range of characteristics.

CHALLENGES

When thousands of trains are on the move in dense traffic, rapid decisions must be made to adjust the operations in the event of disruptions. These can trigger a cascade of further necessary changes. The projected uptake in demand makes rail traffic management an increasingly complex task. Not only are train schedules highly optimized and susceptible to disturbances of any kind, but they also have to accommodate the construction of new and maintenance of existing infrastructure. Operating such a complex system is no easy task, and especially dealing with unexpected events, such as breakdowns of trains or delays caused by passengers, requires re-scheduling trains in real-time.

Traffic management is usually carried out in operation centers, taking care of a specific region of the network. Human operators are constantly monitoring the traffic, dispatching the trains that are regularly scheduled, and reacting to deviations from the schedule. While the operators are supported by various tools, this task mainly remains a manual one and requires specialized and experienced personnel.

The task of railway scheduling can formally be described as a Vehicle Scheduling Problem (VSP), as introduced by (Bodin and Golden, 1981). Finding solutions to the VSP has been a long-standing research topic in Operations Research (OR) and leveraged techniques from integer programming (Foster and Ryan, 1976) to pathfinding (Potvin and Rousseau, 1993). While the task of planning railway schedules takes place well in advance of operations, traffic management takes place in real-time and has to take into account deviations from the schedule due to unexpected events.

To take these extended requirements into account (Li et al., 2007) proposed an extension to the VSP, the Vehicle ReScheduling Problem (VRSP). However, taking into account the full complexity of modern railway networks the VRSP is an NP-hard problem. In addition, the need for dynamic re-scheduling in real-time makes finding solutions to the VRSP even more challenging and novel approaches are needed.

Further, the railway domain is traditionally operated by human operators. The introduction of new software tools, especially AI-based systems, should carefully consider human factors as well as the societal and ethical impacts. Experience with earlier decision support software showed that the understandability of the assistance provided and its limits are key to human operators (Endsley, 2023a). Operators need to be able to rely on and trust such a system in time-critical situations in particular. Therefore, the explainability of an AI-based decision support system will play a crucial role in human operators' acceptance. In addition, European railway networks usually play a critical role in public transport. Hence, the public constitutes an important stakeholder group, and the robustness, efficiency, and transparency of automated Traffic Management Systems (TMSs) are key factors for the acceptance by this group Westin et al. (2016).

In summary, the introduction of AI-based decision support systems for railway traffic management and the productive collaboration between human operators depends on two factors: One, the system's ability to detect and appropriately react to deviations from the planned schedule in real-time. Reactions may range from suggestions for re-routing of trains to automatically dispatch trains. This requires novel approaches to cope with the complexity of railway networks and dense traffic in a time-critical situation. Two, the system's ability to provide meaningful explanations and forms of interaction to human operators.

RELATED WORKS

The need for advanced railway traffic management systems as well as for standardization of protocols to ensure the interoperability of systems across neighboring railway operators has been recognized by the European Union almost two decades ago (Bloomfield, 2006). Since then, computerized decision support systems for dynamic traffic control systems and algorithms that enable them have been subject to research (D'Ariano, 2009; Corman and Meng, 2015). Nowadays, railway operators use a variety of software to support human operators in traffic management, like SBB's Rail Control System (SBB, 2020).

In recent years, the potential for AI-based approaches for railway traffic management has been recognized (Parvez Farazi et al., 2021; Gorsane et al., 2023). In particular, deep reinforcement learning in combination with agent-based modeling, Multi-Agent Reinforcement Learning (MARL), has demonstrated great potential for the coordination of multiple agents in stochastic environments. First, remarkable results have been achieved in playing games like StarCraft II (Vinyals et al., 2017; Samvelyan et al., 2019) and Dota 2 (OpenAI et al., 2019). Soon after, cooperative MARL has also been applied to the real-world problem of traffic signal control (Arel et al., 2010; Tan et al., 2020).

The open research community around Flatland developed a simulation environment to allow AI researchers to easily test novel approaches to the VRSP (Mohanty et al., 2020). Through machine learning challenges, generalization and scalability of learned behavior in the context of simple railway dispatching tasks could be demonstrated (Laurent et al., 2021). Further, researchers showed the effectiveness of deep reinforcement learning for real-time railway traffic management (Lövétei et al., 2022), and how strategies like communication between agents can further improve the performance of MARL (Roost et al., 2020).

With the increasing use of AI-based systems and their potential for real-time decision support in critical domains, standards for the development and deployment of such systems were defined, like the ISO/IEC 23053 "Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)". Further, principles and frameworks for assessing societal and ethical dimensions of such systems have been developed (e.g., Cannarsa 2021; Dubber et al. 2020; HLEG 2020) and can support both the responsible application of AI and foster acceptance by human operators and the public. Finally, co-learning approaches and human-computer interaction research have addressed challenges of human-AI collaboration and can help develop productive joint human-AI control systems with meaningful human control and tasks, e.g., (Westin et al., 2016; Endsley, 2023a)).

AI4REALNET USE CASES

The AI4REALNET industrial partners defined two use cases for the railway infrastructure (a more detailed description can be found in Bessa et al. 2024):

- **Automated re-scheduling in railway operations.** An automated AI-based system to manage and optimize railway schedules in real-time, ensuring efficient rail network use while minimizing passenger delays. The system is constantly monitored by a human operator who can adjust the system's configuration and identify the need for adaptation and re-training. Adapting the schedule includes interventions, such as changing the speed curves of trains, changing the order of trains at the infrastructure element, changing the routes of trains, or changing the platform of a commercial stop in a station.

- **AI-assisted human re-scheduling in railway operations.** An AI assistant supports the human dispatcher in analyzing the real-time state of all the trains and tracks in the dispatcher's area and recommends possible dispatching options in case of deviations from the pre-planned schedule.

3.3. AIR TRAFFIC MANAGEMENT

In the quest for safer, more efficient, and greener airspace operations, the ATM community is struggling with how to best transition to higher levels of automation and exploit state-of-the-art Machine Learning (ML) and other AI methods in ways that keep humans at the center of operations, especially in planning and tactical operations where the application of AI is most challenging due to acceptance (and trust) issues and certification hurdles. The long-term vision of the Single European Sky ATM Research (SESAR) program anticipates that tasks will eventually be performed collaboratively by hybrid human-AI teams. Here, the goal is not automation and/or using AI *per se* but optimizing the overall performance of the socio-technical ATM system and maximizing human performance and engagement at all times, as specified by the SESAR 3 Multi-annual Work Programme 2022-31.

CHALLENGES

SESAR foresees that the use of hybrid human-AI teams will have several benefits and challenges. AI-based human operator support tools are set to increase capacity in airspaces and ensure the integration of new entrant aircraft types, such as hybrid hydrogen-electric aircraft, which may present challenges related to traffic mixes. In terms of operational efficiency, improvements are anticipated by enabling better traffic predictions and forecasts, which contribute to the punctuality of air traffic operations and necessitate real-time data exchange between air-ground systems. Safety and security will be maintained at least at the same level as the current ATM system. AI-based solutions aim to achieve a positive impact on the operational mitigation of aviation's environmental impact, such as directing traffic over shorter and more energy-efficient high-altitude routes. The development of a framework for Human-AI teaming involves designing AI systems to support both task work and teamwork. This requires considerations about human-AI team performance and processes, team trust, team biases, team situational awareness, team training needs, human-AI interaction methods, interface, transparency and explainability, and Human-System Integration processes, measures, and testing. Explainable Artificial Intelligence (XAI) is crucial because AI/ML algorithm decisions are often opaque, non-intuitive, and ununderstandable by human operators, limiting their applicability in the ATM sector. The objective is to improve the transparency of automated systems in the ATM domain by investigating methods based on XAI in operational use cases, such as predicting air traffic conflict resolution and delay propagation and validating the robustness and transparency of the system. Finally, AI will enrich aviation datasets with new types of data, unlocking air/ground AI-based applications, fostering data-sharing, and building an inclusive AI aviation/ATM partnership. This will support decision-makers, pilots, Air Traffic Controllers (ATCOs), and other stakeholders by increasing ATCO productivity, reducing workload, and enhancing complexity capabilities, ultimately bringing benefits in cost-efficiency.

Currently, the ATM community is implementing the data-exchange infrastructure that should facilitate global communication via high-bandwidth digital datalinks, global surveillance via ADS-B and satellite datalinks, and global navigation via several satellite constellations (e.g., GPS, GALILEO, GLONASS, and BeiDou). In turn, the SESAR program is investigating what (new) forms of automation can be deployed that will eventually take advantage of the overhauled data exchange infrastructure.

RELATED WORKS

ATM consists of several entities that all need to work together seamlessly to achieve safe and efficient air traffic operations. Those entities operate at different time scales, ranging from long-term strategic flight planning (years to months prior to operation), pre-tactical operations (days to hours before operation) towards tactical operations in the execution phase of flight.

In the long-term, AirSpace Management (ASM) applies and enhances the Flexible Use of Airspace (FUA) concept by developing the European airspace into one continuum that is flexible and reactive to changes in airspace users' needs, with the ultimate objective of optimizing the European Network capacity and performance. Here, collaborative decision-making (CDM) among all stakeholders is vital. Next, Air Traffic Flow and Capacity Management (ATFCM) focuses on balancing the management of capacity and demand, planning strategically, and applying tactically as a result of the physical airport or airspace limitations. ATFCM is the primary means of ensuring flight punctuality and efficiency whilst maintaining or improving safety. At the sharp end of airspace operations, Air Traffic Control (ATC) is responsible for directing aircraft on the ground and through a given section of controlled airspace and providing advisory services to aircraft in non-controlled airspace. The primary purpose of ATC worldwide is to prevent collisions, organize and expedite the flow of traffic in the air, and provide information and other support for pilots during the execution of flights.

ATM innovation projects that focus on integrating intelligent, AI-based forms of automation can be found in the areas of ASM, ATFCM, and ATC. The majority of projects are centered on prediction, forecasting, and ML and AI explainability and interpretability. AI methods related to prediction and forecasting have been applied to 1) classifying and predicting operational flight delays from historical data using Gradient Boosted Decision Trees (GBDT) (Dalmau et al., 2021), 2) convective weather prediction for pre-tactical operations using ensemble neural networks (Jardines et al., 2021), 3) airspace capacity management using Bayesian Networks, 4) tactical Air Traffic Control support using Message Passing Neural Networks (MPNN) and Multi-Agent Reinforcement Learning (Dalmau and Allard, 2020), and 5) dynamic airspace sectorization using fuzzy clustering and evolutionary algorithms (Gerdes et al., 2018). AI explainability and interpretability in ATM have been mostly centered on human-machine interfaces supporting visual analytics (Andrienko et al., 2022; Kravaris et al., 2023) and/or encoding AI/ML outputs into state-of-the-art decision-support tools that are familiar to operational users (Westin et al., 2022; IJtsma et al., 2022; van Rooijen et al., 2020).

AI4REALNET USE CASES

The AI4REALNET industrial partners defined two use cases for the airspace infrastructure (a more detailed description can be found in Bessa et al. 2024):

- **Airspace sectorisation assistant.** An AI assistant, capable of operating under various levels of automation, will provide recommendations or even execute decisions on splitting the sector best horizontally, vertically, or both to balance the ATCO workload while ensuring safety and efficient traffic flows. It will also act bidirectionally by allowing the human operator to nudge the AI-generated recommendations in more favorable directions.

- **Flow and airspace management assistant.** The activation/deactivation of military airspace in some airports can induce deviations from the flight plan routes. In this sense, to optimize the lateral deviation of the flights due to avoidance of an eventual temporary military-activated area, an AI assistant can analyze and suggest a decision in sectorization and routing of the main flows in the flight information region (FIR).

	Grid2Op	Flatland	BlueSky
Single or multi-agent?	Single	Multi-agent	Both
Observation space: type & size	Discrete & continuous (very large, > 4,000 dimensions)	Discrete & continuous (very large, > 32,000)	Continuous (very large, > 20,000)
Competitive or collaborative	Collaborative		Both
Sequential or episodic?	Sequential		
Stochastic or deterministic environment?	Stochastic		
Discrete or continuous action space or mixed?	Discrete & continuous		
System represented as a graph?	Yes		No
Size of action space	Very large (> 65,000 different discrete actions & 200 continuous actions)	Very large (grows exponentially with the number of trains)	Large (limited by the number of available sectors and on the number of flights in the sector)

FIGURE 2 - COMPARISON OF THE THREE DIGITAL ENVIRONMENTS FROM AI4REALNET.

4. AI-FRIENDLY DIGITAL ENVIRONMENTS

For developing, validating, and benchmarking novel AI-based systems, open-source AI-friendly digital environments reproducing realistic operating conditions of energy and mobility network infrastructures are fundamental, even to conduct a controlled risk assessment of the systems. Leveraging from a digital environment allows organizations to promote AI developments internally in an organization (in-house AI communities) and connect and co-develop with similar AI communities externally.

This also aims to create a new mindset towards sharing data, community construction of digital environments for AI development/test, evolving the rather rigid critical infrastructures business model towards a more dynamic network joining technological platforms, mobility/energy providers, and customers, and finding appropriate answers to new legal issues concerning liability and ethics, considering that the three infrastructures are high-risk sectors in the EU AI Act.

A comparison of the three digital environments is depicted in Figure 2, and a short description of each environment is presented below.

4.1. POWER GRIDS: GRID2OP

RTE developed the open-source Grid2Op[3] environment to model and study a large class of power system-related problems and facilitate the development and evaluation of controllers (or agents) that

[3] `github.com/rte-france/Grid2Op`

act on power grids. Any control algorithm in interaction with a virtual version of the electrical grid can be used to overcome gaps between research communities.

Through different L2RPN competitions (Marot et al., 2021), calibrated virtual environments have been instantiated for testing over robustness to adversarial attacks, adaptability for increasing renewable energy share, or agent alert trustworthiness. Such "autonomous" agent scenarios can already be visualized and analyzed through the Grid2viz [4] module. Moreover, it is also possible for a human to play live scenarios, assisted by an AI agent with InteractiveAI [5], a generic platform to instantiate AI assistant Interface and interactions for real-time critical system operators. It helps diagnose situations or risks. It can also provide recommendations of remedial actions a human can choose from by embedding trained AI agents.

Chronix2Grid [6] package allows the generation of synthetic but realistic consumption, renewable production, electricity loss (dissipation) and economic dispatched productions chronic for a given power grid.

4.2. RAILWAY NETWORKS: FLATLAND

The Flatland environment[7] is a comprehensive framework developed (by industry partners like SBB, DB, and the AI community) for easy development and experimentation on the vehicle rescheduling problem for railway networks. Flatland represents railway networks as 2D grid environments with restricted transitions between neighboring cells. On the 2D grid, multiple train runs must be performed for a given set of goals and circumstances. Trains are represented as agents that make decisions on movement and navigation.

Flatland is a discrete-time simulation, i.e., it performs all actions with constant time steps. A single simulation step synchronously moves the time forward by a constant increment, thus enacting exactly one action per agent. The Flatland environment is tailored towards RL. It provides observations and rewards to any controlling agent and expects one discrete action per agent per step. Flatland, in its current state, provides a set of global and local observations. It provides generators for generating railway networks and demand for trains (scenarios), an evaluation system, and mechanisms to inject disturbances into rail operations. These disturbances are represented as malfunctions of trains, i.e., trains being unable to move on the track for several time steps. The occurrence of these is distributed according to configurable distribution at scenario definition. After three competitions with Flatland, a comprehensive set of basic (mostly) AI solutions for Flatland exists and can be used as baseline/benchmark models.

[4] grid2viz-neurips.herokuapp.com
[5] https://github.com/IRT-SystemX/InteractiveAI
[6] github.com/BDonnot/ChroniX2Grid
[7] github.com/flatland-association/flatland-rl, github.com/flatland-association/flatland-book

4.3. AIR TRAFFIC MANAGEMENT: BLUESKY

The BlueSky environment is an open-source ATM simulator developed in 2013 with TU Delft as its main developer.[8] It contains open source, open data aircraft performance models and a global navigation database including airports; it is also compatible with Base of Aircraft Data (BADA) (containing performance and operating procedure coefficients for 295 different aircraft types). Although BlueSky started as a simulator aimed at conventional aviation, it has been extended to include several drone/urban air mobility models and functionality in recent years. It has since been applied in several UAM/UTM-related projects.

Through its modular setup, an extension of each of the components of BlueSky (e.g., autopilot, FMS, performance model, conflict detection and resolution, environmental modeling, visualization, etc.) can be reimplemented or extended. In the same way, it is also possible to add completely new functionality to the simulator. By default, BlueSky has its own Qt/OpenGL-based interface that allows the user to control the simulation and get an overview of the simulated traffic. Through its client/server network implementation, BlueSky can also easily interface with separate ATM user interface applications and piloted blip driver stations.

5. RESEARCH PROPOSALS AND RELATED WORKS

In this section, we overview the state-of-the-art and research directions for the AI areas of interest in the AI4REALNET project. Motivated by the critical infrastructures under study in AI4REALNET, which are characterized by various peculiarities, we delineate several research areas that capture and leverage these peculiarities as discussed in the previous sections.

First, network-based critical systems are characterized by a complex structure with a distributed nature and a long-term decision process. This implies the necessity of addressing the decision-making problem while accounting for multiple agents located in a distributed manner. Moreover, the decision process is characterized by a long horizon, where decisions made at a certain time impact far-future performance. The fundamental methodological approach for conquering this complexity is Reinforcement Learning (RL). In Section 5.1, we discuss RL-based solutions that capture the multi-agent distributed long-horizon nature of the problem, including distributed and hierarchical RL. We propose developing novel algorithms that address the challenges of the decision processes in network-based critical infrastructures. Moreover, since an essential requirement of RL is the availability of a reward

[8] `github.com/TUDelft-CNS-ATM/bluesky`

function, and given that the decision processes involve humans employing a certain decision strategy, it is interesting to investigate approaches that recover a representation of the objectives optimized by humans. This is made possible through inverse RL.

Having established the basic frameworks, the subsequent sections analyze several aspects necessary for making RL solutions suitable for deployment in environments where AI is intended to interact seamlessly with humans.

One of the most relevant requirements when deploying an AI system in collaboration with humans is the acceptance of the AI solution, which is strictly related to the ability of humans to understand and trust the AI decision process. This necessitates enforcing explainability constraints on the deployed solutions, allowing human users and experts to trace the motivations behind decisions. This is discussed in Section 5.2, where we review the state-of-the-art and propose novel research directions for the AI4REALNET project.

Tightly connected to the previous topic, automation transparency further constrains the relationship between the AI system and humans. In Section 5.3, we discuss in more detail the frameworks for human-AI collaboration, list the requirements, and explore approaches that implement design strategies enforcing transparency in the decision process.

A further step towards developing AI systems that comfortably interact with humans involves investigating techniques to develop AI systems guided by already existing human knowledge. Knowledge-assisted AI brings significant benefits to the AI learning process learning to hybrid systems that attempt to capture the advantages of both learning and non-learning traditional systems. Section 5.4 discusses the corresponding approaches and perspectives.

Finally, a crucial step towards automation in an AI-human environment is co-learning. As discussed in Section 5.5, frameworks and methods for enabling a bidirectional exchange of information between humans and AI systems are explored, facilitating a fruitful exchange of knowledge that ultimately improves the overall learning process.

5.1. RL-BASED APPROACHES

In this part, we revise RL techniques that can be adopted in the AI4REALNET project and we discuss how we can push the boundaries on these fields to reach the scope of the project.

5.1.1. DISTRIBUTED REINFORCEMENT LEARNING

Distributed Reinforcement Learning (DRL) extends the classical formulation of Reinforcement Learning by distributing the learning process across different agents that simultaneously act in the environment. It can be considered as a special case of the more general Multi-Agent Reinforcement Learning (MARL, Shapley, 1953a; Albrecht et al., 2024). Indeed, there are three main settings of MARL problems: (i) *cooperative*, in which all the agents cooperate to optimize the same shared objective (i.e., team games), (ii) *competitive*, in which all the agents compete (i.e., zero-sum games), and (iii) *mixed*, which combines the previous two (i.e., general-sum games). Distributed RL can be seen as a *cooperative MARL* problem. Several peculiarities distinguish DRL from standard RL, like credit assignment and the different learning goals, as it is difficult to track how agents have contributed to the different rewards.

Cooperative problems require implementing communication between agents to achieve the same objective efficiently. Agents can share different types of information: (i) sampled data, e.g., observations and actions; (ii) predicted data, e.g., Q values; (iii) knowledge, e.g., model parameters. Depending on how agents communicate, we can distinguish three main approaches to DRL. During training and inference, agents can access some centrally shared mechanism or information between them. For example, a single central agent may receive information from all other agents and dictate the actions of the agents. This is referred to as *centralized training and execution*. Conversely, in *decentralized training and execution*, each agent can ignore the existence of other agents and learn its optimal policy in a completely local way using single-agent RL techniques. Alternatively, a mix of the previous two in which agents are trained with a centralized approach and their policies are executed fully decentralized (*centralized training and decentralized execution*). An example of a state-of-the-art algorithm that uses centralized training and execution can be found in (Yu et al., 2022), where an extension of a popular actor-critic RL algorithm is introduced with a centralized critic that is shared among agents. Typically, a centralized approach can offer more efficient agent coordination and theoretical guarantees to convergence (Albrecht et al., 2024). This is why almost all state-of-the-art algorithms require centralization, at least during training.

The mathematical frameworks to formalize DRL may vary given the several dichotomies above. We can indeed consider Markov Games (MG, Section 2.2.1), Partially Observable Markov Games (POMG, Section 2.2.2) and Decentralized Partially Observable Markov Decision Processes (Dec-POMDPs, Section 2.2.3), depending on the characteristics of the problem.

The main reason for distributing the learning process is to break down the complexity of the RL problem at the cost of introducing some bias. In particular, the goal of DRL is to mitigate the so-called *curse of dimensionality*, as large state and action spaces have a negative impact on the *sample complexity* of the learning algorithms (i.e., the number of samples needed to learn how to behave in a problem with a given precision).

All the problems presented in Section 3 as use cases of the AI4REALNET project are affected by the curse of dimensionality. As an example, consider that a configuration of the Grid2Op environment for simulating a realistic power grid benchmark counts 118 substations, 186 powerlines, 99 loads, and

62 generators, each contributing to the state with more than one continuous variable.[9] The problem is more relevant when we move to real-world applications. For example, the French power grid is composed of more than $25,000$ substations and $10,000$ powerlines.

Having a central information structure or controller can be problematic or unfeasible in many situations, including the use cases of AI4REALNET. To break the curse of dimensionality, ideally, we would like to make only local information available to the agents, not on the entire network, thus avoiding a centralized approach. For this reason, a trade-off is necessary between a fully centralized approach with complete availability of information and a fully decentralized approach with limited availability of information among agents. This explains why many state-of-the-art algorithms are not considered appropriate for these applications.

Research Directions In the context of AI4REALNET, a first research direction may consider developing algorithms that, given the original MDP $\mathcal{M}$, find different subsets of it such that each subset can be considered almost as smaller and more tractable MDP which can be solved in a decentralized fashion. More formally, $\mathcal{M}$ can be defined as a set of MDPs $\mathcal{M} = (\mathcal{M}_i)_{i=1}^m$, each defined as $\mathcal{M}_i = \langle \mathcal{S}_i, \mathcal{A}_i, P_i, R_i, H \rangle$. The objective would be to discover each $\mathcal{M}_i$ and distribute the learning process across them. Information-theoretic concepts such as mutual information can be used to analyze how state and action variables relate to each other. Each MDP $\mathcal{M}_i$ would be composed only by those variables that are informative about each other. To the best of our knowledge, no previous work is trying to tackle DRL problems with such an approach. Most similar works revolve around feature selection in the sense that a subset of features of the original state and action space is selected to solve the original MDP. For instance, (Castelletti et al., 2011) addresses variable selection in high-dimensional real control problems by proposing an algorithm that, starting from the set of variables needed to explain the reward, recursively selects the other variables based on a statistical measure of significance that accounts for non-linear dependencies. (Beraha et al., 2019) shows that using mutual information for feature selection in supervised learning allows direct control of the ideal prediction error, thus providing a theoretical argument that could also benefit an extension of that work to our DRL decomposition.

A second line of research in the context of the AI4REALNET project can look at the problem from different perspectives. Indeed, we know that, in the real world, searching for fully isolated problems is not realistic. A new high-level agent could be added to the set of agents whose task would be to communicate contextual information related to the environment they do not observe to the agents under its supervision. Possible lines of research may thus investigate the best way to communicate between this high-level agent and the agents under its supervision.

[9] See `grid2op.readthedocs.io`.

5.1.2. HIERARCHICAL REINFORCEMENT LEARNING

Hierarchical Reinforcement Learning (HRL) is another extension of the classical RL formulation, in which the sequential decision-making problem is decomposed into a hierarchy of simpler subtasks. A high-level policy learns to perform the main task by choosing optimal subtasks as the higher-level actions. A subtask may itself be a standard RL problem with a lower-level policy, which, in turn, learns to solve it (Pateria et al., 2021).

The main reason for creating a hierarchy of subproblems is to decompose the long horizon of the original task into multiple shorter horizons in terms of the sequences of the subproblems. A long horizon may, for example, create issues in the credit assignment and cause slow and inefficient learning (sometimes referred to as *curse of horizon*). HRL can alleviate such issues by providing temporal abstraction, i.e., each subtask is a higher-level action that lasts for a longer timescale compared to a lower-level action and is associated with a collective reward related to a specific subgoal.

All the problems presented in Section 3 of the AI4REALNET projects are affected by the curse of the horizon. As an example, consider that each time step of the time series that is injected into the Grid2Op environment lasts for five minutes (Serré et al., 2022; Manczak et al., 2023), thus resulting in a one-day scenario of 288 time steps in which the entire power grid should always be perfectly balanced without line overloads.

The mathematical framework used to formalize HRL problems is Semi-Markov Decision Processes (SMDP, section 2.1.4). A possible formalization of temporally extended actions is the *options* framework (Sutton et al., 1999). Each option solves a specific subtask. Formally, an option is a triple $o = (\mathcal{I}_o, \beta_o, \pi_o)$, in which $\mathcal{I}_o, \beta_o$ represent initiation and a termination condition respectively, i.e., a set of states selected by the high-level option in which the subtask starts or ends, π_o is the low-level policy that learns to solve the subtask. Options collectively create the space of all possible subtasks of a given problem, called the subtask space.

The theoretical objective of HRL problems is the composition of two different objectives. One is to learn a *hierarchical policy*, i.e., a mapping from states to lower-level actions accounting for the entire hierarchy, that maximizes the expected cumulative reward conditioned on a subtask space. The other is *subtask discovery*, i.e., find a subtask space that provides the best possible conditioning. Learning a hierarchical policy typically involves learning a fixed or variable number of lower-level policies associated with a given set of tasks. Subtask discovery can happen concurrently with or independently from learning a hierarchical policy. In the alternative, subtasks can be handcrafted, for instance, defined by specific real problems.

There are several main issues peculiar to HRL problems, for instance, the non-stationarity induced by the simultaneously changing policies at different levels of the hierarchy and the complexity of learning at various levels, i.e., how to propagate reward between levels, how to decompose the value function,

how to design state/action space at each level.
An example of a popular work is (Dietterich, 2000), in which the original MDP is decomposed into a hierarchy of smaller MDPs and, consequently, the value function of the original MDP is decomposed into an additive combination of the value functions of the smaller MDPs. The authors also provide algorithms with convergence guarantees based on such decomposition.

Research Directions The AI4REALNET project involves developing reinforcement learning algorithms to be deployed in critical infrastructures. These environments are often characterized by complex continuous state-action spaces. In the current literature, hierarchical reinforcement learning (HRL) approaches are mostly based on value-based methods. These methods, however, struggle when scaled to large or continuous state and action spaces.
On the other hand, a significant portion of the reinforcement learning (RL) literature focuses on a different class of approaches that are naturally suitable for continuous state-action spaces, namely, *policy search* (PS). Surprisingly, these techniques are rarely adapted for dealing with hierarchical structures. Therefore, as a research direction, we identify the generalization of PS approaches to hierarchical architectures. Unlike value-based approaches, PS explicitly represents the policy and, possibly, the value function (actor-critic approaches).
Here, we propose to develop approaches in which a hierarchy of policies is defined. Policies are intended to achieve specific goals at the lowest level, while at higher levels, they are intended to activate the suitable low-level policy.

5.1.3. INVERSE REINFORCEMENT LEARNING

Reinforcement learning is a powerful paradigm enabling agents to autonomously learn tasks through interaction. However, RL often requires iterative refinement of a reward formulation to allow the agent to fulfill its goal. One of the primary reasons for RL systems failing to perform on a given task can be traced back to a misleading reward definition, which regulates both the exploration of an agent (Fortunato et al., 2019; Colas et al., 2018) and its learned behavior along with the convergence speed (Laud and DeJong, 2003; Dong et al., 2020). To overcome such limitation, we can adopt Inverse RL (IRL) solutions, focused on determining the underlying reward function that an agent is optimizing, given its observed behavior. Indeed, unlike traditional RL, where the objective is to learn a policy to maximize a known reward function, IRL aims to infer what drives an agent's actions by observing its interactions with the environment. This approach is particularly useful in scenarios where the reward function is not explicitly provided but must be deduced from expert demonstrations. By understanding the rewards that guide expert behavior, IRL enables the creation of agents that can mimic complex decision-making processes, improving their ability to perform tasks in a manner consistent with expert strategies (Ng et al., 2000). The objective of IRL is to leverage these demonstrations to bridge the gap between observing how an expert performs a task and formalizing the underlying motivations that drive their decisions.

IRL is generally applied in domains where the reward formulation is complex and often encompasses multiple requirements, such as robotics (Vasquez et al., 2014; Krishnan et al., 2019), complex games (Muelling et al., 2014) and self-driving vehicles (Shimosaka et al., 2014; You et al., 2019).

To infer a reward function from the available demonstration, various algorithms like *Maximum Margin* IRL (Ng et al., 2000), *Bayesian* IRL (Ramachandran and Amir, 2007) or *maximum entropy* IRL (Ziebart et al., 2008) can be applied. Maximum Margin IRL requires optimal demonstrations, and it emphasizes a clear distinction between optimal and sub-optimal actions through margin maximization in its reward predictions. Bayesian IRL, at the cost of additional computation, eases the need for optimal data by incorporating prior knowledge and leveraging probabilistic methods to handle uncertainty, resulting in more robust and reliable modeling of an agent's behavior even when data are noisy or sub-optimal. Finally, maximum entropy IRL ensures diversity and generality in behaviors by maximizing the entropy of actions based on the observed data.

In general, IRL methods exhibit considerable computational complexity and suffer from issues related to the availability and quality of data. Consequently, their efficacy is strongly influenced by the quality of the expert demonstrations used. To mitigate these limitations, (Michini and How, 2012; Levine et al., 2011) extend Bayesian IRL to approximate unseen states, which improves the generalization capability of the model by making better use of the available data. Similarly, (Shiarlis et al., 2016; Audiffren et al., 2015; Boularias et al., 2011) extend Maximum Entropy IRL to handle imperfect trajectories and enhance the method's reliability. Additionally, *Adversarial* IRL enables the training of a policy based on the expert's data (Ho and Ermon, 2016; Finn et al., 2016) by framing the learning process similarly to how it is done in a Generative Adversarial Network. Here, the policy acts as the generator and is considered converged when the discriminator can no longer distinguish between generated and true expert data.

Research Directions The use cases addressed by the AI4REALNET project are based on 3 complex real-world domains presented in Section 3: (i) *Power Grids*, (ii) *Railway Networks* and (iii) *Air Traffic Management*. Each domain has its own digital environment, summarised in Figure 2, which will then be used to train an AI agent. Despite their differences, these environments share multiple constraints and complex objectives, further complicating the design of a reward formulation.

While designing a reward formulation that captures the expected behavior in a simple task can already be challenging, doing so for a real-world task requires addressing additional complexity due to multiple constraints, objectives, and safety requirements. Thus, while the project already incorporates *Human-in-the-loop* techniques to leverage human expertise during the training phase of an AI agent, incorporating IRL methods could promote greater reliability and robustness in behaviors that are crucial for the successful deployment of AI into the real world. While IRL enables the development of human-inspired policies that may facilitate human acceptance and foster trustworthiness, it also introduces new challenges and concerns that must be addressed. Its scalability is limited by computational complexity and the challenge of representing high-dimensional state and action spaces. The data collection phase must adhere to requirements, including ethical considerations, and depend on human availability. Additionally, further investigation is needed to explore the integration of human-AI co-learning based on human preferences and its impact on IRL techniques.

5.2. EXPLAINABLE AI

Integrating AI into real-world systems necessitates understanding AI choices to ensure human trust. Without trust, AI decisions may be disregarded (Ahn et al., 2021). This concept is further explored in Section 5, which discusses the human perception of human-AI interaction. Trust becomes especially important in critical systems where the complexity of black-box models and their lack of transparency can lead to potential rejection by users. Consequently, in the AI4REALNET project, the explainable AI (XAI) component is pivotal for supporting human-AI interaction and fostering trustworthiness on the human side.

As described in recent and popular survey papers (Barredo Arrieta et al., 2020; Das and Rad, 2020; Minh et al., 2022), explainable AI is a large and growing subfield of AI. Explainable AI concerns itself with explaining the models learned using AI techniques and explaining the decisions of those models. Such explanations can serve several purposes. For example, a better understanding of a learned model and its decisions can help find bugs in training algorithms, identify hidden biases, provide recourse to adversely affected users, and assess trustworthiness. As attention to AI ethics and regulation grows, many of these factors become increasingly important.

As a relatively young sub-discipline of AI, there are many different approaches to explainability, and the terminology to describe such approaches differs between authors. An important class of models is the collection of models that are understandable by themselves due to their simple and structured internal structure. Such models are referred to *intrinsically* explainable (Das and Rad, 2020), *transparent* (Barredo Arrieta et al., 2020), or *inherently interpretable* (Minh et al., 2022). Such models are often deemed to need no further explanation. On the other hand, more complex models, such as deep neural networks, are not understandable by themselves. For such models, *post-hoc* (Barredo Arrieta et al., 2020; Das and Rad, 2020) or *post-modelling* (Minh et al., 2022) explanations can be constructed. This typically involves separate methods that aim to communicate how an already trained model makes its decision, e.g., visually or in language.

Explanations can furthermore be divided based on whether explanations are local or global. Where global explanations aim to make the entire input-output characteristic more understandable, local methods focus on interpreting the decision by the model for a specific input value. Yet another important criterion is whether a given explanation technique is model-specific or model-agnostic, with the first class developed to work only with a certain class of models (e.g., convolutional neural networks), whereas the latter has no such restrictions.

In modern machine learning practice, considerable research effort is devoted to post-hoc, model-agnostic explanations (Das and Rad, 2020), with local explanations popularly chosen for deep learning models (Barredo Arrieta et al., 2020). One reason for this is that intrinsically understandable models often suffer from an interpretability-accuracy trade-off (Das and Rad, 2020; Minh et al., 2022; Barredo Arrieta et al., 2020), in those models which are not intrinsically understandable, such as deep neural networks often perform better. Modern deep neural networks also have highly complex input-output mappings, making global explanations daunting. Similarly, in this position paper, we will focus on several methods in the categories of feature importance methods and counterfactual explanations and discuss how explanations can be illustrated using visualization-based techniques.

In the context of the AI4REALNET project and towards explainable AI we plan to investigate the use of visualization-based explanation (sec. 5.2.3) to provide counterfactual and contrastive explanations (sec. 5.2.1) as well as feature importance based explanations (sec. 5.2.2).

All three of these aspects share a number of challenges that pose research questions within the AI4REALNET project. Evaluation of explanations is a remaining challenge across the field, and the use cases with collaborative tasks for AI and human agents could provide a useful benchmark. Furthermore, most post-hoc explanation methods have been developed for supervised learning settings, leaving a research gap in developing and evaluating such methods in the deep reinforcement learning setting or similar settings requiring planning and decision-making.

5.2.1. COUNTERFACTUAL AND CONTRASTIVE EXPLANATIONS

Counterfactual and contrastive explanations are both based on a comparison between actual and hypothetical input values. Sometimes, a distinction is drawn with contrastive explanations focusing on why label y_1 is given rather than some label y_2, while counterfactual explanations focus on which changes to input x_1 would cause a change in the predicted label y_1 (Stepin et al., 2021). However, the notions are often used interchangeably, so Stepin et al. (2021) propose using the term 'counterfactual explanations' to refer to both terms jointly.

Counterfactual explanations have the potential benefits of being easier to understand or interpreted by users (Stepin et al., 2021; Barredo Arrieta et al., 2020), in particular, being preferred by users over case-based reasoning (Verma et al., 2020). Furthermore, counterfactual explanations are often 'actionable', meaning they describe which change of inputs would lead to the prediction of a (desired) label (Verma et al., 2020; Stepin et al., 2021). Furthermore, human explanations are usually contrastive (Miller, 2019).

While survey papers such as the ones by Stepin et al. (2021) and Verma et al. (2020) discuss many works on counterfactual explanations, many of these focus on supervised learning settings, and only a few of these focus on decision-making scenarios such as planning or reinforcement learning. The papers that do focus on such a scenario are all model-specific.

Such papers include the work by Kim et al. (2019), who describe the numeric differences between two sets of plan traces using a contrastive explanation using temporal logic. Another approach is taken by Chakraborti et al. (2017) and Sreedharan et al. (2018), who take into account an operator's model when explaining why a plan A is optimal rather than a proposed plan B - by addressing misconceptions in the internal model of the system user or by tuning the explanation to the level of expertise of the user. In the case of the second paper, these explanations are provided in linguistic form.

Lastly, Sukkerd et al. (2018) and Zhao and Sukkerd (2019) consider multi-objective scenarios. The first proposes an explainable planning framework that, in particular, can explain trade-offs in a multi-objective framework by contrasting solutions to reasonable Pareto-optimal alternatives. The second can answer questions of the form 'why X' or 'why not Y' by contrasting the proposed plan to an alternative plan where some action X does not (or some action Y does) occur. In both cases, explanations are offered in linguistic form.

Research Directions Counterfactual and contrastive explanations tend to be easy to interpret and actionable. Notwithstanding the recent advances in counterfactual and contrastive explanations, the field still contains many open challenges. Considering the common challenges above, this type of explanation has only been provided by model-specific methods in decision-making problems, leaving model-agnostic methods in this setting a research gap. Furthermore, the actionability of explanations can be further improved, e.g., by considering which attributes are most easily changed or by providing a course of action that could be taken to yield the desired label (Verma et al., 2020).

5.2.2. FEATURE IMPORTANCE METHODS

Feature attribution methods are the most extensively explored explanation techniques in literature (Samek et al., 2021). Hence, attribution methods play a pivotal role in understanding the decisions of machine learning models, offering insights into the factors driving their predictions. These methods can be categorized as post-hoc methods and aim to attribute importance to individual input features, shedding light on which features contribute most significantly to the model's output. Feature attribution methods aid in discerning whether a feature has exerted a positive or negative impact on a model's prediction. Moreover, the significance of individual features can be deduced from the absolute values of the feature attributions. By unraveling the black-box nature of complex models such as deep neural networks, feature attribution enhances interpretability and transparency, enabling users to comprehend and trust model decisions.

These methods can be broadly categorized into perturbation-based approaches and gradient-based approaches. Perturbation-based methods analyze the effect of perturbing individual input features on the model's output. Input Perturbation, for example, selectively masks or alters input features and measures the resulting variations in model outputs, thereby attributing importance to different features based on their impact on prediction. Prominent examples of perturbation-based methods are LIME and SHAP. Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016) generates local surrogate models around specific instances by perturbing input features and observing changes in predictions, providing interpretable explanations. Shapley Additive Explanations (SHAP) (Shapley, 1953b), on the other hand, perturbs input features to compute Shapley values, which represent the average contribution of each feature across all possible coalitions of features, offering a principled approach to feature attribution. Similarly, Feature permutation (Fisher et al., 2019) involves systematically shuffling or permuting individual features to assess their impact on model performance, thereby attributing importance to each feature based on its influence on prediction accuracy.

Gradient-based methods rely on the gradients of the model's output concerning its input to attribute feature importance. By computing these gradients, methods such as Integrated Gradients (Sundararajan et al., 2017) attribute importance to input features based on their impact on the model's predictions. It integrates the gradients along the path from a baseline to the input, highlighting features that contribute most significantly to model output changes. Similarly, Saliency methods (Simonyan et al., 2013) visualize the gradients as heatmaps, emphasizing regions of the input space that have the strongest influence on the model's decision and thereby offering intuitive explanations of model behavior. Additionally, Input X Gradient (Shrikumar et al., 2016) multiplies input features by their gradients, emphasizing features with higher gradients as more influential.

The development of new methods in XAI is ongoing, with the aforementioned methods representing just a subset. However, evaluation remains a critical factor in assessing the reliability and effectiveness of these techniques. While qualitative assessments are common, they often lack objectivity and comparability between different methods. Evaluation metrics have been studied to achieve a quantitative evaluation. They offer direct quantification of explanation quality across various dimensions. These metrics, such as faithfulness, robustness, complexity, randomization, and axiomatic properties, provide comprehensive assessments of explanation methods' performance. Faithfulness measures the consistency of explanations with model predictions, while robustness assesses their stability under input perturbations. Complexity evaluates the conciseness of explanations, randomization gauges their resistance to data and model randomness, and axiomatic metrics ensure adherence to specific properties. These evaluation criteria are pivotal for guiding the development and adoption of XAI methods, fostering transparency and trustworthiness in AI systems.

Research Directions While feature attribution methods are widely applied for supervised learning tasks, challenges remain to be addressed. To remedy the mentioned gap of applications to DRL agents, in the context of the AI4REALNET project, we plan to develop and examine multiple feature attribution methods designed to explain the decisions made by DRL agents for critical applications such as power grid control. Additionally, we will provide a set of relevant metrics for evaluating various facets of explainability, including robustness, complexity, randomization, and faithfulness, to shed light on the effectiveness and nuances of different attribution techniques in power system control contexts.

5.2.3. VISUALIZATION-BASED EXPLANATIONS

Visualization-based explanations aim to fill the gap between human understanding and AI decision-making by providing a graphical representation of the explanations. Saliency maps are a popular visualization technique in classification tasks that highlights the importance of different regions on graphical inputs. These maps are widely used to provide contextualized local explanations for decisions made by opaque models (Alqaraawi et al., 2020; Petsiuk et al., 2021; Simonyan et al., 2013; Liu et al., 2016). Their widespread adoption can be attributed to their inherently human-friendly communication nature, which makes them understandable from a human perspective even without the need for much expertise in the problem addressed.

In RL, (Vouros, 2022) argues that explanations can be offered by identifying and highlighting the most influential features in the decision-making process. Conversely, (Atrey et al., 2019) underscore the limitations of saliency maps in establishing a causal relationship between observed features and the resulting decisions of agents.

To combine the efficacy of saliency maps while overcoming their limitations, previous work in Explainable RL has focused on developing explanations through graphical depictions for the decision-making process, each addressing distinct objectives. (Olson et al., 2021) generate a counterfactual state representation to highlight the minimal changes required in the observation to choose a different action. Additionally, (Dorfer et al., 2022) and (Fuxjäger et al., 2023) contextualize the agent's decision by highlighting the areas of the observed space that require immediate attention by the agent, as well as by forecasting the expected consequences of a decision on the system. (Sequeira and Gervasio, 2020) and (Amir and Amir, 2018) summarize the agent's behavior on a trajectory through a short video highlighting key agent-environment interactions. Finally, (Madumal et al., 2020) proposes visualizing the agent's decision with a causal graph. This method significantly improves the transparency of the algorithm by providing a means to understand how and why models make their decisions.

Research Directions While visualization defines the means of communication, the content of the explanation must be defined based on the domain. Representing the domains in a common way can help abstract the low-level details intrinsic to each domain, thereby standardizing the explanation methodology. The AI4REALNET project will contribute to visualization enhancements by proposing an extension of existing visualization libraries for power grids. This extension will integrate feature attribution values into the interactive graphical representation of power grids, enhancing the interpretability and usability of the framework. Interactive visualizations will be developed to represent the decision-making processes of RL agents in power system control scenarios. Such graph representation from the power grid domain (Fuxjäger et al., 2023) will be extended to the other domains. While this intuitively applies to the *railway network*, some challenges may arise when representing the *air traffic management* through a graph. Modeling the network as a graph allows a human to navigate the representation and obtain a broader or more detailed view of the system, depending on their needs.

5.3. AUTOMATION TRANSPARENCY

5.3.1. THE HUMAN SIDE OF HUMAN-AI COLLABORATION

Automation is shifting functions from humans to technology. This changes the human role in achieving work objectives. It is a well-known problem that humans often get a role they cannot fulfill (Bainbridge, 1983). With AI, this problem has not disappeared. Rather, it increases because AI automates sophisticated cognitive functions (Endsley, 2023b). AI outputs that are not understandable for humans are only one of the related problems. Others are that humans lose skills and tacit knowledge when functions are allocated to AI, and the humans no longer perform them themselves. Or they rely too much on an AI and therefore do not use it appropriately. To avoid such negative effects of automation, the AI, the human-AI interaction, and the human-AI collaboration need to be carefully designed. Moreover, an appropriate design may not only prevent negative effects but even enable synergies between humans and AI, creating a symbiosis. This is because a clever combination of knowledge from both humans and technology, based on their complementary strengths and weaknesses, can result in human-AI performance that outperforms any human and any AI individually (e.g., Grote et al. 1995).
Against this background, the following chapter focuses on the human side of human-AI collaboration and the resulting requirements for automation transparency. It goes even deeper on the human side by approaching a collaborative new trend that perceives the human implicitly to enhance the symbiosis between both actors (human and AI). The first section discusses the transparency requirements from a content perspective. Five key human processes are described: human decision-making, human

learning, human trust, human motivation, and human implicit status. The main characteristics of each process are elaborated on, and the requirements for automation transparency are derived. The second section focuses on methodological aspects. Two complementary methods of cognitive (system) engineering are described: Ecological Interface Design (EID) and the Joint Control Framework (JCF). These methods are suitable for analyzing and designing human-AI interaction and human-AI collaboration to derive consequences for both AI design and human task design. The final section discusses proposals for future research concerning AI4REALNET, i.e., in critical network control.

5.3.2. KEY HUMAN PROCESSES

HUMAN DECISION MAKING

Today's AI-based decision-support systems are mainly based on recommendations. However, recommendations provided by AI are usually not sufficient, even if they are enriched using explanations (XAI) and transparency (Eisbach et al., 2023; Miller, 2023). Several studies showed that explanations do not lead automatically to better decisions (Ngo and Krämer, 2022; Zhang et al., 2020). Therefore, rather than just providing decisions, joint human-AI decision-making based on the complementary capabilities of humans and AI is required (Endsley, 2023b; Miller, 2023). From a psychological perspective, joint decision-making needs to consider the human decision-making processes with its cognitive elements as well as with its related biases such as the anchoring effect or the confirmation bias (Eisbach et al., 2023; Ha and Kim, 2023; Wang et al., 2019). In this chapter, the peculiarities and elements of human decision-making are described.
Human decision-making is diverse and varies depending on the human's level of experience (kle, 1993). Basically, two types of decision-making can be distinguished: analytical and pattern recognition (Kahnemann, 2012; kle, 1993). Novices with low experience make decisions rather analytically by comparing options and arguments thoroughly, leading to time-consuming, deliberate decisions. In contrast, experienced experts tend to make decisions quickly and largely effortlessly and without conscious control (Kahnemann, 2012; kle, 1993). This is known as naturalistic decision-making and describes how experts make decisions in real-life situations, especially in difficult circumstances (Klein, 2008; kle, 1993). Common difficulties include time pressure, high stakes and high risks, dynamic conditions, and incomplete, contradictory, and ambiguous information. In addition, real-world goals are often ill-defined and conflict with other goals. In such situations, experts - in contrast to inexperienced novices – make decisions using experience-based pattern recognition, which is faster than the analytical procedures and more informed (Klein et al., 2003; kle, 1993).
In more detail, the cognitive process of decision-making is complex and encompasses many different macrocognitive functions and sub-processes (Klein, 2018; kle, 1993). Macrocognition is defined as a process of "adapting cognition to complexity" (Hoffman et al., 2009)[p. 87]. In other words, it is a

dynamic application of thinking to what is happening in a complex environment.

Macrocognition encompasses functions and processes (Klein et al., 2003), regardless of whether in individuals, human-human teams, or human-AI teams. Functions represent actions rather than states and must be performed in real-world contexts (Klein, 2018). Decision-making emerges from six macrocognitive functions, which are always active in parallel rather than being performed sequentially: 1) Permanently decisions need to be taken. 2) Sensemaking organizes perceptions into meaningful units. Situation awareness arises as a result of sensemaking (Klein et al., 2003). 3) Plans are adapted to changing circumstances through (re)planning. 4) This requires an adaptation, which allows one to respond flexibly to new situations. 5) Problem detection identifies potential challenges. 6) Coordination synchronizes activities and resources. Although these six functions are distinct, they are interrelated. In cognitive work, no task depends exclusively on a single function or process (Klein, 2018).

These macro-cognitive functions are supported by macro-cognitive processes such as 1) maintaining common ground (with the relevant humans or/and AI), 2) developing mental models (of the current situation and the limitations and capabilities of the AI), 3) mentally stimulating and story building (of possible actions), 4) managing uncertainty and risk, 5) identifying leverage points to control the process, and 6) managing attention to be aware of potential problems (Klein, 2018).

These macro-cognitive functions and processes are distributed, including individuals, teams, and even organizations. A decision made somewhere affects other individuals, teams, or organizations, often without anyone even being aware of the others and the interrelated effects (Wäfler and Rack, 2021).

The anchoring effect describes a remarkably robust cognitive bias that influences human judgment and so decision-making (Furnham and Boo, 2011; Pohl, 2006). It describes the phenomenon that initially presented information “anchors” people’s attention and perception, making them blind to other information (Tversky and Kahneman, 1974).

Confirmation bias refers to the tendency to seek confirmation of one’s own assumptions by selectively searching for, interpreting, and remembering information in a way that systematically hinders the possibility of rejecting one’s own assumptions (Pohl, 2006). Confirmation bias, therefore, leads to information that contradicts one’s own assumptions being neglected, which causes distorted decisions. To overcome the confirmation bias in human-AI collaboration, Ha and Kim (2023) suggests providing the human a priori information (e.g., a set of data that is taken into account when computing decisions) before showing the final decisions generated by AI. According to these authors, this might be the only way to overcome the confirmation bias effectively. In contrast, there are still no ways to fully overcome the anchoring effect when AI suggests recommendations (Pohl, 2006; Wilson et al., 1996; Furnham and Boo, 2011).

An AI supporting the human decision-making process must provide transparency regarding the macro-cognitive functions and processes (e.g., by providing an explanation regarding the state of the process to be controlled, regarding emerging problems, or regarding leverage points) as well as regarding biases in human decision-making (e.g., by mirroring patterns in human decision-making behavior).

HUMAN LEARNING

Human learning is a complex process that leads to lasting changes in humans, influencing their perceptions of the world and their interactions across physical, psychological, and social dimensions. It is fundamentally shaped by the ongoing, interactive relationship between the learner's characteristics and the learning content, all situated within the specific environmental context of time and place, as well as the continuity over time (Alexander et al., 2009). Humans effectively learn through a dynamic process described in the Experiential Learning Theory, conceived by Kolb (1984). This approach describes a cyclical four-stage model essential for converting experiences into substantive knowledge. The four iterative stages encompass the following:

- Concrete Experience: This initial phase involves hands-on engagement with tasks, where humans encounter new experiences or reinterpret existing ones, laying the foundation for learning. This requires the preconditions and the ability to be fully open-minded and unbiased to new experiences.

- Reflective Observation: In this stage, humans reflect on their experiences, contemplating what was successful or identifying potential improvements. This reflective practice is critical for internalizing learning outcomes. This means humans must look at and reflect on their experiences from many perspectives.

- Abstract Conceptualization: Humans then conceptualize their reflections, crafting new ideas or adjusting existing mental models. This phase is where abstract understanding materializes, enabling the construction of novel mental models or conceptual frameworks. For these, humans should be able to create concepts that integrate their observations into comprehensive cognitive representations.

- Active Experimentation: The culminating stage involves applying abstract cognitive representations in real-world scenarios to observe the outcomes of these applications. Taking into account real-world feedback is key for empirically testing ideas and refining mental models. With active experimentation, humans should be able to use their cognitive representations to make decisions and solve problems.

This cyclic process is characterized by a sequential learning progression, emphasizing the necessity of engaging with each stage systematically to ensure a thorough learning experience. This sequence facilitates the conversion of experiences into actionable knowledge through a recurring cycle of practice, reflection, model development, and experimentation (Kolb and Kolb, 2009).

In the context of human-AI collaboration, there are at least three learning objects that are important for humans: (1) learning about the process and the task (i.e., about the subject matter of decision-making, e.g., Air Traffic Management (ATM)), (2) learning about the AI (e.g., about the AI's capabilities and boundaries), and (3) learning about one's own behavior (e.g., about own decision-making patterns and biases).

Learning about the process and task Acquiring more expertise in professional decision-making means improving performance in the macrocognitive functions and processes of decision-making. Appropriate learning improves task performance (Klein, 2018). For instance, in problem detection – a critical aspect of macro-cognitive functions – AI could empower humans to act more proactively by enabling them to systematically go through the entire cycle of experiential learning. Thus, they make new experiences, conceptualize their reflections, craft new ideas, or adjust existing mental models, thereby learning to detect potential problems earlier and more frequently.

Learning about the AI To improve performance, it is important to know the collaboration partner (i.e., the AI) and to have an accurate mental model of it (Gao et al., 2023; Wilkison et al., 2007). Adapting Rook (2013)'s definition of mental models to the AI context, a mental model is an individual's cognitive representation of the functioning of an AI-based on experiences with it. This model significantly influences how individuals interact with the AI, anticipate its behavior, and make decisions.

For humans to have an accurate mental model of a particular AI, they need to be able to continuously update their mental model – a process that is essential for learning (Stöttinger et al., 2018). Updating a mental model involves adjusting the understanding of the AI by continuously incorporating new information and experiences (Uitdewilligen et al., 2013; Valadao et al., 2015). Over time, this dynamic learning process is essential for generating accurate mental models of AI.

Learning about oneself Various cognitive, perceptual, and motivational biases often affect human judgment and decision-making. Interestingly, while humans can better identify biases in others, they frequently fail to recognize their own, a phenomenon called the "bias blind spot". It is primarily attributed to two factors: an overreliance on introspective evidence, which is unreliable due to the nonconscious nature of biases, and a conviction that one's own perceptions are accurate reflections of reality, leading to the belief that differing perspectives are biased (Pronin, 2007).

Humans' tendency to deny their own biases, even while recognizing biases in others, reveals a deep limitation in self-awareness. AI can promote self-awareness by enhancing and developing human self-reflection (Jelodari et al., 2023). Real-time feedback on behavior and decision-making styles can provide insightful information and reveal patterns and biases that may not be obvious to individuals (Lieder et al., 2019). The benefits of AI in overcoming human biases could be expanded by providing a nuanced understanding of how an individual's behavioral tendencies may change under different conditions, such as different times of day (e.g., taking more risks in decisions at the end of a shift). In this way, AI offers a personalized reflection of an individual's cognitive and behavioral patterns.

In summary, integrating experiential learning theory into human-AI interaction is critical to enhancing human expertise. This enables humans to create precise mental models of the decision-making subject and the AI, enabling better collaboration with AI. In addition, deepening self-awareness and self-reflection are important aspects of effective decision-making and overcoming biases. This holistic approach to learning enables humans to realize AI's full potential and enhances human capabilities.

Regarding these three learning areas, an AI that supports human learning must provide transparency across the four stages of the learning cycle: Transparency 1) regarding concrete experience (e.g., by triggering exploration through providing explanations regarding different factors of the process that can be related to each other), 2) regarding reflective observation (e.g., by triggering reflection through animating the human to formulate hypotheses regarding the interrelations of different factors of the process), 3) regarding abstract conceptualization (e.g., by explaining data-based evidence for and against hypothesis the human formulated), and 4) regarding active experimentation (e.g., by supporting the human to explore learnings in real-world scenarios and different contexts without negative consequences and by providing immediate feedback on the outcomes).

HUMAN TRUSTING

Human trust in a particular AI is a combination of the human's knowledge, beliefs, emotions, and experiences with that AI. These factors shape human expectations of the reliability and effectiveness of an AI. Such expectations lead to a positive or negative assessment, which determines humans' trust in the AI. In essence, the construct of trust encompasses the degree of confidence a human has in the automated system's ability to perform accurately and cooperatively in various contexts (Cahour and Forzy, 2009). In contrast to trustworthiness, an attribute of a particular AI, trust is a dynamic process that builds on trustworthiness but does not directly result from it (Hoffman, 2017). Actual trust in AI is influenced by various other aspects (Kaplan et al., 2023), including personal experience with AI (Hoffman et al., 2018).

In general, humans need to establish appropriate trust in an AI, which mainly means having a realistic understanding of the AI's boundaries or the AI's scope of application. Appropriate trust means that, with increasing experience a human should trust the AI for specific task or objectives in certain contexts or problem scenarios, while also appropriately mistrust the AI for other tasks or objectives in specific contexts or problem situations. Consequently, trust does not develop in the traditional sense by gradually increasing to a more advanced state. Instead, it morphs, meaning that the state of trust changes without necessarily developing or improving linearly. As humans interact with the AI over time, their trust morphs in response to new hints (e.g., about the AI's error bounds), experiences, or contexts and becomes appropriate trust. When an appropriate trust is established, humans can confidently rely on the AI (Hoffman et al., 2018).

The consequences of inappropriate positive trust and inappropriate negative trust have serious consequences, especially in high-stakes environments (Lee and See, 2004). Highlighting the nuanced dynamics of these relationships, Parasuraman and Riley (1997), along with contributions from Parasuraman and Manzey (2010), delve into the complex interplay between positive and negative trust in the realm of human-automation interaction. They emphasize the critical influence that these states of trust have a critical impact on the effectiveness of human-AI collaboration. In particular, they note that excessive mistrust can lead to a number of errors, such as ignoring valid AI recommendations, underutilizing available automation capabilities, or overlooking beneficial advice. Such pitfalls can lead to suboptimal or even dangerous scenarios in domains where accuracy and safety are critical, such as aviation, public transportation, and power grid operations.

Negative trust may lead to the creation of workarounds. A workaround is essentially a human method of overcoming a perceived problem with or a perceived limitation of a system. This happens because humans perceive AI as not helpful in achieving objectives (Koopman and Hoffman, 2003). Consequently, humans do not cooperate with the AI but try to find and execute alternative ways to achieve their objectives.

Overtrust in AI is another form of inappropriate trust. Overtrust is a scenario in which trust in AI exceeds the level justified by its capabilities and reliability (Jacovi et al., 2021). This overtrust, especially when skepticism is more appropriate, can lead to significant errors. Parasuraman and Manzey (2010) describe several challenges related to overtrust. They manifest in different ways, such as commission errors, where humans accept incorrect information from the system as accurate, or omission errors, where the AI fails to give a critical warning and the human is unaware of this omission. In addition, this overtrust can culminate in automation complacency, where humans, perpetuated by overtrust in the AI, fail to adequately monitor or verify the AI's output, potentially leading to critical missings or misjudgments (Parasuraman and Manzey, 2010). The implications of such overtrust are profound and underscore the need for a balanced approach to trusting AI, where trust in the AI is balanced with an awareness of its limitations and a commitment to intervene when necessary.

An AI supporting humans in gaining appropriate trust must provide transparency regarding its capabilities and limits. Achieving this with the means of explanations is very limited, as explanations cannot be verified by humans but must be trusted blindly. Therefore, AI must facilitate an exploratory process, allowing human to explore the AI and thus enabling them to refine their understanding and adjust their trust accordingly. Such an approach ensures that trust is not blindly given but is informed by direct experience and a deep comprehension of the AI's capabilities and limitations, leading to a more nuanced and effective human-AI collaboration.

HUMAN MOTIVATION

The tendency of not using IT-tools (Fildes et al., 2009) as well as algorithm aversion is quite common (Niehaus et al., 2022; Schaap et al., 2023). Therefore, intrinsic motivation to use AI needs to be deliberately promoted. Whether or not humans are intrinsically motivated depends on work conditions. Since both AI and the way in which humans interact with AI are working conditions, the prerequisites for intrinsic motivation must be considered when designing AI and human-AI interaction. Appropriate work conditions need to foster user engagement with the information provided by AI (Eisbach et al., 2023). This is required to develop calibrated trust (Eisbach et al., 2023) and foster proactive behavior (Erinaldi et al., 2020). Both are especially in safety-relevant systems pivotal to the ability to anticipate possible future events (Brizon and Wybo, 2009; Duchek, 2020). However, motivational aspects are usually not considered in the current design of AI. In contrast, current AI-based decision-support systems often lead to increased monitoring tasks, providing users with AI-generated recommendations instead of giving them an essential role in decision-making. Such monitoring tasks overstrain humans because they are monotonous and fatiguing (Bainbridge, 1983; Kaber et al., 2009; Parker and Grote, 2022). Sometimes, they even exceed human capabilities (e.g., as AI takes more information into account than a human can oversee) (Bainbridge, 1983; Endsley, 2023b). To overcome these problems, the design of AI and human-AI collaboration must incorporate basic principles of intrinsic work motivation as described below.

Intrinsic work motivation can be defined as "the doing of an activity for its inherent satisfactions rather than for some separable consequence. When intrinsically motivated, a person is moved to act for the fun or challenge entailed rather than because of external products, pressures, or rewards" (Ryan and Deci, 2000)[p. 54]. Hence, intrinsic work motivation is a prerequisite for human interest and task engagement. However, intrinsic work motivation is not a personality trait. Rather, it is highly influenced by the task's design characteristics to be performed (Fried and Ferris, 1987; Hackman and Oldham, 1976; Parker and Grote, 2022; Rai and Maheshwari, 2021). According to the widely used task characteristics model (e.g., (Niehaus et al., 2022; Parker et al., 2017; Parker and Grote, 2022), three psychological states must be considered to support intrinsic work motivation (Hackman and Oldham, 1976): 1) Experiencing meaningfulness of the work, 2) experiencing responsibility for outcomes of the work, 3) knowledge of the actual results of the work activities.

Experienced meaningfulness of the work in the context of human-AI collaboration means that the user (not the developer/designer of the AI) must experience the interaction with the AI as meaningful (Parker and Grote, 2022; Sadeghian and Hassenzahl, 2022). In human-AI collaboration, it is suggested that task allocation should not only rely on the humans' abilities and performance. Rather, tasks allocated to the human need to be perceived as meaningful (Sadeghian and Hassenzahl, 2022). Consequently, not only the what and the how of task execution needs to be addressed, but especially the why. Therefore, all interaction elements on the AI side, such as providing information or asking for information, must have a comprehensible purpose for the human. Furthermore, humans experience meaningfulness when the interrelations between their own activities and the activities of others (including the AI's activities) are comprehensible and well-reasoned.

Experienced responsibility for outcomes of the work can be achieved through human autonomy (Hackman and Oldham, 1976; Morgeson et al., 2005) and control over the process (Wäfler et al., 1999), and over the AI (Endsley, 2023b). This is because people only feel responsible for outcomes that they have at least partially influenced themselves. Hence, autonomy is given when the outcome depends on "the individuals' own efforts, initiatives, and decisions" (Hackman and Oldham, 1976)[p. 258]. Within the collaboration with an AI, this means that the human must have the possibility to initiate the interaction and must experience decisional control (Schaap et al., 2023). The control is given when the following three preconditions are met: transparency, predictability, and influenceability (Wäfler et al., 1999). An essential component of control is the awareness and understanding of the current situation, as emphasized in the situation awareness model (Endsley, 1995, 2023c). A clear perception and adequate comprehension of the environment enable individuals to anticipate possible future events and make appropriate decisions. To have control over the process and AI, and therefore over decisions, not only supports intrinsic work motivation but also avoids algorithm aversion(Schaap et al., 2023). If the human does not have decisional control, the acceptance of the AI-made decisions and the agent is low (Schaap et al., 2023).

Knowledge of the actual results of the work activities is also a precondition for intrinsic work motivation (Hackman and Oldham, 1976). Imagine a long jumper who never gets feedback on how far the jumps are. The jumper would certainly lose intrinsic motivation to jump. Feedback on performance is crucial to maintaining motivation. This is why pedometers motivate people to move: Knowing if you've reached your 10,000 steps is motivating. To have a motivational effect, the feedback must be timely. If the long jumper only receives feedback on his average performance once a year, this has no motivating effect. In human-AI collaboration, it can be assumed that humans are more motivated to use AI when AI provides comprehensible feedback about the effectiveness of his or her performance. This feedback may refer to both the process and the AI. The former refers to the human impact and effectiveness of process control. The latter refers to the effectiveness of using AI, i.e., whether humans utilize the AI's potential (Parker and Grote, 2022).

An AI supporting human motivation must provide transparency regarding meaningfulness (e.g., by providing explanations for the why of suggested decisions), autonomy (e.g., by providing the possibility to explore different options for solutions), and feedback (e.g., by providing explanations regarding the effects of the solution chosen as compared with those options not chosen).

HUMAN IMPLICIT SYMBIOSIS

First mentioned by Licklider (1960) in 1960, “Man-Computer Symbiosis” was intelligent technology that should be developed to augment human intelligence, not replace it, turning the typical user-centered application into a distributed system. This idea at the time did not have many followers and was lost through most of the 20th century. However, with the rise of AI, computing power density, wearable devices, and IoT, it is becoming a relevant concept in a world where humans need more and more interaction with machines daily. The Macmillan dictionary states that symbiosis is “a close relationship between two different things or people from which both get benefits.”. This is already an adaptation from the early biology concept that stated symbiosis is “a close connection between two different living creatures from which both usually get benefits”. In this symbiotic relationship, neither the machine will be a tool for the human, nor the human will be just an aid and/or supervisor of the machine’s automated processes. Both sides will be able to make autonomous decisions while aware of the current state of their counterpart, and they will be able to alter their behavior according to the current environmental conditions and the final goal.

In this way, there is a need to make the interaction between humans and machines more fluid and natural. Interfaces and input commands on a keyboard are far from what human-to-human interaction usually is, where more complex phenomena such as speech, body language, and psychophysiological states happen and interfere with the process. For example, in the same way humans perceive that a machine/AI is not behaving as it should, the machine could also perceive humans beyond the system interfaces (explicit interaction). Here, we introduce a key concept: implicit interaction. Such interaction aims to take the interaction between humans and AI to the next level by providing another layer of knowledge about the human - the capability of the machine/AI to perceive the human psycho and physiological states. Joining this implicit knowledge with the explicit interaction, both parties work better together to reach a common goal for mutual advantage, always being aware of each other state and plans to create a truly symbiotic relationship.

A strong effort in this sense is being made in the industry sector (Industry 5.0), where collaborative robots (cobots) perform more complex tasks with some interaction with humans. However, it raises significant challenges related to workers’ safety since most systems are reactive and do not anticipate human movement (Buerkle et al., 2021). It can easily lead to accidents, bringing up the importance of predicting human intention for the robot to adapt and change its behavior. Such prediction is impossible in explicit ways due to the lack of movement from the user, who would easily benefit from an implicit symbiosis to enhance the human-cobot relationship. In this field, probabilistic models are already aimed at overcoming these limitations. However, predicting human behavior remains highly complex (Bi et al., 2019; Wang et al., 2021).

In the grip systems operator's area, there is still a large lack of systems that provide such symbiotic relations. Several research studies are being conducted to understand the operators themselves by evaluating the cognitive load (Wenskovitch et al., 2022), performance (Anderson et al., 2023) or even stress (Rodrigues et al., 2021). These studies cover different types of operators (e.g., air traffic controllers and power grid operators) and try to understand their behavior in different operating scenarios and more difficult situations that can be partially supported by AI systems. Normally, psychophysiological metrics are collected in these studies, such as heart rate variability, reaction time, tried social stress, scoring, surveys, and interviews, allowing researchers to design a combined Machine Learning model for a deeper understanding of human behavior. Most of these studies aim to understand the operator in a better way and provide systems that interact explicitly with the humans through the interface but do not close the loop - the human behavior is not implicitly re-used by the machine for a higher level of symbiosis.

In the AI4REALNET project, we advocate the need to take this next step and close the loop between the human and the AI, where the symbiosis is implicit, and the human implicitly interacts with the AI directly, reinforcing its learning.

5.3.3. DESIGNING FOR TRANSPARENCY

Automation transparency is expected to become increasingly important for system developers, policymakers, and operational users. To date, however, it remains unclear how to achieve transparency best *by design* in identifying what information needs to be conveyed and offering design guidance on how to organize it best (e.g., levels of transparency).

In the context of the AI4REALNET project, we plan to adopt a decision-centric view on AI transparency by focusing on the important decisions that have operational consequences that the human operator cares about. Two kinds of decisions are considered. Firstly, decisions regarding the process that is being controlled. Secondly, decisions regarding the delegation of tasks and functions to AI. Given the socio-technical nature of the work environment, Cognitive (systems) Engineering principles and two complementary design and analysis methods will be considered – Ecological Interface Design (EID) and the Joint Control Framework (JCF).

EID principles are used to develop interfaces based on identifying and portraying the functional work domain constraints governed by causal and intentional laws, rules, and principles (Borst et al., 2015). Such constraints are independent of specific actors and agents (e.g., humans or (AI-based) automation) operating in the same workspace. This phase is known as the Work Domain Analysis (WDA) and results in a functional (abstraction) hierarchy ranging from the higher-level functional purpose of the overall system to the location and status of system resources. As such, an ecological interface typically offers *domain transparency* by integrating contextual information but lets agents (either human or automated) 'finish the design' by empowering them to decide upon a specific course of action (Borst et al., 2015; Paassen et al., 2018; Klomp et al., 2015).

The EID approach opens up the potential for mutual learning. The AI could learn from the human, by analyzing selected (manual) actions (decisions on what to do / how to do it); and then starting to propose or even execute those. The decisions and actions that the AI performs can also be portrayed on an ecological interface, which shows what constraints have been considered. Imagine that the operator notices that traffic is not flowing well in a sector with strong winds – then the operator can observe the system through the ecological interface and check whether the wind is being considered (correctly).

To analyze how the AI and the human operator work together in terms of activity patterns (e.g., sensing, exchanging information, acting, etc.) the Score notation in JCF can be used. The JCF includes the same functional abstraction levels as EID traditionally relies on but adds an overarching level that describes the current 'Frame' that sets the stage for the technical system and its associated goals. Framing is needed to account for human subjectivity – the operator frame can differ from the frame that an AI has and also from that of a system designer (who specifies a functional purpose – however, their framing before operations may differ from that of operators, who may, for instance, encounter situations that were unforeseen, or use systems for unplanned ends). The JCF models control processes over time, describing perceptions, decisions, and actions between a subject and their object of concern (Lundberg et al., 2021), as well as the autonomy of humans and AI respectively (Lundberg and Johansson, 2021). Thus, the interactions are mapped on the same levels as are used in EID to, for instance, model control of drone traffic (Westberg et al., 2022).

Using the score notation, the AI process toward the process of concern (e.g., air traffic) can be described. Secondly, the Operator process in monitoring the AI (or automation) can also be described. This includes work-as-done (Lundberg et al., 2024) and designs for how work should be done (Nylin et al., 2022). Further, intent can also be described, including options for AI decisions and what the AI assumes humans intend. This can be useful for pro-active monitoring of an AI (Hammarbäck et al., 2023, 2024).

5.3.4. AUTOMATION TRANSPARENCY RELATED RESEARCH PROPOSAL

Five key conclusions can be drawn from the above regarding automation transparency and the associated research proposals. These are described in the form of theses.

Thesis 1: The introduction of AI into decision-making processes requires a careful allocation of func-

tions between humans and AI to avoid negative impacts on human performance and to enable synergies between humans and AI that go beyond the capabilities of humans and AI alone.

Research is required to understand better the complementarity of humans and AI, i.e., the qualitatively different strengths and weaknesses of both in decision-making, and to understand better how humans and AI are to be combined to create synergies.

Thesis 2: To empower humans to be real partners for AI in human-AI decision-making goes far beyond the explainability of an AI's inner workings and of AI-generated decisions. Rather, automation transparency must support human cognitive processes related to decision-making, learning, trust, and motivation.

Research is required regarding the key contents humans need to have transparency about when collaborating with AI so that critical human cognitive processes are best supported. Even if the aim is to design human-AI collaborative decision-making, this does not only refer to cognitive processes of human decision-making. It also refers to cognitive processes of human learning as the human side of co-learning, to cognitive processes of human trusting as the human side of an AI's trustworthiness, and to cognitive processes of human motivation.

Thesis 3: While explainability is important, it is only one form for providing automation transparency. Others are exploration, animation, mirroring, or intuitive interface design.

Research is required to enable contextualized forms where AI can provide transparency on topics relevant to human-AI collaborative decision-making. These forms may be an explanation (i.e., AI explains a subject matter), exploration (i.e., AI supports the human to explore a subject matter, e.g., by creating a protected environment in which decisions can be tried out and corresponding effects are tested, or in which capabilities and limits of an AI can be explored), animation (i.e., AI animates the human to reflect on a subject matter, e.g., by triggering the human to formulate hypotheses the explain observed phenomena), mirroring (i.e., AI mirrors individualized patterns in human behavior to make the human aware of own biases and variabilities in decision-making), or intuitive interface design (i.e., the subject matter is intuitively comprehensible for humans).

Thesis 4: AI-based decision-support must go beyond providing recommendations (with or without explanations or transparent decision models). Cognitive forcing and evaluative AI are other forms of decision support more suitable for supporting human cognitive processes related to decision-making, learning, trust, and motivation.

Miller (2023) outlines various forms of collaborative human-AI decision-making ranging from human-on-the-loop to human-in-the-loop and machine-in-the-loop. These forms are AI providing recommendations (with or without explanations or transparent decision models), cognitive forcing (i.e., the human takes an initial decision, AI provides explanations and recommendations regarding this human-initialized decision), and evaluative AI (i.e., the human formulates hypothesis, AI provides human with evidence for and against this hypothesis). Cognitive forcing and evaluative AI are suitable to support exploration, animation, and mirroring, as described in thesis 3. However, research is still needed to develop forms of AI that enable cognitive forcing and evaluative AI.

Thesis 5: The analysis and design of AI, human-AI interaction, and human-AI collaboration require methods of cognitive (system) engineering that can model the decision-making process as well as requirements for function allocation resulting from human cognitive processes related to decision-making, learning, trusting and motivation.

Ecological Interface Design (EID) and the Joint Control Framework (JCF) are well-suited methods for analyzing and designing AI, human-AI interaction, and human-AI collaboration. Research must integrate key aspects of human cognitive processes related to decision-making, learning, trust, and motivation.

As described in Chapter 3, decision-making in critical network control is challenging and becoming increasingly demanding. Therefore, the need for research resulting from these 5 theses for AI4REALNET relates to the question of how AI can effectively support relevant cognitive processes of the human decision-maker. These cognitive processes are described in Figure 3. In the focus are necessary, preventive or corrective decisions to control the network. For the human decision-maker to make appropriate decisions effectively and efficiently, AI must support the development of appropriate situational awareness, e.g., by alerting humans to emerging problems in the network. To gain adequate situation awareness, humans need to monitor the network. This also needs to be supported by AI, e.g., by supporting humans in learning where the critical points in the network are so that humans can manage their attention accordingly. Monitoring, developing appropriate situational awareness, and decision-making require knowledge represented by mental models. In critical network control, these central mental models refer to representations of the environment (i.e., knowledge about the network to be controlled), representations of human-human collaboration (i.e., knowing what impact one's own decisions have on other human decision-makers who aim at keeping other areas of the network under control), representations of AI (i.e. knowing the capabilities and the limitations of the AI to build adequate trust into the AI; knowing how best to interact with the AI), and representations of oneself (i.e., knowing the own patterns and biases in decision-making). Furthermore, these mental models need to be developed and continuously improved. This is a continuous human learning process that also needs to be supported by AI. Lastly, AI needs to support human motivation to engage in network control by using the AI. This includes, for example, making the meaningfulness of AI behavior comprehensible or giving humans feedback on the effects of their decisions.

Supporting cognitive processes as described above by AI means to provide the human with corresponding transparency. This can be implemented by providing explications. Research is needed to clarify which content and forms of explications are suitable, for example, to make meaningfulness comprehensible. However, explications are not the best way to ensure transparency for many of the

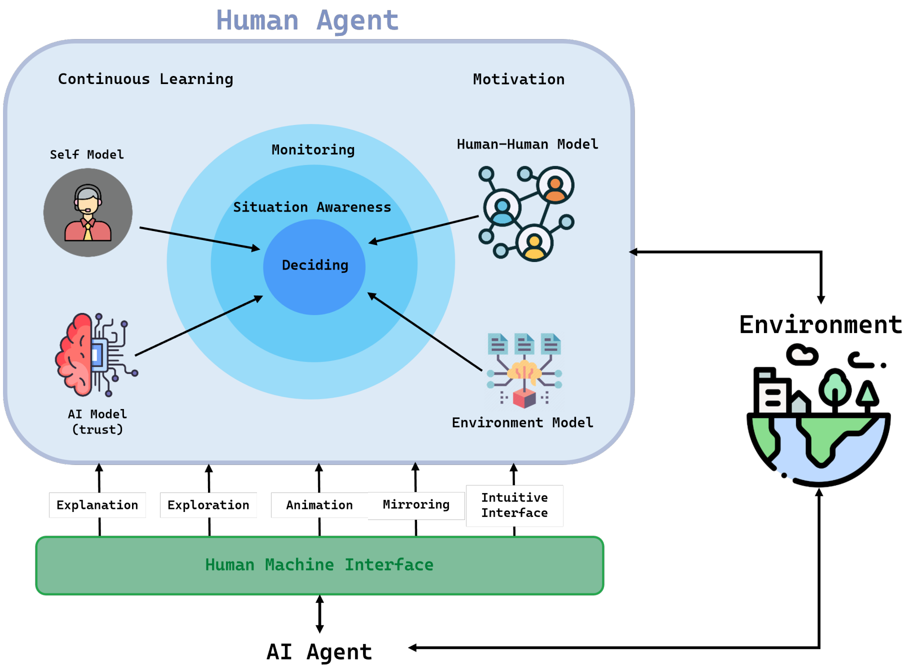

FIGURE 3 - MODEL OF HUMAN DECISION-MAKING

mentioned cognitive processes. While explication might be suitable for real-time decision-making support, exploration might better support learning processes. In this way, AI-supported exploration enables humans to develop more adequate mental models of, for example, leverage points in the network to be controlled. Evaluative AI (Miller, 2023) can also support exploration, i.e., the human formulates hypotheses about causal relationships in the network, and the AI searches the data for arguments in favor and against. Exploring the AI is also a prerequisite for developing mental models of the AI. It makes it possible to familiarize oneself with the capabilities and limitations of AI, which in turn enables the development of appropriate trust in AI. For example, by recognizing and mirroring patterns in human decision-making behavior (e.g., the tendency to make riskier decisions at the end of a shift), AI can support people in developing an appropriate self-model.

A method is required for the targeted design of human-AI interaction to support decision-making in operating and controlling critical networks and the associated cognitive processes described above. A promising way to achieve this is a combination of Ecological Interface Design (EID) and the Joint Control Framework (JCF). Research is required to merge these two approaches.

5.4. KNOWLEDGE-ASSISTED AI

Knowledge-assisted AI describes a somewhat diffuse subfield. Within the AI4REALNET project, we are particularly interested in *hybrid approaches that combine learning elements with existing conventional planners or human domain knowledge*. Such methods occur in various bodies of work but are often referred to using different terminology. Except for the term “knowledge-assisted”, closely related methods are also referred to as “neuro-symbolic” methods, “hybrid” methods, or “informed” machine learning, for example. Here, we give an overview of modern approaches to knowledge-assisted AI regardless of terminology.

We will focus on “generic” approaches that specify approaches that can work in different application domains using some formalization of the additional sources of information. We will thus exclude work specific to only one application scenario. Furthermore, we will pay special attention to knowledge-assisted approaches for RL since it is such a central technology in the AI4REALNET project. We will also focus on formalized knowledge instead of direct interaction with human experts - this mode will be covered in WP3 instead.

To start with, Von Rueden et al. (2021) provides a survey of what they term “informed” machine learning. They specifically define informed machine learning as *learning from a hybrid information source that consists of data and prior knowledge*, and furthermore emphasize such knowledge *comes from an independent source, is given by formal representations, and is explicitly integrated into the machine learning pipeline*. They survey over 150 papers in the field and present a taxonomy that considers these key elements: the knowledge *source* (scientific knowledge, world knowledge, or expert knowledge), the *representation* (algebraic or differential equations, simulation results, spatial invariances, logic rules, knowledge graphs, probabilistic relations, or human feedback) and the *integration* into the machine learning pipeline (to augment training data, constrain they hypothesis set, modify learning algorithms, or filter final hypotheses). Where it considers reinforcement learning algorithms, they mainly identify human feedback as representation, such as feedback that replaces rewards (Knox and Stone, 2009), preferences with respect to action sequences (Christiano et al., 2017), and natural language instructions (Kaplan et al., 2017).

A related overview, focusing on neural-symbolic methods, is provided by Van Harmelen and Ten Teije (2019). This work attempts to classify the abstract design patterns behind neural-symbolic methods. Design patterns of interest to AI4REALNET include patterns where symbolic methods further process the output of learning elements, such as the patterns termed ‘explainable learning systems’ and ’learning an abstraction for reasoning’, and patterns where learning elements take symbolic prior information as additional input, termed ‘learning with (derived) symbolic information as prior’. This pattern overlaps with the definition of ‘informed machine learning’ mentioned earlier.

Individual neuro-symbolic methods are surveyed by Yu et al. (2023); Hitzler et al. (2022); Garcez et al. (2022); Sarker et al. (2021). Out of these, the survey by Yu et al. (2023) is perhaps most relevant as the most recent survey and as the only survey that covers reinforcement learning in some more detail. It proposes the categories “learning for reasoning”, “reasoning for learning”, and bi-directional “learning-reasoning”. Out of these, the last two are relevant for knowledge-assisted AI. The reinforcement learning methods in these categories will be covered later in this section. What is interesting about the survey by Sarker et al. (2021) is that they identify a potential class of methods referred to as [Neuro[Symbolic]] that have a deep integration of symbolic methods as deliberative components inside a deep neural network; however, none of the papers they surveyed belongs to.

Neither of the surveys presented so far covers many reinforcement learning approaches. In the literature, we find several groups of approaches. The first group uses prior information to define an initial policy, whereas the second group uses prior information to inform the reward and/or value functions. A third group uses function approximators that contain a mixture of neural and symbolic components. Lastly, a fourth group takes a hierarchical approach with a more symbolic type of planning at the high level and (neuronal) learning at the lower level.

In the first group, we have, for example, knowledge-assisted deep deterministic policy gradients (Zhao et al., 2020), where exploration actions are sometimes selected from a prior policy (obtained by optimizing a low-fidelity model). Knowledge-assisted deep Q learning (Zhao et al., 2022) is a closely related approach, which has also been applied to a power network control problem and which additionally considers an approximation to obtain the guiding policy even for very large problems. Shielding (Alshiekh et al., 2018) could be seen as a special case of this category where a 'shield' based on a temporal logic description is used to prohibit taking dangerous actions. In this category, we could also include methods that use knowledge about invariances of the environment (Van der Pol et al., 2020).

A second group of methods uses knowledge to aid the value function. For example, 'value refinement networks' (Wöhlke et al., 2022) use planning in a coarse model to obtain an initial value function, which is then refined using a learned convolutional neural network to provide more fine-grained Q-values. Several methods also combine a knowledge-based guiding policy with a knowledge-based initial reward- or value function. Dai et al. (2022) uses an additional reward based on the distance of the chosen action to the action from the guiding policy. Xie et al. (2024) uses heuristics to shape the reward function. Gao et al. (2022) use physics knowledge represented as PDE to constrain the policy and value function to control a power network.

In a third group, there are approaches that combine both neural and symbolic components in a single function approximator (serving, e.g., as policy or value function). For example, Garnelo et al. (2016); Garcez et al. (2018) use a pipeline consisting of a neural network (trained with unsupervised learning) to extract a symbolic representation from sensory inputs, which is then further processed by a symbolic system fine-tuned with reinforcement learning. Höpner et al. (2022) considers a value function that uses a knowledge graph to generalize learned information to sparsely visited states using relations between tokens in the same higher-level category.

A fourth class of methods contains hierarchical methods that typically use a form of symbolic reasoning or planning at the higher level, whereas the low-level execution of behavior is performed by neural modules. High-level planning typically relies on knowledge of the high-level dynamics of the problem, whereas low-level behavior is purely learned from data. The work by Lyu et al. (2019); Yang et al. (2018); Araki et al. (2021) are examples in this category. Variations are proposed by Vaezipoor et al. (2021), where the high-level description in the form of linear temporal logic is directly fed into a neural architecture, and Mitchener et al. (2022), where the weights of the symbolic system are also tuned by reinforcement learning, bringing data-driven methods also to the high-level decision making.

We can now turn back to the taxonomy by Von Rueden et al. (2021) and ask how reinforcement learning methods fit in this taxonomy. Looking at different knowledge representations, we have seen methods that exploit knowledge of (spatial) invariances (Van der Pol et al., 2020), logic rules about high-level transitions (Lyu et al., 2019; Yang et al., 2018; Araki et al., 2021; Vaezipoor et al., 2021), knowledge graphs (Höpner et al., 2022), differential equations (Gao et al., 2022) and human feedback (Knox and Stone, 2009; Christiano et al., 2017; Kaplan et al., 2017). Moreover, physics-informed machine learning (Karniadakis et al., 2021) has been emerging as an approach to integrate the principles of physics (e.g., conservation laws, symmetry, and boundary conditions) with machine learning algorithms to inform and guide the learning process and enhance their predictive capabilities and interpretability, as well as to solve partial differential equations based on physical simulation (e.g., time domain simulations of the power system dynamic response to disturbances).

Probabilistic relations and simulation results were not covered here – probabilistic environment models do play a central role in model-based reinforcement learning, where they are often used as simulators, but this is a huge field in itself which is better covered elsewhere, such as in the recent survey by Moerland et al. (Moerland et al., 2023). Model-based reinforcement learning also still leaves the problem of extracting the actual policy or value function open. Similarly, algebraic and differential equations also have obvious applications in model-based approaches.

Future research directions Within the context of the AI4REALNET project, there are a number of specific challenges that can be investigated further. Many scenarios in the AI4REALNET domains rely on a network topology structure and critical safety constraints. Using these pieces of information to guide (reinforcement) learning efficiently to near-optimal solutions holds promise across such scenarios.

More generally, future approaches could focus on less studied representations, such as the use of differential or algebraic equations directly in policies or value functions. Furthermore, less-explored design patterns, such as the use of symbolic methods as deliberative components inside a neural network, seem to be an underexplored niche. There also seems to be space for a more modern approach that directly integrates constraints into the architecture of the neural network, analogous to Towell and Shavlik (1994) but adapted to deep architectures.

5.5. CO-LEARNING

The question of how automated systems will interact with humans has been asked for many years, going back to Woods in 1996, who argued that automated systems can be integrated into human teams when those systems are perceived as individual and independent agents (Woods, 1996). The following year, Rich and Sidner showed that highly automated systems can participate in human teamwork, but must adhere to the principles of human-human cooperation (Rich and Sidner, 1997). With the recent advances in AI capabilities, research into the design of human-AI teams has gained momentum and is known under many names, with a plethora of concepts, ideas, and initial implementations. Aliases are commonly “Human-AI” combined with a suffix indicating a collaborative nature – “Teaming” and “Cooperation” are among the most common. In the following and within the context of AI4REALNET, we predominantly use the term “co-learning”, interchangeable with other aliases. The co-learning concept described in the following section differentiates itself from existing research in that it aims to achieve continual and mutual learning in a human-AI team, viewing the system holistically with the goal of exploiting strengths while mitigating weaknesses, thereby achieving performances superior to that which the agents could achieve individually.

Co-learning has been defined as an emergent process in which group members engage in bidirectional communication, exchanging feedback and adapting their behavior over time (Abich and Sikorski, 2023). Co-learning necessitates a mutual and explicit learning objective with the over-arching goal of improving team performance (Schoonderwoerd et al., 2022). In almost all definitions, a condition imposed on any co-learning system is that the AI entity be autonomous and capable of perceiving and acting on its surroundings in ways previously limited to human agents (Wynne and Lyons, 2018). Wynne & Lyons additionally state that the agent must have a unique role to play in the team and not serve merely as a tool. It can be said that co-learning aims to recreate the fruitful cooperation humans intuitively cultivate in human-human teams.

Research Directions Most existing literature is limited to theoretical works or initial analyses of human interaction with artificial agents in collaborative settings. The theoretical works discuss the challenges of co-learning and propose solution strategies, providing a valuable conceptual foundation on which further work can be done. The many analyses of human interaction are predominantly limited to non-learning agents or in some cases pseudo-AI-agents in “Wizard-of-Oz”-type studies, where the ”AI-agent” is in fact controlled by a human. The results of such studies shed light on important psychological phenomena, such as the impact of high communication loads on a human’s ability to integrate information from teammates into their decisions (Westby and Riedl, 2023), or that the level of AI assistance tolerated by a human is very individual and linked to their Need for Cognition (NFC)

(Swaroop et al., 2023). To the best of our knowledge, there are no works that develop *and* implement a conceptual framework for co-learning that considers the needs of the human agents and derives requirements for the artificial agent.

In their 2019 paper titled "Six Challenges for Human-AI Co-Learning", van den Bosch et al. provide a detailed description of co-learning as well as proposing a series of requirements and challenges. The authors propose six models that an agent must have and continually refine to achieve mutual learning in a human-AI team. More specifically, an agent requires taxonomy, team, task, self, "Theory-of-Mind," and communication models to be capable of interacting with human agents in a manner that conforms with human cognition (Van den Bosch et al., 2019).

For any human team to function, a common language and a shared understanding of team dynamics is required. In the system proposed by Van den Bosch et al., the interaction between human and artificial agents is managed by the human agent via the *team model*, which defines work agreements, team organization, hierarchy, task distribution, and delegation. The *taxonomy model* manages the shared language, pertaining to concepts and relations important for a common understanding of the task. With a common taxonomy and the agent's place in the team defined, it can begin to solve tasks. To do so, a *task model* is required, which is comprised of knowledge about the task and the relations between states, actions, and outcomes, including solution strategies and representation of state knowledge. Two models exist that describe the inner states of the team members: the *self-model*, depicting the inner state of the artificial agent, and the *"Theory of Mind"-model*, which covers knowledge about the inner state of other agents. Both models contain information about the goals, values, capabilities, resources, and intentions of the agents. The Theory-of-Mind-model differentiates itself from the self-model in that the information can be provided directly by the human agent or inferred by the artificial agent through behavioral observation. It also considers aspects of emotion and personality. The knowledge of self and of others enables productive alignment and adaptation within the team, which occurs through the final model – the *communication model*, which is informed by the team and taxonomy model and exchanges information with the self- and Theory-of-Mind models. The sharing of information enables the AI agent to be able to process human communication and send information using the defined vocabulary, under consideration of the human's inner state, within the context of agreed-upon team dynamics while communicating its approach to the task as well as its own inner state. Communication of the inner state resulting from the self-model is of particular significance, given that it cannot be inferred from behavioral cues it would be in a human-human team (Van den Bosch et al., 2019).

A descriptive design for a co-learning system based on this concept of interconnected models within the artificial agent is presented in Figure 4, where arrows display interaction and information flow. Humans interact with artificial agents via a human-machine interface (HMI) and with the environ-

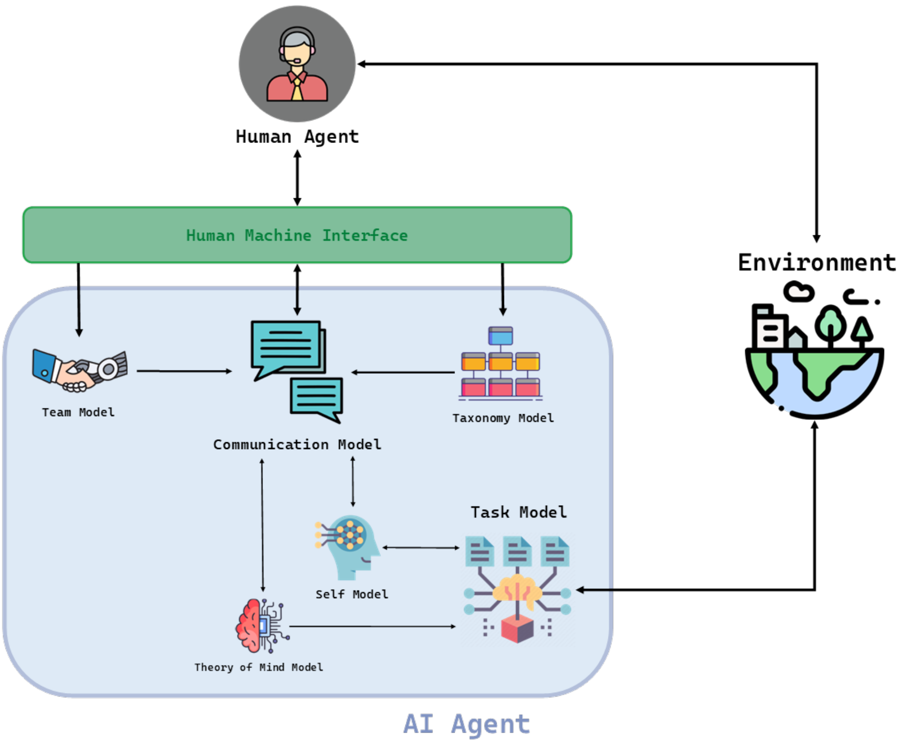

FIGURE 4 - PRELIMINARY DESIGN FOR HUMAN-AI CO-LEARNING BASED ON Van den Bosch et al. (2019).

ment directly. Via the HMI, the human agent can control the team and taxonomy models, setting the framework for collaboration and communication. The team and taxonomy models inform the communication model, ensuring efficient communication. The HMI also serves as the communication interface between artificial and human agents. Via communication with the human agent and through observation, the theory-of-mind model is maintained. The self-model informs the communication model, allowing the artificial agent to communicate its inner state. Both theory-of-mind- and self-model inform the task model, which interacts directly with the environment.

While this co-learning concept is merely descriptive, by no means definitive, and has no concrete concepts for technological implementation, it nonetheless provides an overview of the functionalities required for human-AI co-learning and provides a suggestion for their interactions. The lack of technological discussion is intentional - the design of system architecture and implementation of its individual parts, as well as impact analyses, are precisely the research gaps that this project can tackle.

AI4REALNET has the unique opportunity and expertise to develop a holistic approach to co-learning, considering both the technological and human side of co-learning systems while being among the first to realize a real-world co-learning system, a significant and important contribution to the advancement of the field.

6. CONCLUSIONS

In conclusion, the AI4REALNET consortium advocates that the focus should be placed on optimizing the degree of decision support of AI to humans, aiming at achieving the best possible team between humans and AI technology (rather than simply deploying AI-based systems). To accomplish this, the objective is not simply the implementation of automation or AI but rather the enhancement of the socio-technical system's overall efficiency, ensuring maximum human performance and engagement consistently.

In this line, the explainability of AI is crucial for developing an accurate mental model, as it clarifies the AI's decision-making process. However, it alone does not ensure effective human learning. Therefore, AI4REALNET emphasizes that AI transparency is a vital complement, offering clear, real-time insights into AI's activities, which is essential for immediate understanding and interaction in dynamic contexts. Transparency should be integrated into the four stages of the learning cycle: 1) during concrete experience, by explaining various factors of the process to encourage exploration; 2) during reflective observation, by prompting reflection and hypothesis formulation about interrelated factors; 3) during abstract conceptualization, by providing data-based evidence for or against the human's hypotheses; and 4) during active experimentation, by enabling safe real-world exploration and immediate feedback on outcomes.

To demonstrate the tangible value of AI-based decision systems in the industry, AI4REALNET focused on six use cases that address key industry requirements across three infrastructures with common properties. Operating such a complex system means dealing with unexpected events (e.g., weather-related events), which requires the definition of remedial and preventive actions in real-time. Taking the full complexities of these networks is, in general, an NP-hard problem that should be solved in real-time in a dynamic environment, which is even more challenging, and novel approaches are required. The AI4REALNET advocates the formulation of these use cases as sequential decision problems (Markov decision processes) but also explores the supervised learning paradigm. Aspects such as scalability in large-scale networks (i.e., mitigate the curse of dimensionality and the curse of the horizon) and sampling efficiency in RL are addressed by exploring distributed and hierarchical RL, as well as knowledge-assisted AI (that is aligned with neurosymbolic learning). Furthermore, inverse RL can enable the development of human-inspired policies that may facilitate acceptance and trustworthiness.

The following expected benefits are foreseen for society:

- **Network capacity**. AI-based human operator support tools to increase capacity in airspace, railway, and electrical networks and postpone investments in additional network capacity (e.g., integrate additional renewable energy sources) or ensure the integration of new entrant aircraft types (e.g., hybrid hydrogen-electric aircraft).

- **Operational efficiency**. Improvements in operational efficiency are achieved by enabling better predictions, contributing to the punctuality of air traffic and railway operations, and minimizing operational costs by activating cheaper electrical network flexibility.
- **Safety and security**. Maintain at least the same level of safety and security as the current network management system. Increase resilience to extreme (natural and man-made) events.
- **Environment**. Achieve a positive impact of AI-based solutions on operational mitigation of aviation's environmental impact, for example, by directing traffic over shorter and more energy efficient (high altitude) routes; facilitate energy transition by reducing renewable energy sources curtailment and improving carbon intensity of actions.

REFERENCES

Karel Van den Bosch, Tjeerd Schoonderwoerd, Romy Blankendaal, and Mark Neerincx. Six challenges for human-ai co-learning. In *Adaptive Instructional Systems: International Conference (AIS)*, pages 572–589. Springer, 2019.

Michael Chui, James Manyika, Mehdi Miremadi, Nicolaus Henke, Rita Chung, Pieter Nel, and Sankalp Malhotra. Notes from the ai frontier: Insights from hundreds of use cases. *McKinsey Global Institute*, 2018.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

Christopher Bishop. *Pattern recognition and machine learning*. Springer, 2006.

Martin L Puterman. Markov decision processes. *Handbooks in operations research and management science*, 2:331–434, 1990.

K.J. Åström. Optimal control of markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10(1):174–205, 1965.

Richard D Smallwood and Edward J Sondik. The optimal control of partially observable markov processes over a finite horizon. *Operations research*, 21(5):1071–1088, 1973.

George E Monahan. A survey of partially observable markov decision processes: theory, models, and algorithms. *Management Science*, 28(1):1–16, 1982.

Michael L Littman. A tutorial on partially observable markov decision processes. *Journal of Mathematical Psychology*, 53(3):119–125, 2009.

Craig Boutilier, Richard Dearden, and Moisés Goldszmidt. Exploiting structure in policy construction. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1104–1113. Morgan Kaufmann, 1995.

Craig Boutilier, Thomas L. Dean, and Steve Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94, 1999.

Thomas Degris and Olivier Sigaud. Factored markov decision processes. *Markov Decision Processes in Artificial Intelligence*, pages 99–126, 2013.

Ian Osband and Benjamin Van Roy. Near-optimal reinforcement learning in factored mdps. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 604–612, 2014.

Ronald A Howard. *Semi-Markovian control systems*. MIT Operations Research Center, 1963.

Sheldon M Ross. Average cost semi-markov decision processes. *Journal of Applied Probability*, 7(3): 649–656, 1970.

Gianluca Drappo, Alberto Maria Metelli, and Marcello Restelli. An option-dependent analysis of regret minimization algorithms in finite-horizon semi-mdp. *Transactions on Machine Learning Research*, 2023.

Lloyd S Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953a.

Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 157–163. Morgan Kaufmann, 1994.

Jerzy Filar and Koos Vrieze. *Competitive Markov decision processes*. Springer, 1997.

John F Nash. Non-cooperative games. *Annals of Mathematics*, 1950.

Eric A. Hansen, Daniel S. Bernstein, and Shlomo Zilberstein. Dynamic programming for partially observable stochastic games. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 709–715. AAAI Press / The MIT Press, 2004.

Qinghua Liu, Csaba Szepesvári, and Chi Jin. Sample-efficient reinforcement learning of partially observable markov games. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:18296–18308, 2022.

Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. The complexity of decentralized control of markov decision processes. *Mathematics of operations research*, 27(4):819–840, 2002.

Christopher Amato, Girish Chowdhary, Alborz Geramifard, N Kemal Üre, and Mykel J Kochenderfer. Decentralized control of partially observable markov decision processes. In *IEEE Conference on Decision and Control*, pages 2398–2405. IEEE, 2013.

Frans A Oliehoek and Christopher Amato. *A concise introduction to decentralized POMDPs*. Springer, 2016.

V. Van Roy, F. Rossetti, K. Perset, and L. Galindo-Romero. Ai watch – national strategies on artificial intelligence: A European perspective, 2021 edition. Technical report, Publications Office of the European Union, 2021.

Maria Luisa Di Silvestre, Salvatore Favuzza, Eleonora Riva Sanseverino, and Gaetano Zizzo. How decarbonization, digitalization and decentralization are changing key power infrastructures. *Renewable and Sustainable Energy Reviews*, 93:483–498, 2018.

Mathaios Panteli, Dimitris N. Trakas, Pierluigi Mancarella, and Nikos D. Hatziargyriou. Power systems resilience assessment: Hardening and smart operational enhancement strategies. *Proceedings of the IEEE*, 105(7):1202–1213, 2017.

J. R. Andrade, C. Rocha, R. Silva, J. P. Viana, R. J. Bessa, C. Gouveia, B. Almeida, R. J. Santos, M. Louro, P. M. Santos, and A. F. Ribeiro. Data-driven anomaly detection and event log profiling of SCADA alarms. *IEEE Access*, pages 73758–73773, 2022.

Antoine Marot, Adrian Kelly, Matija Naglic, Vincent Barbesant, Jochen Cremer, Alexandru Stefanov, and Jan Viebahn. Perspectives on future power system control centers for energy transition. *Journal of Modern Power Systems and Clean Energy*, 10(2):328–344, 2022a.

Antoine Marot, Alexandre Rozier, Matthieu Dussartre, Laure Crochepierre, and Benjamin Donnot. Towards an ai assistant for human grid operators. In *Hybrid Human Artificial Intelligence (HHAI)*, 2022b.

Jan Viebahn, Matija Naglic, Antoine Marot, Benjamin Donnot, and Simon H Tindemans. Potential and challenges of ai-powered decision support for short-term system operations. In *CIGRE Session*, volume 2022, 2022.

Anton R Fuxjäger, Kristian Kozak, Matthias Dorfer, Patrick M Blies, and Marcel Wasserer. Reinforcement learning based power grid day-ahead planning and ai-assisted control. *arXiv preprint arXiv:2302.07654*, 2023.

F. Heymann, H. Quest, T. Lopez Garcia, C. Ballif, and M. Galus. Reviewing 40 years of artificial intelligence applied to power systems – a taxonomic perspective. *Energy and AI*, 15:100322, 2024.

Jan Viebahn, Sjoerd Kop, Hariadi Budaya Joost van Dijk, Marja Streefland, Paul Champion Davide Barbieri, Mario Jothy, Vincent Renault, and Simon Tindemans. Gridoptions tool: Real-world day-ahead congestion management using topological remedial actions. In *CIGRE Session*, 2024.

Z. Z. Zhang, G. S. Hope, and O. P. Malik. Expert systems in electric power systems – a bibliographical survey. *IEEE Transactions on Power Systems*, 4(4):1355–1362, 1989.

Z. A. Vale and A. Machado e Moura. An expert system with temporal reasoning for alarm processing in power system control centers. *IEEE Transactions on Power Systems*, 8(3):1307–1314, 1993.

K. Demaree, T. A. Athay, K. W. Cheung, Y. Mansour, E. Vaahedi, A. Y. Chang, B. R. Corns, and B. W. Garrett. An on-line dynamic security analysis system implementation. *IEEE Transactions on Power Systems*, 9(4):1716–1722, 1994.

J. A. Huang, A. Valette, M. Beaudoin, K. Morison, A. Moshref, M. Provencher, and J. Sun. An intelligent system for advanced dynamic security assessment. In *International Conference on Power System Technology*, 2022.

Mladen Kezunovic, Pierre Pinson, Zoran Obradovic, Santiago Grijalva, Tao Hong, and Ricardo Bessa. Big data analytics for future electricity grids. *Electric Power Systems Research*, 189:106788, 2020.

Antoine Marot, Benjamin Donnot, Gabriel Dulac-Arnold, Adrian Kelly, Aidan O'Sullivan, Jan Viebahn, Mariette Awad, Isabelle Guyon, Patrick Panciatici, and Camilo Romero. Learning to run a power network challenge: a retrospective analysis. In *Proceedings of Machine Learning Research*, volume 133, pages 112–132, 2021.

Florin Capitanescu. Critical review of recent advances and further developments needed in AC optimal power flow. *Electric Power Systems Research*, 136:57–68, 2016.

Ehsan Dehnavi, Saeed Afsharnia, Amir Abbas Shayegani Akmal, and Moein Moeini-Aghtaie. A novel congestion management method through power system partitioning. *Electric Power Systems Research*, 213:108672, 2022.

Antoine Marot, Benjamin Donnot, Camilo Romero, Balthazar Donon, Marvin Lerousseau, Luca Veyrin-Forrer, and Isabelle Guyon. Learning to run a power network challenge for training topology controllers. *Electric Power Systems Research*, 189:106635, 2020.

Medha Subramanian, Jan Viebahn, Simon H. Tindemans, Benjamin Donnot, and Antoine Marot. Exploring grid topology reconfiguration using a simple deep reinforcement learning approach. In *IEEE Madrid PowerTech*, 2021.

Anandsingh Chauhan, Mayank Baranwal, and Ansuma Basumatary. PowRL: A reinforcement learning framework for robust management of power networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 14757–14764, 2023.

Malte Lehna, Jan Viebahn, Antoine Marot, Sven Tomforde, and Christoph Scholz. Managing power grids through topology actions: A comparative study between advanced rule-based and reinforcement learning agents. *Energy and AI*, 14:100276, 2023.

Bo Zhou, Hongsheng Zeng, Yuecheng Liu, Kejiao Li, Fan Wang, and Hao Tian. Action set based policy optimization for safe power grid management. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track: European Conference*, pages 168–181. Springer, 2021.

Matthias Dorfer, Anton R Fuxjäger, Kristian Kozak, Patrick M Blies, and Marcel Wasserer. Power grid congestion management via topology optimization with alphazero. *arXiv preprint arXiv:2211.05612*, 2022.

Qiuhua Huang, Renke Huang, Tianzhixi Yin, Sohom Datta, Xueqing Sun, and Jason Hou et al. Toward intelligent emergency control for large-scale power systems: Convergence of learning, physics, computing and control. *arXiv:2310.05021*, 2023.

Nastaran Gholizadeh, Nazli Kazemi, and Petr Musilek. A comparative study of reinforcement learning algorithms for distribution network reconfiguration with deep Q-learning-based action sampling. *IEEE Access*, 11:13714–13723, 2023.

Volodymyr Mnih, Koray Kavukcuoglu, and David Silver et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. Deterministic policy gradient algorithms. In *International conference on machine learning*, pages 387–395. Pmlr, 2014.

Vijay Konda and John Tsitsiklis. Actor-critic algorithms. *Advances in neural information processing systems*, 12, 1999.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. PMLR, 2018.

F. Heymann, K. Parginos, R. J. Bessa, and M. Galus. Operating ai systems in the electricity sector under European's AI Act – insights on compliance costs, profitability frontiers and extraterritorial effects. *Energy Reports*, 10:4538–4555, 2023.

Mica R. Endsley. Ironies of artificial intelligence. *Ergonomics*, 66:1656–1668, 2023a.

Antoine Marot, Benjamin Donnot, Karim Chaouache, Adrian Kelly, Qiuhua Huang, Ramij-Raja Hossain, and Jochen L. Cremer. Learning to run a power network with trust. *Electric Power Systems Research*, 212:108487, 2022c.

Ricardo Bessa, Mouadh Yagoubi, Milad Leyli-abadi, Maroua Meddeb, Toni Wäfler, and Bruno Lemetayer et al. AI4REALNET framework and use cases. Technical Report D1.1, AI4REALNET Deliverable, 2024.

SBB. Facts and figures - transportation, 2024.

CH-FOT. Expansion step 2035, 2024.

Lawrence Bodin and Bruce Golden. Classification in vehicle routing and scheduling. *Networks*, 11(2): 97–108, 1981.

Brian A Foster and David M Ryan. An integer programming approach to the vehicle scheduling problem. *Journal of the Operational Research Society*, 27(2):367–384, 1976.

Jean-Yves Potvin and Jean-Marc Rousseau. A parallel route building algorithm for the vehicle routing and scheduling problem with time windows. *European Journal of Operational Research*, 66(3):331–340, 1993.

Jing-Quan Li, Pitu B Mirchandani, and Denis Borenstein. The vehicle rescheduling problem: Model and algorithms. *Networks: An International Journal*, 50(3):211–229, 2007.

Carl Westin, Clark Borst, and Brian Hilburn. Strategic conformance: Overcoming acceptance issues of decision aiding automation? *IEEE Transactions on Human-Machine Systems*, 46(1):41–52, 2016.

R. Bloomfield. Fundamentals of european rail traffic management system - ertms. In *IET Professional Development Course on Railway Signalling and Control Systems*. Institution of Engineering and Technology, 2006.

Andrea D'Ariano. Innovative decision support system for railway traffic control. *IEEE Intelligent Transportation Systems Magazine*, 1(4):8–16, 2009.

Francesco Corman and Lingyun Meng. A review of online dynamic models and algorithms for railway traffic management. *IEEE Transactions on Intelligent Transportation Systems*, 16(3):1274–1284, 2015.

SBB. Rail control system (rcs) – verkehrssteuerung der zukunft, 2020.

Nahid Parvez Farazi, Bo Zou, Tanvir Ahamed, and Limon Barua. Deep reinforcement learning in transportation research: A review. *Transportation Research Interdisciplinary Perspectives*, 11:100425, 2021.

Rihab Gorsane, Khalil Gorsan Mestiri, Daniel Tapia Martinez, Vincent Coyette, Beyrem Makhlouf, Gereon Vienken, Minh Tri Truong, Andreas Söhlke, Johann Hartleb, Amine Kerkeni, Irene Sturm, and Michael Küpper. Reinforcement learning based train rescheduling on event graphs. In *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2023.

Oriol Vinyals, Timo Ewalds, and Sergey Bartunov et al. Starcraft ii: A new challenge for reinforcement learning, 2017.

Mikayel Samvelyan, Tabish Rashid, Christian Schroeder de Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philip H. S. Torr, Jakob Foerster, and Shimon Whiteson. The starcraft multi-agent challenge, 2019.

OpenAI, Christopher Berner, Greg Brockman, and Brooke Chan et al. Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680*, 2019.

Itamar Arel, Cong Liu, Tom Urbanik, and Airton G Kohls. Reinforcement learning-based multi-agent system for network traffic signal control. *IET Intelligent Transport Systems*, 4(2):128–135, 2010.

T. Tan, F. Bao, Y. Deng, A. Jin, Q. Dai, and J. Wang. Cooperative deep reinforcement learning for large-scale traffic grid signal control. *IEEE Transactions on Cybernetics*, 50(6):2687–2700, 2020.

Sharada Mohanty, Erik Nygren, Florian Laurent, Manuel Schneider, Christian Scheller, Nilabha Bhattacharya, Jeremy Watson, Adrian Egli, Christian Eichenberger, Christian Baumberger, Gereon Vienken, Irene Sturm, Guillaume Sartoretti, and Giacomo Spigler. Flatland-rl : Multi-agent reinforcement learning on trains, 2020.

Florian Laurent, Manuel Schneider, and Christian et al. Scheller. Flatland competition 2020: Mapf and marl for efficient train coordination on a grid world. In *Proceedings of the NeurIPS 2020 Competition* *and Demonstration Track*, volume 133 of *Proceedings of Machine Learning Research*, pages 275–301. PMLR, 2021.

István Lövétei, Bálint Kővári, Tamás Bécsi, and Szilárd Aradi. Environment representations of railway infrastructure for reinforcement learning-based traffic control. *Applied Sciences*, 12(9):4465, 2022.

Dano Roost, Ralph Meier, Stephan Huschauer, Erik Nygren, Adrian Egli, Andreas Weiler, and Thilo Stadelmann. Improving sample efficiency and multi-agent communication in rl-based train rescheduling. In *Swiss Conference on Data Science (SDS)*. IEEE, 2020.

Michel Cannarsa. Ethics guidelines for trustworthy ai. *The Cambridge handbook of lawyering in the digital age*, pages 283–97, 2021.

Markus D. Dubber, Frank Pasquale, and Sunit Das. *The Oxford Handbook of Ethics of AI*. Oxford University Press, 2020.

EU Commission HLEG. Assessment list for trustworthy ai (altai). *EU Commission, Brussels*, 2020.

Ramon Dalmau, Franck Ballerini, Herbert Naessens, Seddik Belkoura, and Sebastian Wangnick. An explainable machine learning approach to improve take-off time predictions. *Journal of Air Transport Management*, 95:102090, 2021.

Aniel Jardines, Manuel Soler, Alejandro Cervantes, Javier García-Heras, and Juan Simarro. Convection indicator for pre-tactical air traffic flow management using neural networks. *Machine Learning with Applications*, 5:100053, 2021.

R. Dalmau and E. Allard. Air Traffic Control Using Message Passing Neural Networks and Multi-Agent Reinforcement Learning. In *SESAR Innovation Days (SID)*, 2020.

Ingrid Gerdes, Annette Temme, and Michael Schultz. Dynamic airspace sectorisation for flight-centric operations. *Transportation Research Part C: Emerging Technologies*, 95:460–480, 2018.

Natalia Andrienko, Gennady Andrienko, Linara Adilova, and Stefan Wrobel. Visual analytics for human-centered machine learning. *IEEE Computer Graphics and Applications*, 42(1):123–133, 2022.

Theocharis Kravaris, Konstantinos Lentzos, Georgios Santipantakis, George A. Vouros, Gennady Andrienko, Natalia Andrienko, Ian Crook, Jose Manuel Cordero Garcia, and Enrique Iglesias Martinez. Explaining deep reinforcement learning decisions in complex multiagent settings: towards enabling automation in air traffic flow management. *Applied Intelligence*, 53(4):4063–4098, 2023.

Carl Westin, Clark Borst, Erik-Jan van Kampen, Tiago Nunes Monteiro, Supathida Boonsong, Brian Hilburn, Matteo Cocchioni, and Stefano Bonelli. Personalized and transparent ai support for atc conflict detection and resolution : an empirical study. pages 1–9, 2022.

Martijn IJtsma, Clark Borst, Marinus M. van Paassen, and Max Mulder. Evaluation of a decision-based invocation strategy for adaptive support for air traffic control. *IEEE Transactions on Human-Machine Systems*, 52:1135–1146, 2022.

S. J. van Rooijen, J. Ellerbroek, C. Borst, and E. van Kampen. Toward individual-sensitive automation for air traffic control using convolutional neural networks. *Journal of Air Transportation*, 28:105–113, 2020.

Stefano V. Albrecht, Filippos Christianos, and Lukas Schäfer. *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches*. MIT Press, 2024.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 35:24611–24624, 2022.

Andrea Castelletti, Stefano Galelli, Marcello Restelli, and Rodolfo Soncini-Sessa. Tree-based variable selection for dimensionality reduction of large-scale control systems. In *IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, pages 62–69. IEEE, 2011.

Mario Beraha, Alberto Maria Metelli, Matteo Papini, Andrea Tirinzoni, and Marcello Restelli. Feature selection via mutual information: New theoretical insights. In *International joint conference on neural networks (IJCNN)*, pages 1–9. IEEE, 2019.

Shubham Pateria, Budhitama Subagdja, Ah-hwee Tan, and Chai Quek. Hierarchical reinforcement learning: A comprehensive survey. *ACM Computing Surveys (CSUR)*, 54(5):1–35, 2021.

Gaëtan Serré, Eva Boguslawski, Benjamin Donnot, Adrien Pavão, Isabelle Guyon, and Antoine Marot. Reinforcement learning for energies of the future and carbon neutrality: a challenge design. *arXiv preprint arXiv:2207.10330*, 2022.

Blazej Manczak, Jan Viebahn, and Herke van Hoof. Hierarchical reinforcement learning for power network topology control. *arXiv preprint arXiv:2311.02129*, 2023.

Richard S Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211, 1999.

Thomas G Dietterich. Hierarchical reinforcement learning with the maxq value function decomposition. *Journal of artificial intelligence research*, 13:227–303, 2000.

Meire Fortunato, Mohammad Gheshlaghi Azar, Bilal Piot, Jacob Menick, Ian Osband, Alex Graves, Vlad Mnih, Remi Munos, Demis Hassabis, Olivier Pietquin, Charles Blundell, and Shane Legg. Noisy networks for exploration, 2019.

Cédric Colas, Olivier Sigaud, and Pierre-Yves Oudeyer. Gep-pg: Decoupling exploration and exploitation in deep reinforcement learning algorithms. In *International conference on machine learning*, pages 1039–1048. PMLR, 2018.

Adam Laud and Gerald DeJong. The influence of reward on the speed of reinforcement learning: An analysis of shaping. In *Proceedings of the International Conference on Machine Learning*, pages 440–447, 2003.

Yunlong Dong, Xiuchuan Tang, and Ye Yuan. Principled reward shaping for reinforcement learning via lyapunov stability theory. *Neurocomputing*, 393:83–90, 2020.

Andrew Y Ng, Stuart Russell, et al. Algorithms for inverse reinforcement learning. In *Icml*, volume 1, page 2, 2000.

Dizan Vasquez, Billy Okal, and Kai O Arras. Inverse reinforcement learning algorithms and features for robot navigation in crowds: an experimental comparison. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1341–1346. IEEE, 2014.

Sanjay Krishnan, Animesh Garg, Richard Liaw, Brijen Thananjeyan, Lauren Miller, Florian T Pokorny, and Ken Goldberg. Swirl: A sequential windowed inverse reinforcement learning algorithm for robot tasks with delayed rewards. *The international journal of robotics research*, 38(2-3):126–145, 2019.

Katharina Muelling, Abdeslam Boularias, Betty Mohler, Bernhard Schölkopf, and Jan Peters. Learning strategies in table tennis using inverse reinforcement learning. *Biological cybernetics*, 108:603–619, 2014.

Masamichi Shimosaka, Takuhiro Kaneko, and Kentaro Nishi. Modeling risk anticipation and defensive driving on residential roads with inverse reinforcement learning. In *International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1694–1700. IEEE, 2014.

Changxi You, Jianbo Lu, Dimitar Filev, and Panagiotis Tsiotras. Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning. *Robotics and Autonomous Systems*, 114:1–18, 2019.

Deepak Ramachandran and Eyal Amir. Bayesian inverse reinforcement learning. In *IJCAI*, volume 7, pages 2586–2591, 2007.

Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al. Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438, 2008.

Bernard Michini and Jonathan P How. Improving the efficiency of bayesian inverse reinforcement learning. In *IEEE International Conference on Robotics and Automation*, pages 3651–3656. IEEE, 2012.

Sergey Levine, Zoran Popovic, and Vladlen Koltun. Nonlinear inverse reinforcement learning with gaussian processes. *Advances in neural information processing systems*, 24, 2011.

Kyriacos Shiarlis, Joao Messias, and Shimon Whiteson. Inverse reinforcement learning from failure. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2016.

Julien Audiffren, Michal Valko, Alessandro Lazaric, and Mohammad Ghavamzadeh. Maximum entropy semi-supervised inverse reinforcement learning. In *International joint conference on artificial intelligence*, 2015.

Abdeslam Boularias, Jens Kober, and Jan Peters. Relative entropy inverse reinforcement learning. In *Proceedings of the international conference on artificial intelligence and statistics*, pages 182–189. JMLR Workshop and Conference Proceedings, 2011.

Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning. *Advances in neural information processing systems*, 29, 2016.

Chelsea Finn, Paul Christiano, Pieter Abbeel, and Sergey Levine. A connection between generative adversarial networks, inverse reinforcement learning, and energy-based models. *arXiv preprint arXiv:1611.03852*, 2016.

Daehwan Ahn, Abdullah Almaatouq, Monisha Gulabani, and Kartik Hosanagar. Will we trust what we don't understand? impact of model interpretability and outcome feedback on trust in ai. *arXiv preprint arXiv:2111.08222*, 2021.

Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

Arun Das and Paul Rad. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371*, 2020.

Dang Minh, H Xiang Wang, Y Fen Li, and Tan N Nguyen. Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, pages 1–66, 2022.

Ilia Stepin, Jose M Alonso, Alejandro Catala, and Martín Pereira-Fariña. A survey of contrastive and counterfactual explanation generation methods for explainable artificial intelligence. *IEEE Access*, 9:11974–12001, 2021.

Sahil Verma, John Dickerson, and Keegan Hines. Counterfactual explanations for machine learning: A review. *arXiv preprint arXiv:2010.10596*, 2, 2020.

Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.

Joseph Kim, Christian Muise, Ankit Jayesh Shah, Shubham Agarwal, and Julie A Shah. Bayesian inference of linear temporal logic specifications for contrastive explanations. International Joint Conferences on Artificial Intelligence, 2019.

Tathagata Chakraborti, Sarath Sreedharan, Yu Zhang, and Subbarao Kambhampati. Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. In *International Joint Conference on Artificial Intelligence*, pages 156–163, 2017.

Sarath Sreedharan, Siddharth Srivastava, and Subbarao Kambhampati. Hierarchical expertise level modeling for user specific contrastive explanations. In *IJCAI*, pages 4829–4836, 2018.

Roykrong Sukkerd, Reid Simmons, and David Garlan. Towards explainable multi-objective probabilistic planning. In *Proceedings of the International Workshop on Software Engineering for Smart Cyber-Physical Systems*, pages 19–25, 2018.

Ellin Zhao and Roykrong Sukkerd. Interactive explanation for planning-based systems. In *Proceedings of the ACM/IEEE International Conference on Cyberphysical Systems*, *Work-in-Progress track*, 2019.

Wojciech Samek, Grégoire Montavon, Sebastian Lapuschkin, Christopher J Anders, and Klaus-Robert Müller. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3):247–278, 2021.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

L. Shapley. A value for n-person games. In *Classics in game theory*. Princeton University Press, 1953b.

Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.*, 20(177):1–81, 2019.

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR, 2017.

Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

Avanti Shrikumar, Peyton Greenside, Anna Shcherbina, and Anshul Kundaje. Not just a black box: Learning important features through propagating activation differences. *arXiv preprint arXiv:1605.01713*, 2016.

Ahmed Alqaraawi, Martin Schuessler, Philipp Weiß, Enrico Costanza, and Nadia Berthouze. Evaluating saliency map explanations for convolutional neural networks: a user study. In *Proceedings of the international conference on intelligent user interfaces*, pages 275–285, 2020.

Vitali Petsiuk, Rajiv Jain, Varun Manjunatha, Vlad I Morariu, Ashutosh Mehra, Vicente Ordonez, and Kate Saenko. Black-box explanation of object detectors via saliency maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11443–11452, 2021.

Ziyi Liu, Junfeng Gao, Guoguo Yang, Huan Zhang, and Yong He. Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. *Scientific reports*, 6(1): 20410, 2016.

George A Vouros. Explainable deep reinforcement learning: state of the art and challenges. *ACM Computing Surveys*, 55(5):1–39, 2022.

Akanksha Atrey, Kaleigh Clary, and David Jensen. Exploratory not explanatory: Counterfactual analysis of saliency maps for deep reinforcement learning. *arXiv preprint arXiv:1912.05743*, 2019.

Matthew L Olson, Roli Khanna, Lawrence Neal, Fuxin Li, and Weng-Keen Wong. Counterfactual state explanations for reinforcement learning agents via generative deep learning. *Artificial Intelligence*, 295:103455, 2021.

Pedro Sequeira and Melinda Gervasio. Interestingness elements for explainable reinforcement learning: Understanding agents' capabilities and limitations. *Artificial Intelligence*, 288:103367, 2020.

Dan Amir and Ofra Amir. Highlights: Summarizing agent behavior to people. In *Proceedings of the international conference on autonomous agents and multiagent systems*, pages 1168–1176, 2018.

Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. Explainable reinforcement learning through a causal lens. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 2493–2500, 2020.

L. Bainbridge. Ironies of automation. In *Proceedings of IFAC*, volume 19, pages 775–779, 1983.

M. R. Endsley. Ironies of artificial intelligence. *Ergonomics*, 66(11):1656–1668, 2023b.

G. Grote, S. Weik, T. Wäfler, and M. Zölch. Criteria for the complementary allocation of functions in automated work systems and their use in simultaneous engineering projects. *International Journal of Industrial Ergonomics*, 16:367–382, 1995.

S. Eisbach, M. Langer, and G. Hertel. Optimizing human-ai collaboration: Effects of motivation and accuracy information in ai-supported decision-making. *Computers in Human Behavior: Artificial Humans*, 1(2):100015, 2023.

T. Miller. Explainable ai is dead, long live explainable ai! hypothesis-driven decision support, 2023.

T. Ngo and N. Krämer. I humanize, therefore i understand? effects of explanations and humanization of intelligent systems on perceived and objective user understanding. *PsyArXiv*, 2022.

Y. Zhang, Q. V. Liao, and R. K. E. Bellamy. Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 295–305, 2020.

T. Ha and S. Kim. Improving trust in ai with mitigating confirmation bias: Effects of explanation type and debiasing strategy for decision-making with explainable ai. *International Journal of Human–Computer Interaction*, pages 1–12, 2023.

D. Wang, Q. Yang, A. Abdul, and B. Y. Lim. Designing theory-driven user-centric explainable ai. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 1–15, 2019.

Decision making in action: Models and methods. Ablex Pub, 1993.

D. Kahnemann. *Schnelles Denken, langsames Denken*. Verlagsgruppe Random House GmbH, 7 edition, 2012.

G. Klein. Naturalistic decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):456–460, 2008.

G. Klein, K. G. Ross, B. M. Moon, D. E. Klein, R. R. Hoffman, and E. Hollnagel. Macrocognition. *IEEE Computer Society*, pages 81–84, 2003.

G. Klein. *Macrocognitive Measures for Evaluating Cognitive Work*. CRC Press, 2018.

R. R. Hoffman, D. O. Norman, and J. Vagners. Complex sociotechnical joint cognitive work systems. *IEEE Intelligent Systems*, 24(3):82–c3, 2009.

Toni Wäfler and Oliver Rack. *Kooperation und künstliche Intelligenz*, pages 77–88. 2021.

A. Furnham and H. C. Boo. A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1):35–42, 2011.

R. F. Pohl. *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*. Psychology Press, reprint edition, 2006.

A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157): 1124–1131, 1974.

T. D. Wilson, C. E. Houston, K. M. Etling, and N. Brekke. A new look at anchoring effects: Basic anchoring and its antecedents. *Human Factors*, 1996.

P. A. Alexander, D. L. Schallert, and R. E. Reynolds. What is learning anyway? a topographical perspective considered. *Educational Psychologist*, 44(3):176–192, 2009.

D. A. Kolb. *Experimental learning: Experience as the source of learning and development*. Prentice-Hall, 1984.

A. Y. Kolb and D. A. Kolb. The learning way: Meta-cognitive aspects of experiential learning. *Simulation & Gaming*, 40(3):297–327, 2009.

Q. Gao, W. Xu, M. Shen, and Z. Gao. Agent teaming situation awareness (atsa): A situation awareness framework for human-ai teaming, 2023.

B. D. Wilkison, A. D. Fisk, and W. A. Rogers. Effects of mental model quality on collaborative system performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 51, pages 1506–1510, 2007.

L. Rook. Mental models: A robust definition. *The Learning Organization*, 20(1):38–47, 2013.

E. Stöttinger, C. L. Guay, J. Danckert, and B. Anderson. Updating impairments and the failure to explore new hypotheses following right brain damage. *Experimental Brain Research*, 236(6):1749–1765, 2018.

S. Uitdewilligen, M. J. Waller, and A. H. Pitariu. Mental model updating and team adaptation. *Small Group Research*, 44(2):127–158, 2013.

D. F. Valadao, B. Anderson, and J. Danckert. Examining the influence of working memory on updating mental models. *Quarterly Journal of Experimental Psychology*, 68(7):1442–1456, 2015.

E. Pronin. Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, 11 (1):37–43, 2007.

M. Jelodari, M. H. Amirhosseini, and A. Giraldez-Hayes. An ai powered system to enhance self‐reflection practice in coaching. *Cognitive Computation and Systems*, 5(4):243–254, 2023.

F. Lieder, F. Callaway, Yash Raj Jain, P. M. Krueger, Priyam Das, Sayan Gul, and T. L. Griffiths. A cognitive tutor for helping people overcome present bias, 2019.

B. Cahour and J.-F. Forzy. Does projection into use improve trust and exploration? an example with a cruise control system. *Safety Science*, 47(9):1260–1270, 2009.

R. Hoffman. A taxonomy of emergent trusting in the human–machine relationship. In *Cognitive Systems Engineering: The Future for a Changing World*, pages 137–163. CRC Press, 2017.

A. D. Kaplan, T. T. Kessler, J. C. Brill, and P. A. Hancock. Trust in artificial intelligence: Meta-analytic findings. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 65(2):337–359, 2023.

R. Hoffman, S. T. Mueller, G. Klein, and J. Litman. Measuring trust in the xai context. *DARPA*, 2018.

J. D. Lee and K. A. See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 2004.

R. Parasuraman and V. Riley. Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2):230–253, 1997.

R. Parasuraman and D. H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(3):381–410, 2010.

P. Koopman and R. R. Hoffman. Work-arounds, make-work, and kludges. *IEEE Intelligent Systems*, 18 (6):70–75, 2003.

A. Jacovi, A. Marasović, T. Miller, and Y. Goldberg. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai, 2021.

R. Fildes, P. Goodwin, M. Lawrence, and K. Nikolopoulos. Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1):3–23, 2009.

S. Niehaus, M. Hartwig, P. H. Rosen, and S. Wischniewski. An occupational safety and health perspective on human in control and ai. *Frontiers in Artificial Intelligence*, 5:868382, 2022.

G. Schaap, T. Bosse, and P. Hendriks Vettehen. The abc of algorithmic aversion: Not agent, but benefits and control determine the acceptance of automated decision-making. *AI & SOCIETY*, 2023.

Erinaldi, ErniMasdupi, and Syahrizal. The effect transformational leadership, self efficacy, intrinsic motivation and job engagement on proactive behavior. In *Proceedings of the Padang International Conference On Economics Education, Economics, Business and Management, Accounting and Entrepreneurship*, 2020.

A. Brizon and J. L. Wybo. The life cycle of weak signals related to safety. *International Journal of Emergency Management*, 6(2):117, 2009.

S. Duchek. Organizational resilience: A capability-based conceptualization. *Business Research*, 13(1): 215–246, 2020.

D. B. Kaber, N. Stoll, K. Thurow, R. S. Green, S. Kim, and P. Mosaly. Human–automation interaction strategies and models for life science applications. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 19(6):601–621, 2009.

S. K. Parker and G. Grote. Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. *Applied Psychology*, 71(4):1171–1204, 2022.

R. M. Ryan and E. L. Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54–67, 2000.

Y. Fried and G. R. Ferris. The validity of the job characteristics model: a review and meta-analysis. *Personnel Psychology*, 40(2):287–322, 1987.

J. R. Hackman and G. R. Oldham. Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*, 16(2):250–279, 1976.

A. Rai and S. Maheshwari. Exploring the mediating role of work engagement between the linkages of job characteristics with organizational engagement and job satisfaction. *Management Research Review*, 44(1):133–157, 2021.

S. K. Parker, F. P. Morgeson, and G. Johns. One hundred years of work design research: Looking back and looking forward. *Journal of Applied Psychology*, 102(3):403–420, 2017.

S. Sadeghian and M. Hassenzahl. The ”artificial” colleague: Evaluation of work satisfaction in collaboration with non-human coworkers. In *International Conference on Intelligent User Interfaces*, pages 27–35, 2022.

F. P. Morgeson, K. Delaney-Klinger, and M. A. Hemingway. The importance of job autonomy, cognitive ability, and job-related skill for predicting role breadth and job performance. *Journal of Applied Psychology*, 90(2):399–406, 2005.

T. Wäfler, A. Windischer, C. Ryser, S. Weick, and G. Grote. Wie sich mensch und technik sinnvoll ergänzen. die gestaltung automatisierter produktionssysteme mit kompass. Vdf Hochschulverlag AG an der ETH Zürich, 1999.

M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

M. R. Endsley. Supporting human-ai teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior*, 140:107574, 2023c.

Joseph CR Licklider. Man-computer symbiosis. *IRE transactions on human factors in electronics*, (1): 4–11, 1960.

Achim Buerkle, William Eaton, Niels Lohse, Thomas Bamber, and Pedro Ferreira. Eeg based arm movement intention recognition towards enhanced safety in symbiotic human-robot collaboration. *Robotics and Computer-Integrated Manufacturing*, 70:102137, 2021.

Luzheng Bi, Cuntai Guan, et al. A review on emg-based motor intention prediction of continuous human upper limb motion for human-robot collaboration. *Biomedical Signal Processing and Control*, 51:113–127, 2019.

Junyao Wang, Yuehong Dai, and Xiaxi Si. Analysis and recognition of human lower limb motions based on electromyography (emg) signals. *Electronics*, 10(20):2473, 2021.

John Wenskovitch, Alexander Anderson, Slaven Kincic, Corey Fallon, Danielle Ciesielski, Jessica Baweja, Molly C Mersinger, and Brett Jefferson. Operator insights and usability evaluation of machine learning assistance for power grid contingency analysis. *Human Factor in Energy: Oil, Gas, Nuclear, and Electric Power*, 2022.

Alexander A Anderson, Brett A Jefferson, Slaven Kincic, John E Wenskovitch, Corey K Fallon, Jessica A Baweja, and Yousu Chen. Human-centric contingency analysis metrics for evaluating operator performance and trust. *IEEE Access*, 2023.

Susana Rodrigues, Duarte Dias, Marta Aleixo, Antonio Retorta, and João Paulo S Cunha. Implementing a quantified occupational health sensing platform in the aviation sector: an exploratory study in routine air traffic control work shifts. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 7162–7165. IEEE, 2021.

Clark Borst, John M. Flach, and Joost Ellerbroek. Beyond ecological interface design: Lessons from concerns and misconceptions. *IEEE Transactions on Human-Machine Systems*, 45:164–175, 2015.

Marinus M. Van Paassen, C. Borst, J. Ellerbroek., M. Mulder, and J. M. Flach. Ecological interface design for vehicle locomotion control. *IEEE Transactions on Human-machine systems*, 48:541–555, 2018.

Rolf Klomp, Clark Borst, Rene van Paassen, and Max Mulder. Expertise level, control strategies, and robustness in future air traffic control decision aiding. *IEEE Transactions on Human-Machine Systems*, 46:255–266, 2015.

Jonas Lundberg, Mattias Arvola, and Karljohan Lundin Palmerius. Human autonomy in future drone traffic: Joint human–ai control in temporal cognitive work. *Frontiers in Artificial Intelligence*, 4, 2021.

Jonas Lundberg and Björn J. E. Johansson. A framework for describing interaction between human operators and autonomous, automated, and manual control systems. *Cogn. Technol. Work*, 23(3): 381–401, 2021.

Jimmy Johansson Westberg, Karljohan Lundin Palmerius, and Jonas Lundberg. Utm city—visualization of unmanned aerial vehicles. *IEEE Computer Graphics and Applications*, 42(5):84–89, 2022.

Jonas Lundberg, Magnus Nylin, Gesa Praetorius, Anders Jansson, and Magnus Bång. Modelling operator control work across traffic management domains: implications for interaction design. *Cognition, Technology & Work*, pages 1–19, 2024.

Magnus Nylin, Jimmy Johansson Westberg, and Jonas Lundberg. Reduced autonomy workspace (raw)—an interaction design approach for human-automation cooperation. *Cognition, Technology & Work*, 24, 2022.

Jimmy Hammarbäck, Jens Alfredson, Björn J. E. Johansson, and Jonas Lundberg. My synthetic wingman must understand me: modelling intent for future manned–unmanned teaming. *Cogn. Technol. Work*, 26(1):107–126, 2023.

Jimmy Hammarbäck, Björn J. E. Johansson, Jens Alfredson, and Jonas Lundberg. Modelling situated intent for human-autonomy teaming: a human-centric approach. *Theoretical Issues in Ergonomics Science*, 0(0):1–22, 2024.

Laura Von Rueden, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, Julius Pfrommer, Annika Pick, Rajkumar Ramamurthy, et al. Informed machine learning–a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):614–633, 2021.

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

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.

Russell Kaplan, Christopher Sauer, and Alexander Sosa. Beating atari with natural language guided reinforcement learning. *arXiv preprint arXiv:1704.05539*, 2017.

Frank Van Harmelen and Annette Ten Teije. A boxology of design patterns for hybrid learning and reasoning systems. *Journal of Web Engineering*, 18(1-3):97–123, 2019.

Dongran Yu, Bo Yang, Dayou Liu, Hui Wang, and Shirui Pan. A survey on neural-symbolic learning systems. *Neural Networks*, 2023.

Pascal Hitzler, Aaron Eberhart, Monireh Ebrahimi, Md Kamruzzaman Sarker, and Lu Zhou. Neuro-symbolic approaches in artificial intelligence. *National Science Review*, 9(6):nwac035, 2022.

Artur d'Avila Garcez, Sebastian Bader, Howard Bowman, Luis C Lamb, Leo de Penning, BV Illuminoo, Hoifung Poon, and COPPE Gerson Zaverucha. Neural-symbolic learning and reasoning: A survey and interpretation. *Neuro-Symbolic Artificial Intelligence: The State of the Art*, 342(1):327, 2022.

Md Kamruzzaman Sarker, Lu Zhou, Aaron Eberhart, and Pascal Hitzler. Neuro-symbolic artificial intelligence. *AI Communications*, 34(3):197–209, 2021.

Huan Zhao, Junhua Zhao, Jing Qiu, Gaoqi Liang, and Zhao Yang Dong. Cooperative wind farm control with deep reinforcement learning and knowledge-assisted learning. *IEEE Transactions on Industrial Informatics*, 16(11):6912–6921, 2020.

Huan Zhao, Zifan Liu, Xuan Mai, Junhua Zhao, Jing Qiu, Guolong Liu, Zhao Yang Dong, and Amer MYM Ghias. Mobile battery energy storage system control with knowledge-assisted deep reinforcement learning. *Energy Conversion and Economics*, 3(6):381–391, 2022.

Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Elise Van der Pol, Daniel Worrall, Herke van Hoof, Frans Oliehoek, and Max Welling. MDP homomorphic networks: Group symmetries in reinforcement learning. *Advances in Neural Information Processing Systems*, 33:4199–4210, 2020.

Jan Wöhlke, Felix Schmitt, and Herke Hoof. Value refinement network (vrn). In *International Joint Conference on Artificial Intelligence*, 2022.

Huatong Dai, Pengzhan Chen, and Hui Yang. Driving torque distribution strategy of skid-steering vehicles with knowledge-assisted reinforcement learning. *Applied Sciences*, 12(10):5171, 2022.

Yuhong Xie, Yuan Zhang, Tao Lin, Zipeng Pan, Si-Ze Qian, Bo Jiang, and Jinyao Yan. Short video preloading via domain knowledge assisted deep reinforcement learning. *Digital Communications and Networks*, 2024.

Jiemai Gao, Siyuan Chen, Xiang Li, and Jun Zhang. Transient voltage control based on physics-informed reinforcement learning. *IEEE Journal of Radio Frequency Identification*, 6:905–910, 2022.

Marta Garnelo, Kai Arulkumaran, and Murray Shanahan. Towards deep symbolic reinforcement learning. *arXiv preprint arXiv:1609.05518*, 2016.

Artur d'Avila Garcez, Aimore Resende Riquetti Dutra, and Eduardo Alonso. Towards symbolic reinforcement learning with common sense. *arXiv preprint arXiv:1804.08597*, 2018.

Niklas Höpner, Ilaria Tiddi, and Herke van Hoof. Leveraging class abstraction for commonsense reinforcement learning via residual policy gradient methods. In *International Joint Conference on Artificial Intelligence, IJCAI 2022*, pages 3050–3056. International Joint Conferences on Artificial Intelligence Organization, 2022.

Daoming Lyu, Fangkai Yang, Bo Liu, and Steven Gustafson. Sdrl: interpretable and data-efficient deep reinforcement learning leveraging symbolic planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2970–2977, 2019.

Fangkai Yang, Daoming Lyu, Bo Liu, and Steven Gustafson. Peorl: Integrating symbolic planning and hierarchical reinforcement learning for robust decision-making. *arXiv preprint arXiv:1804.07779*, 2018.

Brandon Araki, Xiao Li, Kiran Vodrahalli, Jonathan DeCastro, Micah Fry, and Daniela Rus. The logical options framework. In *International Conference on Machine Learning*, pages 307–317. PMLR, 2021.

Pashootan Vaezipoor, Andrew C Li, Rodrigo A Toro Icarte, and Sheila A Mcilraith. Ltl2action: Generalizing ltl instructions for multi-task rl. In *International Conference on Machine Learning*, pages 10497–10508. PMLR, 2021.

Ludovico Mitchener, David Tuckey, Matthew Crosby, and Alessandra Russo. Detect, understand, act: A neuro-symbolic hierarchical reinforcement learning framework. *Machine Learning*, 111(4):1523–1549, 2022.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, May 2021.

Thomas M Moerland, Joost Broekens, Aske Plaat, Catholijn M Jonker, et al. Model-based reinforcement learning: A survey. *Foundations and Trends in Machine Learning*, 16(1):1–118, 2023.

Geoffrey G Towell and Jude W Shavlik. Knowledge-based artificial neural networks. *Artificial intelligence*, 70(1-2):119–165, 1994.

David D Woods. Decomposing Automation: Apparent Simplicity, Real Complexity. *Automation and Human Performance*, 1996.

Charles Rich and Candace L. Sidner. Collagen: when agents collaborate with people. In *Proceedings of the First International Conference on Autonomous Agents*, page 284–291. Association for Computing Machinery, 1997.

Julian Abich and Eric Sikorski. Taking a Constructivist Approach to Human-AI Co-learning Design. In *Make it Happen*, 2023.

Tjeerd A.J. Schoonderwoerd, Emma M. Van Zoelen, Karel Van Den Bosch, and Mark A. Neerincx. Design patterns for human-AI co-learning: A wizard-of-Oz evaluation in an urban-search-and-rescue task. *International Journal of Human-Computer Studies*, 164:102831, 2022.

Kevin T. Wynne and Joseph B. Lyons. An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*, 19:353–374, 2018.

Samuel Westby and Christoph Riedl. Collective Intelligence in Human-AI Teams: A Bayesian Theory of Mind Approach, 2023.

Siddharth Swaroop, Zana Buçinca, and Finale Doshi-Velez. Adaptive interventions for both accuracy and time in AI-assisted human decision making, 2023.